Takashi Washio
Ken Satoh
Hideaki Takeda
Akihiro Inokuchi (Eds.)

LNAI 4384

New Frontiers in Artificial Intelligence

JSAI 2006 Conference and Workshops
Tokyo, Japan, June 2006
Revised Selected Papers

Springer

Lecture Notes in Artificial Intelligence 4384

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Takashi Washio   Ken Satoh
Hideaki Takeda   Akihiro Inokuchi (Eds.)

New Frontiers in Artificial Intelligence

JSAI 2006 Conference and Workshops
Tokyo, Japan, June 5-9, 2006
Revised Selected Papers

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Takashi Washio
Osaka University
The Institute of Sientific and Industrial Research (ISIR)
8-1, Mihogaoka, Ibarakishi, Osaka, 567-0047, Japan
E-mail: washio@sanken.osaka-u.ac.jp

Ken Satoh
Hideaki Takeda
National Institute of Informatics
Foundations of Information Research Division
Chiyoda-ku, 2-1-2, Hitotsubashi, Tokyo 101-8430, Japan
E-mail: {ksatoh,takeda}@nii.ac.jp

Akihiro Inokuchi
Tokyo Research Laboratory (TRL), IBM Japan
1623-14 Shimotsuruma, Yamato-shi, Kanagawa, 242-8502, Japan
E-mail: inokuchi@jp.ibm.com

Library of Congress Control Number: 2006940650

CR Subject Classification (1998): I.2, H.2.8, H.3, F.1, H.4, H.5.2, I.5, J.1, J.3, K.4.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-69901-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-69901-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11978893 06/3142 5 4 3 2 1 0

Preface

The progress in information technology including artificial intelligence (AI) in the last few decades is remarkable, and has attracted many young researchers to this field. This trend is now accelerated along with the recent rapid growth of computer communication networks and the worldwide increase in researchers. In this context, we have observed many outstanding AI studies in Japanese domestic conferences. They have high technical originality, quality and significance. The annual conference of JSAI (Japan Society for Artificial Intelligence) is one of the key and representative domestic meetings in the field of intelligent information technology. Award papers in this conference have an excellent quality of international standards. The annual conference of JSAI also organizes co-located international workshops to provide excellent study reports.

The objectives of this book are to present the award papers of the 20th annual conference of JSAI 2006 and selected papers from the three co-located international workshops and to promote the study exchange among researchers worldwide. Eight papers were awarded among more than 200 presentations in the conference, and 21 papers were selected from a total of 44 presentations in the workshops of Logic and Engineering of Natural Language Semantics 2006 (LENLS 2006), Learning with Logics and Logics for Learning (LLLL 2006) and Risk Mining (RM 2006). The award papers in the 20th annual conference of JSAI 2006 were selected from presentations covering the wide field of artificial intelligence through the processes of candidate recommendations, detailed open discussions and voting by Program Committee members of the conference. The LENLS workshop series is organized under the aim of bringing together researchers working on information structure and/or dynamic semantics for natural language. LENLS 2006 focused on formal pragmatics in particular. The LLLL 2006 workshop was held to bring together researchers who are interested in the areas of machine learning and computational logic, and to have intensive discussions on various relations between the two thereby making their interchange more active. RM 2006 was held with the aim of sharing and comparing experiences on risk mining techniques applied to risk detection, risk clarification and risk utilization in real fields.

We hope this book introduces the excellent Japanese studies on AI to the world and contributes to the growth of the worldwide community of AI researchers.

November 2006 Takashi Washio

Organization and Editorial Board

The award papers were selected by the Program Committee of the annual conference of JSAI (Japan Society for Artificial Intelligence) 2006. The paper selection of each co-located international workshop was made by the Program Committee of each workshop. Upon the decisions of the paper awards and the paper selections, each chapter was edited by the Program Chairs of the 20th annual JSAI conference and the co-located international workshops. The entire contents and structure of the book were managed and edited by the chief editors.

Volume Editors

Primary Volume Editor: Takashi Washio (Osaka University, Japan)
Ken Satoh (National Institute of Informatics, Japan)
Hideaki Takeda (National Institute of Informatics, Japan)
Akihiro Inokuchi (Tokyo Research Laboratory, IBM Japan)

Chapter Editors (Program Chairs)

Award Paper Chapter:	Hideaki Takeda (National Institute of Informatics, Japan)
LENLS 2006 Chapter:	Eric McCready (Aoyama Gakuin University, Japan)
LLLL 2006 Chapter:	Akihiro Yamamoto (Kyoto University, Japan)
	Kouichi Hirata (Kyushu Institute of Technology, Japan)
	Ken Satoh (National Institute of Informatics, Japan)
RM 2006 Chapter:	Shusaku Tsumoto (Shimane University, Japan)
	Takashi Washio (Osaka University, Japan)

Table of Contents

New Frontiers in Artificial Intelligence: Proceedings of the 20th Annual Conferences of the Japanese Society for Artificial Intelligence

Part I Awarded Papers

Part II Logic and Engineering of Natural Language Semantics

Part III Learning with Logics and Logics for Learning

Part IV Risk Mining

Overview of Awarded Papers – The 20th Annual Conference of JSAI

Hideaki Takeda

National Institute of Informatics and the University of Tokyo
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
takeda@nii.ac.jp

In this chapter, we proudly introduce eight awarded papers, selected from the papers presented in the 20th annual conference of Japan Society for Artificial Intelligence (JSAI2006).

The conference was held at Tower Hall Funabori located in the east part of Tokyo from June 7 until June 9, 2006. 276 papers were presented in about 60 sessions and over 500 people participated in the conference. Sessions vary from the fundamental issues to the state-of-the-art applications. Session for regular papers are as follows; Logic and learning, Reinforcement learning and agent learning, Agent theory, Auction/Game/Economics, Agent learning, Agent planning, Agent simulation and interaction, Genetic algorithm, Image processing, Information extraction and classification, Clustering/self-organization, Classification learning, Text mining, Graph mining, Mining algorithm, Web mining, Pre- and post-processing for data mining, Practices of data mining, Cognitive modeling, Language processing and dialogue, Robot/sensor network, Web information system, Semantic Web, Knowledge modeling and knowledge sharing, Support of knowledge management, knowledge modeling/ontology, Web service, Human interface and communication support, Education support, Learning support environment, and Musical and auditory information processing.

In addition to these sessions, we have organized sessions that include Socio-infomation infrastructure for intelligent support, Even space information support project, Carrier design for researchers, Computing for semantics and understanding, Natural language processing for knowledge-oriented text mining, Language - computer - communication, Experience media, Intercultural collaboration and AI, Data mining for risk information, and Human-agent interaction. The research area covered by these sessions is wider than the other AI conferences such as IJCAI, AAAI, and ECAI. It indicates that Japanese AI community is so active and eager to explore new topics that its topics are expanding beyond the traditional AI areas.

Among them we selected eleven papers as awarded papers. In order to decide award papers, we asked the program committee members and chairs to review the papers. PC members mainly reviewed the papers and the chairs reviewed both the papers and the presentations. It was very competitive result in this year so that we decided to award top eleven papers. Unfortunately the authors of three of eleven papers were reluctant to rewrite their papers in English (the original papers were written in Japanese). As a result, we have eight awarded papers here.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 3–4, 2007.

As well as sessions, the awarded papers varies from the fundamental issues such as the fundamental consideration of data mining (Tsuyoshi Ide) to the state-of-the-art application such as "knowledge globe" (Hideyuki Kubota *et al.*) and discussion support (Hideyuki Tomobe and Katashi Nagao). In particular, the variety of the applications suggest the potential of the AI approach such as bioinformatics (Hiroshi Yamakawa *et al.*) and musical information processing (Keiji Hirata and Satoshi Tojo). Another uniqueness is mixture of different disciplines; from robotics (Masayuki Furuyama *et al.* and Tetsunari Inamura *et al.*) to psychology (Asako Miura *et al.*). AI becomes the "meeting point" for different disciplines.

Finally we would like to thank all the attendees who joined the discussion that contributes selection of the awarded papers.

JSAI2006 Program Committee

Hideaki Takeda, Chair
(National Institute of Informatics and the University of Tokyo)
Katashi Nagao, Vice-chair (Nagoya University)
Takayuki Ito (Nagoya Institute of Technology),
Akihiro Inokuchi (IBM, Japan)
Michita Imai (Keio University)
Ryohei Orihara (Toshiba Cooperation)
Daisuke Katagami (Tokyo Institute of Technology),
Ken Kaneiwa
(The National Institute of Information and Communications Technology)
Takanori Komatsu (Future University-Hakodate),
Ken Satoh (National Institute of Informatics),
Kiyoaki Shirai (Japan Advanced Institute of Science and Technology),
Masahiro Terabe (Mitsubishi Research Institute, Inc.),
Takuichi Nishimura
(National Institute of Advanced Industrial Science and Technology)
Taro Fujimoto (Fujitsu Laboratories Ltd.)
Takashi Miyata (Justsystems Corporation)
Yutaka Matsuo
(National Institute of Advanced Industrial Science and Technology)
Shigeo Matsubara (NTT Communication Science Laboratories)
Tsuyoshi Murata (Tokyo Institute of Technology)
Masaki Murata
(The National Institute of Information and Communications Technology)
Satoshi Morinaga (NEC Cooperation)
Ikuko Eguchi Yairi
(The National Institute of Information and Communications Technology)
Makoto Yokoo (Kyushu University)
Takashi Washio (Osaka University)

Translational Symmetry in Subsequence Time-Series Clustering

Tsuyoshi Idé

IBM Research, Tokyo Research Laboratory
1623-14 Shimotsuruma, Yamato, 242-8502 Kanagawa, Japan
goodidea@jp.ibm.com

Abstract. We treat the problem of subsequence time-series clustering (STSC) from a group-theoretical perspective. First, we show that the sliding window technique introduces a mathematical artifact to the problem, which we call the pseudo-translational symmetry. Second, we show that the resulting cluster centers are necessarily governed by irreducible representations of the translational group. As a result, the cluster centers necessarily forms sinusoids, almost irrespective of the input time-series data. To the best of the author's knowledge, this is the first work which demonstrates the interesting connection between STSC and group theory.

1 Introduction

Learning representative patterns from time series data is one of the most interesting tasks in data mining. Since the seminal work of Das et al. [1], subsequence time-series clustering (STSC) had enjoyed popularity as the simplest and the most reliable technique of stream mining. In STSC, time series data is represented as a set of subsequence vectors generated using a sliding window (see Fig. 1 (a)), and the generated subsequences are grouped using k-means clustering (Fig. 1 (b)). The cluster centers (the mean vectors of the cluster members) are thought of as representative patterns of the time series.

Currently, however, k-means STSC is considered to make little sense as a pattern discovery technique, since, as first pointed out by Keogh et al. [9], k-means STSC is "meaningless" in that the resultant cluster centers tend to form sinusoidal pseudo-patterns almost independent of the input time series. This *sinusoid effect* proved that even the simplest algorithms such as k-means STSC could be too dangerous to be used unless the mathematical structures are fully understood. We believe that the sinusoid effect raised a question to the general trend in the stream mining community, where seemingly plausible analysis tends to be accepted without theoretical justifications.

In a previous paper [6], we theoretically studied the origin of the sinusoid effect. The original k-means STSC task was reduced to a spectral STSC task, and sinusoidal cluster centers were explicitly obtained by solving an eigen problem. In this paper, we discuss mathematical properties of STSC in more detail. In

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 5–18, 2007.

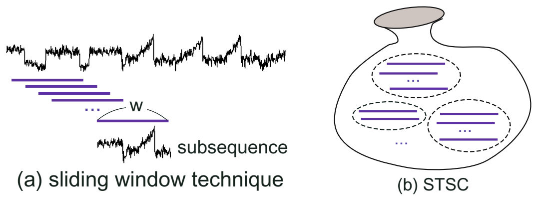


Fig. 1. (a) Sliding window technique to generate subsequences. (b) The generated subsequences are grouped as independent data items.

particular, we will point out that the cluster centers are inevitably governed by irreducible representations of the translational group, because of a hidden translational symmetry introduced by the sliding window technique. To the best of the author's knowledge, this is the first work that points out the interesting connection between STSC and group theory.

The layout of this paper is as follows: In Section 2, we reformulate STSC as the problem of linear algebra in a vector space, and introduce the notion of linear operators. In Section 3, we review the connection between k-means and spectral STSC. In Section 4, we introduce the concept of translational group, and explain its implications in spectral STSC. In Section 5, we derive the solution to spectral STSC from a group-theoretical perspective. In Section 6, we summarize the paper.

2 Lattice Model for Time Series Analysis

In this section, we introduce a lattice model for time series analysis, and show that this model provides us with a very handy way to express the subsequences of time series data. Throughout this paper, we use $\mathbb{R}$ and $\mathbb{C}$ to represent the sets of real and complex numbers, respectively.

2.1 Vector Space in Dirac's Notation

A common approach to express time-series data is to use a scalar function such as $x(t)$. However, this notation is not very effective in describing symmetry properties of the problem. We believe that this has made it difficult to pinpoint the origin of the sinusoid effect. Instead, we introduce a lattice model in Dirac's notation to represent time series data. While Dirac's notation [12,10] is mathematically equivalent to the standard vector-matrix notation, it is much more powerful for describing linear operators, which play an essential role in this paper. In this subsection, we illustrate the notion of Dirac's notation, by following [12].

Let $\mathcal{H}_0$ be a vector space spanned by n linearly independent bases $\{|1\rangle, |2\rangle, ..., |n\rangle\}$. By definition, any vector in $\mathcal{H}_0$ is represented as a linear combination of these bases. For example, a vector $|a\rangle \in \mathcal{H}_0$ may be expressed as

$$|a\rangle = \sum_{l=1}^{n} a_l |l\rangle,$$

where a_ls are constants (generally complex numbers).

To introduce the metric into $\mathcal{H}_0$, we require that each of $|l\rangle$ has a unique counterpart in a dual space of $\mathcal{H}_0$. We denote the counterpart by $\langle l|$, and define that $c|a\rangle$ dual-corresponds to $c^*\langle l|$ where c is a complex constant and * denotes complex conjugate. Now the inner product between vectors $|a\rangle, |b\rangle \in \mathcal{H}_0$ is defined as $\langle a|b\rangle \equiv \langle a| \cdot |b\rangle$, which is generally a complex number. Regarding the other choice $\langle b|a\rangle$, we assume

$$\langle b|a\rangle = [\langle a|b\rangle]^* \tag{1}$$

as a premise. For example, the inner product between the above $|a\rangle$ and $|b\rangle = \sum_{l=1}^{n} b_l |l\rangle$ will be $\langle a|b\rangle = \sum_{l,l'} a_{l'}{}^* b_l \langle l'|l\rangle$, which is computable if $\langle l'|l\rangle$s are given.

As usual, we also require that the squared norm is nonnegative, i.e. $\langle a|a\rangle \geq 0$ for $\forall |a\rangle \in \mathcal{H}_0$. Using the norm, any nonzero vector in $\mathcal{H}_0$ can be normalized as $|a\rangle/\sqrt{\langle a|a\rangle}$ to have the unit norm. We assume that the bases $\{|l\rangle\}$ have been chosen to be orthonormal, i.e.,

$$\langle l|l'\rangle = \delta_{l,l'},$$

where $\delta_{l,l'}$ is Kronecker's delta.

2.2 Linear Operators in $\mathcal{H}_0$

Let $\mathcal{L}$ be the set of linear operators which transforms a vector in $\mathcal{H}_0$ into another vector. We distinguish the operators from ordinary numbers by using ˆ hereafter. By definition, $\forall \hat{o} \in \mathcal{L}$ has an expression

$$\hat{o} = \sum_{l,l'=1}^{n} o_{l,l'} |l\rangle\langle l'|, \tag{2}$$

where $o_{l,l'} \in \mathbb{C}$ is called the (l, l') element of $\hat{o}$. Since $\{o_{l,l'}\}$ uniquely specifies $\hat{o}$ under a given orthonormal basis set, the $n \times n$ matrix $[o_{l,l'}]$ can be thought of as a matrix representation of $\hat{o}$. For $\forall |a\rangle \in \mathcal{H}_0$, we denote the dual element of $\hat{o}|a\rangle$ as $\langle a|\hat{o}^\dagger$, and call $\hat{o}^\dagger$ the Hermitian conjugate of $\hat{o}$. Considering the fact that $\sum_{l,l'=1}^{n} (o_{l,l'}\langle l'|a\rangle)^* \langle l|$ is the dual element of $\hat{o}|a\rangle$, we see that $\hat{o}^\dagger$ has the expression as

$$\hat{o}^\dagger = \sum_{l,l'=1}^{n} o_{l',l}{}^* |l\rangle\langle l'|. \tag{3}$$

A linear operator $\hat{o}$ such that $\hat{o}^\dagger = \hat{o}$ is called Hermitian.

As an example, consider a Hermitian operator $\hat{\vartheta}(w) \in \mathcal{L}$

$$\hat{\vartheta}(w) \equiv \sum_{l'=1}^{w} |l'\rangle\langle l'|, \tag{4}$$

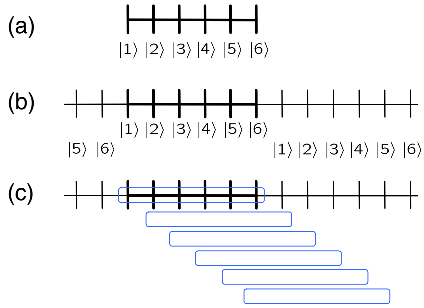


Fig. 2. (a) One-dimensional lattice with $n = 6$. (b) One-dimensional lattice under the periodic boundary condition. (c) Subsequences when $w = n = 6$.

where $w \leq n$. To understand how this operator works, imagine an equi-interval one-dimensional lattice as shown in Fig. 2 (a), where each of the basis is attached to each site (lattice point). Since $\hat{\vartheta}(w)|l\rangle$ vanishes for $l > w$, and $\hat{\vartheta}(w)|l\rangle = |l\rangle$ otherwise, we see that $\hat{\vartheta}(w)$ works as the "cut-off operator". For example, if $n = 6$ and $w = 3$, $\hat{\vartheta}(3)$ simply cuts off the portion which is not contained by the 3-dimensional lattice of $\{|1\rangle, |2\rangle, |3\rangle\}$, remaining the rest unchanged.

It is interesting to see how $\hat{\vartheta}(n)$ works. Clearly, this operator remains any vector in $\mathcal{H}_0$ the same. In other words, this is the *identity operator*. Explicitly, we define the identity operator in $\mathcal{H}_0$ as

$$\hat{1} \equiv \sum_{l'=1}^{n} |l'\rangle\langle l'|. \tag{5}$$

This operator is very useful when one wants to change the representation. For example, $\forall |a\rangle \in \mathcal{H}_0$ is equivalent to $\hat{1}|a\rangle$, so that $|a\rangle = \sum_{l=1}^{n} |l\rangle\langle l|a\rangle$ holds. Since $\langle l|a\rangle$s are just scalar, this equation gives the representation of $|a\rangle$ by $\{|1\rangle, ..., |n\rangle\}$.

2.3 Lattice Model for Time Series Data

Now let us associate $\mathcal{H}_0$ with time-series data. Consider an equi-interval time series data with length n of $\{x_t \in \mathbb{R} \mid t = 1, 2, ..., n\}$. From a geometrical viewpoint, this can be viewed as a vector in the n-dimensional space $\mathcal{H}_0$:

$$|\Gamma\rangle = \sum_{l=1}^{n} x_l |l\rangle. \tag{6}$$

In Fig. 2 (a), this definition amounts to that each x_l is attached to the l-th site. We call this expression the site representation of the time-series data. The coefficients x_l can be obtained by $x_l = \langle l|\Gamma\rangle$. By using $\hat{1}$, one can explicitly compute the squared norm of $|\Gamma\rangle$ as

$$\langle \Gamma | \Gamma \rangle = \langle \Gamma | \hat{1} \cdot \hat{1} | \Gamma \rangle = \sum_{l=1}^{n} \langle \Gamma | l \rangle \langle l | \sum_{l'=1}^{n} | l' \rangle \langle l' | \Gamma \rangle = \sum_{l=1}^{n} \langle \Gamma | l \rangle \langle l | \Gamma \rangle = \sum_{l=1}^{n} |x_l|^2,$$

which is just the squared sum for real valued time series data. We may simply denote this as $\| \, |\Gamma\rangle \, \|^2$.

Hereafter we impose the *periodic boundary condition* (PBC) on time-series data. As indicated in Fig. 2 (b), $\forall l, |l+n\rangle = |l\rangle$ holds under the PBC. As long as $n \gg w$, the result of STSC will not be greatly affected by this artificial condition.

2.4 Translation Operator on the Lattice

As another instance of linear operators, let us focus on the translation operator $\hat{\tau}(l) \in \mathcal{L}$

$$\hat{\tau}(l) \equiv \sum_{l'=1}^{n} |l' + l\rangle \langle l'|. \tag{7}$$

The operator $\hat{\tau}(l)$ shifts the basis in the site representation with l steps. To see this, for example, consider $\hat{\tau}(l)|2\rangle$. Thanks to the orthogonality, it follows

$$\hat{\tau}(l)|2\rangle = \sum_{l'=1}^{n} |l' + l\rangle \langle l'|2\rangle = \sum_{l'=1}^{n} |l' + l\rangle \delta_{l',2} = |2 + l\rangle.$$

By Eq. (3), it follows $\hat{\tau}(l)^\dagger = \hat{\tau}(-l)$, meaning that $\hat{\tau}(l)$s are *unitary* operator. In general, $\hat{o} \in \mathcal{L}$ is said unitary if $\hat{o}\hat{o}^\dagger = \hat{o}^\dagger \hat{o} = \hat{1}$ holds.

The translation operator provides us with a handy way to express subsequences in STSC. If we use the expression of Eq. (6), the p-th subsequence with length w (i.e. *the window size* is w) is given by $\sum_{l=p+1}^{p+w} x_l |l\rangle$. While this subsequence should be originally viewed as a vector in $\mathcal{H}_0$, it is clearly redundant in that only w dimensions are used out of the n dimensions. It is more reasonable to think of $|s_p\rangle$ as a vector in a subspace $\mathcal{H} \equiv \{|1\rangle, ..., |w\rangle\}$. Explicitly, we define $|s_p\rangle$ as

$$|s_p\rangle = \sum_{l=1}^{w} x_{l-p} |l\rangle = \hat{\vartheta}(w) \hat{\tau}(-p) |\Gamma\rangle \tag{8}$$

under the PBC.

Here we define the notion of *translational invariance* of linear operators:

Definition 1 (Translational invariance). *An operator $\hat{o} \in \mathcal{L}$ is said to be translationally invariant if*

$$\hat{o} = \hat{\tau}(l)^\dagger \hat{o} \hat{\tau}(l) \tag{9}$$

holds for $\forall l \in \{0, 1, .., n-1\}$.

The intuition behind this is that the matrix element of $\hat{o}$ between $\forall |a\rangle$ and $|b\rangle \in \mathcal{H}_0$ remains the same as that between $\hat{\tau}(l)|a\rangle$ and $\hat{\tau}(l)|b\rangle$. Since $\hat{\tau}(-l) = \hat{\tau}(l)^{-1}$ by definition, the invariance condition is equivalent to $\hat{o}\hat{\tau}(l) = \hat{\tau}(l)\hat{o}$. In other words, any operators invariant to translations must commute with $\hat{\tau}(l)$s.

3 Spectral Clustering of Subsequences

In this Section, we derive an eigen equation whose eigen vectors corresponds to the k-means cluster centers. It essentially follows the formulation in [6], but is the first treatment of spectral STSC with Dirac's notation.

As before, we use the whole space $\mathcal{H}_0 = \{|1\rangle, ..., |n\rangle\}$ and its subspace $\mathcal{H} = \{|1\rangle, ..., |w\rangle\}$ with $w \leq n$. Notice that we do *not* assume any periodicity in $\mathcal{H}$ unless $w = n$, despite the fact $\mathcal{H}_0$ is always periodic. The k-means STSC task is to group a set of vectors $\{|s_q\rangle \in \mathcal{H} \,|q = 1, 2, ..., n\}$, where the subsequences are thought of as vectors in $\mathcal{H}$.

It is well-known that the k-means algorithm attempts to minimize the sum-of-squared (SOS) error [4]. In our notation, the SOS error is written as

$$E = \sum_{j=1}^{k} \sum_{p \in \mathcal{C}_j} \left|\left| \, |s_p\rangle - |m^{(j)}\rangle \, \right|\right|^2 = \sum_{p=1}^{n} \langle s_p|s_p\rangle - \sum_{j=1}^{k} \frac{1}{|\mathcal{C}_j|} \sum_{p,r \in \mathcal{C}_j} \langle s_p|s_r\rangle, \tag{10}$$

where $\mathcal{C}_j$ and $|\mathcal{C}_j|$ represent the members of the j-th cluster and the number of members, respectively. The centroid of $\mathcal{C}_j$ is denoted by $|m^{(j)}\rangle$. To get the rightmost expression, we used the definition of the centroid $|m^{(j)}\rangle = \frac{1}{|\mathcal{C}_j|} \sum_{p \in \mathcal{C}_j} |s_p\rangle$.

Since the first term in the rightmost side does not depend on clustering, let us focus on the second term, which will be denoted by E_2. To remove the restricted summation, we introduce an indicator vector $|u^{(j)}\rangle \in \mathcal{H}$, where $\langle s_q|u^{(j)}\rangle = 1/\sqrt{|\mathcal{C}_j|}$ for $s_q \in \mathcal{C}_j$ and 0 otherwise, to have

$$E_2 = -\sum_{j=1}^{k} \sum_{p,r=1}^{n} \langle u^{(j)}|s_p\rangle\langle s_p|s_r\rangle\langle s_r|u^{(j)}\rangle = -\sum_{j=1}^{k} \langle u^{(j)}|\hat{\rho}^2|u^{(j)}\rangle,$$

where we introduced a linear operator $\hat{\rho}$

$$\hat{\rho} \equiv \sum_{p=1}^{n} |s_p\rangle\langle s_p| \tag{11}$$

to get the rightmost expression. Note that, in contrast to Eq. (5), $\hat{\rho}$ is not the identity operator since $|s_p\rangle$s are not orthonormal.

The k-means clustering task has now been reduced to seeking the solution $\{|u^{(j)}\rangle\}$ which minimizes E_2. If we relax the original binary constraint on $|u^{(j)}\rangle$, and instead take

$$\sum_{p=1}^{n} \langle u^{(i)}|s_p\rangle\langle s_p|u^{(j)}\rangle = \langle u^{(i)}|\hat{\rho}|u^{(j)}\rangle = \delta_{i,j} \tag{12}$$

as the new restriction on the optimization problem, the k-means task now amounts to

$$\hat{\rho}|u^{(j)}\rangle = \lambda_j |u^{(j)}\rangle, \tag{13}$$

where λ_j is the eigenvalue corresponding to the eigen vector $|u^{(j)}\rangle$. The site representation allows us to solve this via standard matrix computations, i.e.,

$$\sum_{l'=1}^{w} \langle l|\hat{\rho}|l'\rangle\langle l'|u^{(j)}\rangle = \lambda_j \langle l|u^{(j)}\rangle.$$

Before the relaxation, the indicator vectors satisfied

$$|m^{(j)}\rangle \equiv \frac{1}{|\mathcal{C}_j|}\sum_{p\in\mathcal{C}_j}|s_p\rangle = \frac{1}{\sqrt{|\mathcal{C}_j|}}\sum_{p=1}^{n}|s_p\rangle\langle s_p|u^{(j)}\rangle = \frac{1}{\sqrt{|\mathcal{C}_j|}}\hat{\rho}|u^{(j)}\rangle.$$

After the relaxation, $|u^{(j)}\rangle$ is the eigen vector of $\hat{\rho}$. Thus it follows that *the k-means cluster centers correspond to the eigenstates of $\hat{\rho}$*, or

$$|m^{(j)}\rangle \propto |u^{(j)}\rangle. \tag{14}$$

This remarkable relation was first derived in [6]. As compared to previous work of spectral clustering [13,11,2,3], our main contribution is that we introduced a new formulation which directly seeks the cluster centers, instead of the standard formulation based on membership indicators.

4 Group-Theoretical Properties of $\hat{\boldsymbol{\rho}}$

In this section, we explain the basics of group theory. Specifically, we derive irreducible representations of the translational group, showing an interesting connection between Fourier components and the translational group.

4.1 Translational Group

We have shown the theoretical connection between k-means STSC and spectral STSC. Since spectral STSC is expressed as the eigenvalue equation of $\hat{\rho}$, it is useful to study mathematical properties of $\hat{\rho}$.

Using the expression of Eq. (8), and arranging site indices accordingly, $\hat{\rho}$ can be written as

$$\hat{\rho} \doteq \sum_{l=1}^{n} \hat{\tau}(l)^{\dagger}|\mathit{\Gamma}\rangle\langle\mathit{\Gamma}|\hat{\tau}(l), \tag{15}$$

where we used a shorthand notation instead of using $\hat{\vartheta}(w)$: We define the symbol "$\doteq$" meaning "the left and the right sides have the same matrix elements when represented in $\mathcal{H}$ (not $\mathcal{H}_0$)".

In this expression of $\hat{\rho}$, it seems that a set of the translational operators

$$\mathcal{T}_n \equiv \{\hat{\tau}(0), \hat{\tau}(1), ..., \hat{\tau}(n-1)\}.$$

plays a key role. If we define $\hat{\tau}(n) = \hat{1}$, we see that $\mathcal{T}_n$ is closed in that any product between two of the operators remains in $\mathcal{T}_n$. For example, when $n = 6$ as shown in Fig. 2 (b), the operation of $\hat{\tau}(3)$ followed by $\hat{\tau}(4)$ must coincide with that of $\hat{\tau}(1)$ ($= \hat{\tau}(7)$) by definition. More exactly, for $\forall|a\rangle \in \mathcal{H}_0$, $\hat{\tau}(4)\hat{\tau}(3)|a\rangle = \hat{\tau}(1)|a\rangle$ must hold. In addition, first, for any integers $l, l', l'' \in \{0, 1, ..., n-1\}$,

$$\hat{\tau}(l)\hat{\tau}(l')\hat{\tau}(l'') = \hat{\tau}(l+l')\hat{\tau}(l'') = \hat{\tau}(l)\hat{\tau}(l'+l'')$$

is clearly satisfied. Second, $\mathcal{T}_n$ has the unit element of $\hat{\tau}(0) = \hat{1}$. Third, any of the elements in $\mathcal{T}_n$ has an inverse element. For example, $\hat{\tau}(2)$ is the inverse element of $\hat{\tau}(n-2)$, since $\hat{\tau}(2)\hat{\tau}(n-2) = \hat{\tau}(n-2)\hat{\tau}(2) = \hat{1}$.

These three properties are nothing but the axioms of *group*:

Definition 2 (Group). *A group $\mathcal{G}$ is a set of linear operators such that (1) any of three elements in $\mathcal{G}$ satisfy the associativity relation, (2) $\mathcal{G}$ includes the unit element, and (3) any of the elements in $\mathcal{G}$ has an inverse element in $\mathcal{G}$.*

Thus $\mathcal{T}_n$ forms a group, which called the (one-dimensional) translational group.

A remarkable property of Eq. (15) is translational invariance of the right hand side (r.h.s.). Recall the definition of the invariance Eq. (9) and the PBC. Then, it follows

$$\hat{\tau}(l)^\dagger(\text{r.h.s.})\,\hat{\tau}(l) = \sum_{l'=1}^{n} \hat{\tau}(l+l')^\dagger|\Gamma\rangle\langle\Gamma|\hat{\tau}(l+l') = \sum_{l''=1}^{n} \hat{\tau}(l'')^\dagger|\Gamma\rangle\langle\Gamma|\hat{\tau}(l''), \quad (16)$$

showing the translational invariance of r.h.s. of Eq. (15).

Since the particular form of the r.h.s. in Eq. (15) is a direct consequence of the sliding window technique, we must say that this translational symmetry is just a mathematical artifact introduced by the sliding window technique. In this sense, we call the translational symmetry of the r.h.s. of Eq. (15) the *pseudo* translational symmetry.

4.2 Representation Theory

In Subsection 2.2, we introduced the notion of matrix representation of linear operators in a way specific to $\mathcal{H}_0$. Here we generalize the concept to groups:

Definition 3 (Representation). *Let $\mathcal{D}$ be a set of $d \times d$ matrices, each of which is associated with an element of a group $\mathcal{G}$. The set $\mathcal{D}$ is said a representation of $\mathcal{G}$, if $D(\hat{o}_i)D(\hat{o}_j) = D(\hat{o}_k)$ holds for any $\hat{o}_i, \hat{o}_j, \hat{o}_k \in \mathcal{G}$ such that $\hat{o}_i\hat{o}_j = \hat{o}_k$. Here $D(\hat{o}) \in \mathcal{D}$ is said the representation matrix of $\hat{o} \in \mathcal{G}$.*

In this definition, d is called the dimension of the representation. For example, each element of $\mathcal{T}_n$ can be expressed as an $n \times n$ matrix of $\langle l|\hat{\rho}|l'\rangle$, based on the basis of $\mathcal{H}_0$. The matrix representation of $\hat{\tau}(1)$ has ones for $l = l' + 1$, zeros otherwise.

As indicated by this example, a representation can be constructed by looking at how a group element operates on the basis of a vector space. Thus it is natural to introduce the notion of representation spaces as

Definition 4 (Representation space). *A vector space is said invariant w.r.t. a group $\mathcal{G}$, if the space remains in the same space after the operation of $\forall \hat{o} \in \mathcal{G}$. A subspace is said a representation space of $\mathcal{G}$, if it is invariant w.r.t. $\mathcal{G}$.*

In the above example, $\mathcal{H}_0$ is a representation space of $\mathcal{T}_n$. Clearly, a representation space is not unique. For example, a different representation will be obtained if we use $2n$-dimensional space defined using a one-dimensional lattice having $2n$ sites. As expected from this example, there is generally no upper bound on the dimension of representation spaces.

One interesting question here is whether or not there exists a lower bound on the dimension of representation spaces. The answer is yes. It is known that, for a given group, there exist a certain number of "minimal" representation spaces, which are called the irreducible representation space of the group. Putting formally,

Definition 5 (Irreducible representation space). *If a representation space does not include any subspaces that are invariant w.r.t. $\mathcal{G}$, it is called an irreducible representation space.*

For example, while we have used the n-dimensional space to represent $\mathcal{T}_n$ so far, it can be shown that a vector $|f_0^n\rangle \equiv \frac{1}{\sqrt{n}} \sum_{l=1}^{n} |l\rangle$ spans a one-dimensional irreducible representation space of $\mathcal{T}_n$. In fact, it is easy to verify $\forall l$, $\hat{\tau}(l)|f_0^n\rangle = |f_0^n\rangle$. Clearly, this space should be irreducible because it is one-dimensional. Since all the representation matrices are ones (1×1 identity matrix), this irreducible representation is also called the identity representation.

If representations of a group are turned to be irreducible, it is known that a strong theorem, which is a fundamental theorem in group theory and is also known as Schur's first lemma, holds. While Schur's first lemma is almost always expressed as mysterious-looking relations between representation matrices (for a proof, see, e.g. [7]), it essentially states the orthogonality of different irreducible representation spaces:

Theorem 1 (Schur). *For a given group $\mathcal{G}$, let $|q, m\rangle$ be a base in an irreducible representation space labeled by q, where m specifies a dimension of the representation space. If $\hat{o} \in \mathcal{L}$ is invariant w.r.t. any element of $\mathcal{G}$, it follows $\langle q, m|\hat{o}|q', m'\rangle \propto \delta_{q,q'}$, i.e. the matrix element is zero if the irreducible representations are different.*

Another fundamental theorem in group theory is one called Schur's second lemma. Combining the first and the second lemmas, one can prove a stronger relation, which is also known as the selection rule in quantum physics (for proof, see [7]), as

Theorem 2 (Selection rule). *Under the same setting as Theorem 1, it follows*

$$\langle q, m|\hat{o}|q', m'\rangle \propto \delta_{q,q'}\delta_{m,m'}$$

for a unitary operator $\hat{o}$ that is invariant w.r.t. any element of $\mathcal{G}$.

4.3 Irreducible Representations of Translational Group

Since the identity operator is invariant w.r.t. any linear operator, we see from Theorem 2 that the bases of irreducible representation spaces are orthogonal to each other:

$$\langle q, m|q', m'\rangle \propto \delta_{q,q'}\delta_{m,m'}.$$

This orthogonal relation reminds us of subspace learning methods such as PCA, where mutually orthogonal directions are explored in terms of maximum variance. It is tempting to associate the irreducible representation spaces with extracted patterns. One important implication here is that group theory serves as a new tool for pattern learning, where each irreducible representation is thought of as an extracted pattern.

Let us find irreducible representations of $\mathcal{T}_n$. First, since $\hat{\tau}(l) \in \mathcal{T}_n$ is unitary, its matrix representation must be a unitary matrix. Second, irreducible representations of $\mathcal{T}_n$ must be all one-dimensional. To see this, suppose that an irreducible representation space is d_q-dimensional. By definition, for $\forall\hat{\tau}(l) \in \mathcal{T}_n$,

$$\hat{\tau}(l)|q, m\rangle = \sum_{m'=1}^{d_q} |q, m'\rangle\langle q, m'|\hat{\tau}(l)|q, m\rangle.$$

Since $\hat{\tau}(l)$ itself is translationally invariant, it follows from Theorem 2 that

$$\hat{\tau}(l)|q, m\rangle = |q, m\rangle\langle q, m|\hat{\tau}(l)|q, m\rangle.$$

This means that the original d_q-dimensional representation spaces has an invariant subspace w.r.t. $\mathcal{T}_n$, which contradicts to the assumption. Therefore we conclude that each of the irreducible representations of $\mathcal{T}_n$ is one-dimensional. This means that *the irreducible representations must be the eigenvectors*. Because of the unitary constraint, the eigenvalue must have the form of $\mathrm{e}^{\mathrm{i}c}$ for a $c \in \mathbb{R}$. Recall the fact that $\hat{\tau}(l) = [\hat{\tau}(1)]^l$ and $\hat{\tau}(n) = \hat{1}$. Thus it follows that

$$\hat{\tau}(l)|q\rangle = [\mathrm{e}^{\mathrm{i}f_q^n}]^l|q\rangle, \tag{17}$$

where the exponent is given by

$$f_q^n = 2\pi q/n \quad (q \text{ is an integer}).$$

Here the dependence of f on q and n is explicitly expressed as sub- and superscripts, respectively. We dropped the unnecessary index m.

As indicated by the exponential factor in Eq. (17), the eigen spaces are given by discrete Fourier transformation (DFT) in $\mathcal{H}_0$:

$$|f_q^n\rangle = \frac{1}{\sqrt{n}}\sum_{l=1}^{n} \mathrm{e}^{\mathrm{i}f_q(l-l_0)}|l\rangle, \tag{18}$$

where l_0 is a real constant, and the subscript q runs over $\mathcal{D}_f^n = \{-\frac{n-1}{2}, ..., 0, 1, .., \frac{n-1}{2}\}$ when n is odd, and $\{-\frac{n}{2} + 1, ..., 0, 1, .., \frac{n}{2}\}$ when n is even. It is easy to verify Eq. (17):

$$\hat{\tau}(l)|f_q^n\rangle = \frac{1}{\sqrt{n}}\sum_{l'=1}^{n} \mathrm{e}^{\mathrm{i}f_q^n(l'-l_0)}|l'+l\rangle = \mathrm{e}^{-\mathrm{i}lf_q^n}|f_q^n\rangle, \tag{19}$$

where we used $\mathrm{e}^{\mathrm{i}f_q^w n} = 1$.

5 Solution to Subsequence Time-Series Clustering

5.1 The $w = n$ Case

When $w = n$, i.e. the length of the subsequences is the same as the whole time series (see Fig. 2 (c)), the operator $\hat{\rho}$ has exact translational invariance. Thus, from the selection rule, it follows

$$\langle f_q^n|\hat{\rho}|f_{q'}^n\rangle \propto \delta_{q,q'}. \tag{20}$$

This means that the matrix representation of $\hat{\rho}$ is diagonal in the space spanned by $\{|f_q^n\rangle|\ q \in \mathcal{D}_f^n\}$. Thus solving the eigen equation Eq. (13) for $|u^{(j)}\rangle$ is trivial, and the eigen vectors are nothing but $|f_q^n\rangle$s.

Using Eqs. (15) and (19), we can calculate the diagonal element as

$$\langle f_q^n|\hat{\rho}|f_q^n\rangle = \sum_{l=1}^{n}\langle f_q^n|\hat{\tau}(l)^\dagger|\Gamma\rangle\langle\Gamma|\hat{\tau}(l)|f_q^n\rangle = \sum_{l=1}^{n}\langle f_q^n|\Gamma\rangle\langle\Gamma|f_q^n\rangle = n|\langle f_q^n|\Gamma\rangle|^2. \tag{21}$$

Thus the i-th top eigen vector occurs at a Fourier component having the i-th largest *power* $|\langle f_q^n|\Gamma\rangle|^2$. For any real time series, $|f_q^n\rangle$ has the same power as $|f_{-q}^n\rangle$, and the eigenvector takes the form of $|f_q^n\rangle + |f_{-q}^n\rangle$. Thus the resulting cluster centers of spectral STSC are pure sinusoids. Note that each of the eigenvectors must be a pure sinusoid in this case, irrespective of the power spectrum.

To validate this result, we performed k-means and spectral STSC for a white noise data having $n = 6000$ (the data used is Normal data in the Synthetic Control Chart data [8]). We did 100 random restarts and chose the best one in the k-means calculation. As shown in Fig. 3 (a), we have the largest power at $|f_q^n| = 0.358$ in (marked by the triangles). Thus the wavelength must be $2\pi/|f_q^n| = 17.6$, which is completely consistent with Fig. 3 (c). In addition, we see that the eigen vector is a good estimator of the k-means cluster center by comparing Figs. 3 (b) and (c).

Note that the sinusoids are obtained as irreducible representations of $\mathcal{T}_n$. Thus we conclude that the sinusoid effect is a direct consequence of the pseudo-translational symmetry introduced by the sliding window technique.

5.2 The $w < n$ Case

Let us consider the general case of $w < n$. In this case, $\hat{\rho}$ does not have exact translational invariance, so we need to introduce DFT in $\mathcal{H}$ rather than $\mathcal{H}_0$:

$$|f_q^w\rangle = \frac{1}{\sqrt{w}}\sum_{l=1}^{w} \mathrm{e}^{\mathrm{i}f_q^w(l-l_0)}|l\rangle. \tag{22}$$

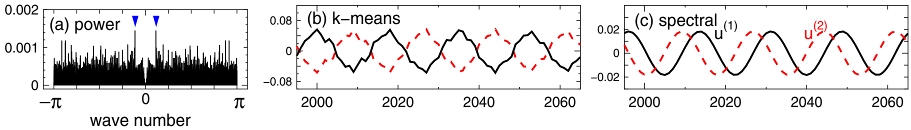


Fig. 3. The results for a white noise data with $w = n = 6000$. (a) The power spectrum. (b) A segment of the k-means cluster centers ($k = 2$), and (c) a segment of the top two eigen vectors of $\hat{\rho}$.

It is straightforward to show $\langle f_q^w | f_{q'}^w \rangle = \delta_{q',q}$, and thus $\{|f_q^w\rangle |\ q \in \mathcal{D}_f^w\}$ forms the complete set in $\mathcal{H}$.

Consider a vector $\hat{\tau}(1)|f_q^w\rangle$. Using the fact $\mathrm{e}^{-\mathrm{i}f_q^w w} = 1$, it is easy to show

$$\hat{\tau}(1)|f_q^w\rangle = \mathrm{e}^{-\mathrm{i}f_q^w}|f_q^w\rangle + \frac{1}{\sqrt{w}}\mathrm{e}^{-\mathrm{i}f_q^w l_0}|B\rangle,$$

where $|B\rangle \equiv |w+1\rangle - |1\rangle$. By applying $\hat{\tau}(1)$ sequentially, we have

$$\hat{\tau}(l)|f_q^w\rangle = \mathrm{e}^{-\mathrm{i}f_q^w l}\left[|f_q\rangle + \frac{1}{\sqrt{w}}\sum_{l'=1}^{l}\mathrm{e}^{\mathrm{i}f_q^w(l'-l_0)}\hat{\tau}(l'-1)|B\rangle\right].$$

While we can omit the second term inside the bracket when both $w = n$ and the PBC hold, that is not the case here. Using this formula, we can calculate the matrix elements of $\hat{\rho}$ in the Fourier representation. The result will be[1]

$$\langle f_q^w|\hat{\rho}|f_{q'}^w\rangle \approx n|\langle f_q^w|\Gamma\rangle|^2\delta_{q,q'} + \sum_{l=1}^{n}\mathrm{e}^{\mathrm{i}l\Delta_{q'q}}J_l^w(q,q'). \tag{23}$$

It is straightforward to get the exact expression of $J_l^w(q,q')$ although we do not show it here. In the above equation, the first term is the same as Eq. (21), and the second term is a perturbation term giving off-diagonal elements. However, under normal conditions, we can assume that the first term is the leading term, since $n \gg 1$ and phase cancellations are unavoidable in the second term. In particular, if the power spectrum of a time-series data set has a single-peaked structure at a certain $|f_q^w|$, the top eigen vector will be well approximated by the $f_{|q|}^w$, irrespective of the details of the spectrum.

On the other hand, if the power spectrum has a multiple peak structure, the second term can play some role, and the eigenvectors will be mixtures of many f_q^ws. In that case, the cluster centers will be far from pure sinusoids. In fact, for the white noise data, where the power spectrum is almost flat, no sinusoids are observed as long as $w \ll n$, as expected.

To conclude, whether the sinusoid effect is observed or not depends on the relative magnitude between the first and the second terms. If the first term is dominant, the sinusoid effect will be observed. As shown, the pseudo-translational

[1] For $w \ll n$, one may approximate as $\sum_{l=1}^{n}\mathrm{e}^{\mathrm{i}l\Delta_{q',q}} \approx n\delta_{q,q'}$.

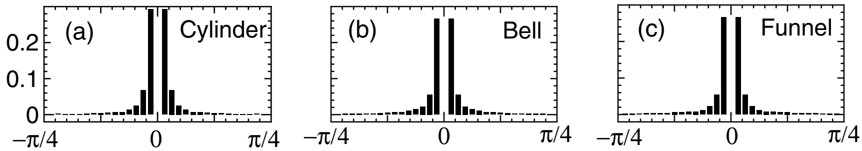


Fig. 4. Averaged power spectra of each instance of the CBF data. The horizontal axis represents f_q^w within $[-\frac{\pi}{4}, \frac{\pi}{4}]$ out of $(-\pi, \pi]$.

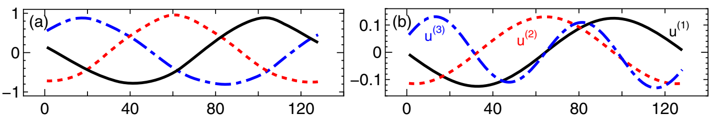


Fig. 5. (a) The k-means cluster centers ($k = 3, w = 128$). (b) The top three eigen vectors of the spectral STSC ($w = 128$).

symmetry is essential to obtain the first term. Therefore, even in this general case, we may claim that the sinusoid effect is a direct consequence of the pseudo-translational symmetry introduced by the sliding window technique.

To validate the theoretical result of Eq. (23), we performed experiments using Cylinder, Bell, and Funnel (CBF) data [8]. The CBF data includes three types of patterns literally having Cylinder, Bell, and Funnel shapes. We randomly generated 30 instances for each type with a fixed length of 128 ($= w$) using Matlab code provided by [8]. We concatenated the instances in order after standardizing each one (zero mean and unit variance). An example segment of the concatenated data was shown in Fig. 1 (a). Again, we did 100 random restarts and chose the best one in the k-means calculation.

Figures 4 (a)-(c) show the power spectra of each instance as a function of f_q^w. To handle the variation of the instances, we simply averaged the resultant spectra for all instances. We see that the most of the weight is concentrated on the $|q| = 1$ component (i.e. the wavelength of w) in all of the cases. The f_0 component is naturally missing because of the standardization.

The results of k-means and spectral STSC are shown in Fig. 5, where the sinusoid effect is clearly observed. The sinusoid of wavelength of w can be understood from the dominant $|q| = 1$ weight in Fig. 4 (a)-(c). Since the single-peaked structure in the power spectrum is naturally expected whenever there are no particular periodicities within the scale of window size, the sinusoid of wavelength w should be observed in a wide variety of the input data. This explains why the sinusoid effect is ubiquitous.

Due to the orthogonality condition, we see that the third singular vector necessarily has a wavelength of about $w/2$ in Fig. 5 (b). This is an example of the difference between the two formulations in how the calculated cluster centers interact with each other. Apart from this, our formulation is completely consistent with the results.

6 Concluding Remarks

We have performed a group-theoretical analysis of the sinusoid effect in STSC. Based on a spectral formulation of STSC, we showed that the sinusoid effect is a group-theoretical consequence of the pseudo-translational symmetry introduced by the sliding window technique.

In Section 4, we claimed that finding irreducible representations can be viewed as pattern learning if the problem under consideration has a certain symmetry related to a group. In this sense, the sinusoid effect can be thought of as a result of (unintentional) pattern learning from time-series subsequences. In another paper [5], we introduced idea that certain correlation patterns are extracted as irreducible representation spaces of a point group. In this way, designing machine learning algorithms so that irreducible representations of a group can be effectively extracted would be an interesting future direction of research.

References

1. G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth. Rule discovery from time series. In *Proc. ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining*, 1998.
2. I. S. Dhillon, Y. Guan, and B. Kulis. Kernel k-means, spectral clustering and normalized cuts. In *Proc. ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining*, pages 551–556, 2004.
3. C. Ding and X. He. K-means clustering via principal component analysis. In *Proc. Intl. Conf. Machine Learning*, pages 225–232, 2004.
4. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd ed.)*. Wiley Interscience, 2000.
5. T. Idé. Pairwise symmetry decomposition method for generalized covariance analysis. In *Proc. IEEE Intl. Conf. Data Mining*, pages 657–660, 2005.
6. T. Idé. Why does subsequence time-series clustering produce sine waves? In *Proc. European Conf. Principles and Practice of Knowledge Discovery in Databases, Lecture Notes in Artificial Intelligence*, volume 4213, pages 311–322, 2006.
7. T. Inui, Y. Tanabe, and Y. Onodera. *Group Theory and Its Applications in Physics, 2nd ed.* Springer-Verlag Telos, 1996.
8. E. Keogh and T. Folias. The UCR time series data mining archive [`http://www. cs.ucr.edu/ eamonn/TSDMA/index.html`]. 2002.
9. E. Keogh, J. Lin, and W. Truppel. Clustering of time series subsequences is meaningless: Implications for previous and future research. In *Proc. IEEE Intl. Conf. Data Mining*, pages 115–122. IEEE, 2003.
10. R. Kondor and T. Jebara. A kernel between sets of vectors. In *Proc. the 20th Intl. Conf. Machine Learning*, pages 361–368, 2003.
11. A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems, 14*, pages 849–856, 2001.
12. J. J. Sakurai. *Modern Quantum Mechanics*. Addison Wesley, 2nd Ed., 1994.
13. H. Zha, C. Ding, M. Gu, X. He, and H. Simon. Spectral relaxation for k-means clustering. In *Advances in Neural Information Processing Systems, 14*, pages 1057–1064, 2001.

Visualization of Contents Archive by Contour Map Representation

Hidekazu Kubota[1,2], Toyoaki Nishida[2], and Yasuyuki Sumi[2]

[1] Japan Society for the Promotion of Science Research Fellow
[2] Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Abstract. This article describes a model for sustainable contents management, its visualization algorithms, and the implemented system, called sustainable knowledge globe (SKG). The focal point of our study is visualization using contour maps. The graphical representation of tree-structured contents increases in complexity with the number of contents. The contour map representations can briefly depict the arrangement and structure of contents in an archive. Three contour map representations are proposed assuming the importance of the arrangement design. Nesting, dendroidal, and island-like contours are amplified from the viewpoint of preservation of the structures and arrangements, in addition to supporting shape and texture design. The comparison and applications of the three algorithms are discussed.

1 Introduction

Sustainable management of contents is indispensable for constructive knowledge evolution. Companies and NGOs store their history, products, and daily discussions in contents archives as a basis of their group activities. Personal archives are also being developed to manage personal memories and experiences [1] [2]. Building a huge archive for long term use is difficult without computational support. It has been recognized that computer visualization is effective for managing huge contents. The focal point of our study is the visualization of an arrangement design that supports the sustainable management of huge contents.

Designing the arrangement of contents is the primary task in intellectual presentation. Exhibition plans in galleries and museums generally appear to have been investigated. The elaborate arrangement of conceptual contents has also been widely discussed in many computational supports for thinking [3] [4] based on the KJ method [5] and a bookmark management system [6]. Nishida has proposed a scenic metaphor called landscape [7]. A landscape is a developing scene where diverse contents are spatially arranged. People can easily remember the location of contents in landscape just as they remember their hometown.

The gradual development of landscape is supposed to bring about sustainable memory. Gradual changes in arrangement are easier for people to remember than sudden changes that disrupt people's mental maps. This gradual development

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 19–32, 2007.

of landscape is schematically developed into a spatiotemporal memory model in Section 2. The topological model of contents arrangement in the landscape is also proposed in Section 2. This model is called the panoramic contents presentation. The spatio-temporal memory model and panoramic contents presentation are implemented into a system called sustainable knowledge globe (SKG). SKG is a contents archive system that uses a sphere model. Its implementation is described in Section 3.

We propose the algorithms for generating contour maps in Section 4 to visualize huge tree data structures that are entangled with each other on the SKG. The characteristic of our algorithms is to support arrangement design by preserving structures and arrangements in addition to providing support from the viewpoint of shape and texture design. The comparison of the algorithms is discussed in Section 5. The conclusion and future works are described in Section 6.

2 Knowledge Landscape

Huge contents consist of documents, proceedings, e-mails, multimedia files, and their synthesized forms. The difficulty in managing such contents lies in their size, diversity, and situational use. People need to explore several past materials according to a current context, and then append new contents to them. It is difficult for people to manage old and new contexts in huge contents if they lack a good overview.

Knowledge landscape is a scenic representation of huge contents that are organized spatially and temporally. An early concept of the landscape illustrated by Nishida [7] is an evolving memory space that expands human biological memory. The landscape enables the user to visually grasp the global nature of knowledge, explore the information space, and accommodate new information at an appropriate place.

Knowledge landscape is generally used as a concept for managing conversation quanta [8]. A conversation quantum is a reusable knowledge material that is an individually functional, interactive, and synthesizable conversation block like a LEGO ® block. The conversation quantum consists of the content that is uttered (conversation movie, dictation text, and contextual information) and its external appearance [9]. A set of conversation quanta has both conversational and visual characteristics. People can retrieve knowledge represented by conversation quanta by using conversational techniques. People can also construct contents by arranging and connecting conversation quanta visually. This paper does not focus on the conversational characteristics of conversation quanta but on their visual characteristics. The visual characteristics of conversation quanta can be approximated to general digital contents such as photos.

We schematically develop the evolving landscape in the spatio-temporal memory model (Fig. 1). Contents are visually placed anywhere the user likes them to be on the landscape (t = 1). Here, we assume that the arbitrary arrangement of contents results in several spatial clues to remind the user of the located contents. The synthesized content is constructed continuously (from t = 0 to t = n) on the landscape by connecting, rearranging, and developing contents.

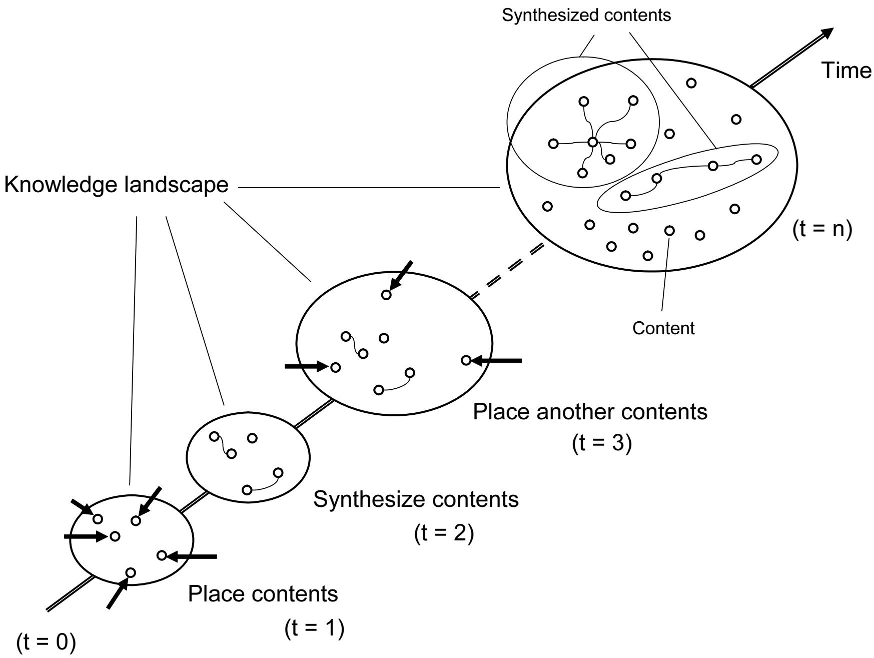


Fig. 1. Spatiotemporal memory model

A panoramic content presentation is a topological model of contents arrangement in the knowledge landscape. The panorama includes tree structured content cards and narrative paths, each of which consists of a sequence of viewpoints (Fig. 2). A content is represented by a content card in the knowledge landscape. The content card is an extended form of a knowledge card [10] that is a reusable representation of content in conversational and editorial situations. The knowledge card allows only an image file or a movie clip, whereas a content card allows any file to make it more expressive.

A content card consists of three parts: an embedded file, a title of a card, and the annotation of the card. A content card is described by XML. The <card> element represents a unit of a content card. The <card> element contains three child elements: <title>, <url>, and <annotation>. The <title> element contains the title text of the card, the <url> element contains the URL of an embedded file (e.g., a document, an image, a movie clip, or a slide), and the <annotation> element contains the annotation text of the card. Each content card is bound to one or more viewpoints that are also described in the same form. The viewpoint represents a scope of view in the panorama that is recorded as information about the eye distance from the globe and the center of gaze. The viewpoint is a very important factor that includes the context and background of the card. The viewpoint (a) in Fig. 2 (2) gives information about the surroundings to the card C. The viewpoint (b) gives detailed view to card C2.

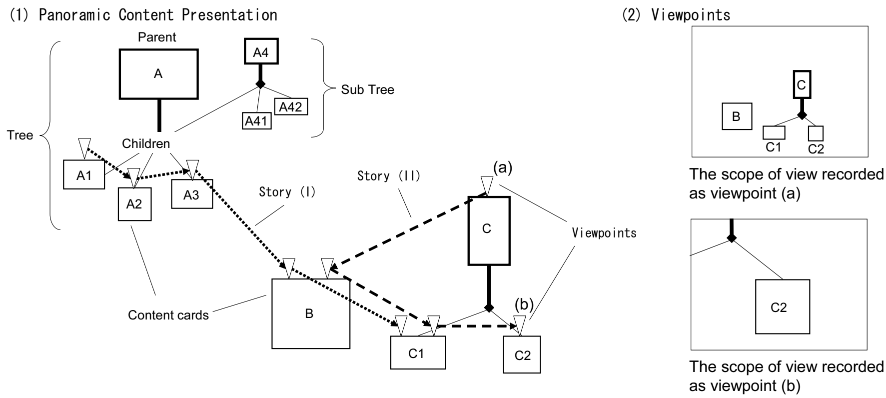


Fig. 2. Panoramic content presentation

A tree structure is a standard way of representing categories. By categorizing the content cards, a set of cards can be easily arranged and retrieved. Cards A1, A2, A3, and A4 in Fig. 2 are the children of the card A, and A4 also represents a subcategory of card A. The relationship between a parent and a child is drawn by an edge. The knowledge landscape allows multiple trees on a virtual globe because the arrangement is limited if it allows only one tree. An independent card that does not belong to a tree is also allowed for the same reason.

Story representation is important to describe a sequence of contents. Here, a story is represented by a sequence of viewpoints that shows a panoramic story line of the contents. The relationship between a previous viewpoint and the next viewpoint is drawn by an arrow that cut across trees. The sequence from A1 to C1 is a story (I) in Fig. 2. The sequence from C to C2 is another story (II). A content card is allowed to belong to multiple stories. The story is so orderly and cross-boundary that people can easily create cross-contextual contents.

3 Sustainable Knowledge Globe

The knowledge landscape is implemented into the sustainable knowledge globe (SKG), which is a sustainable content builder [11]. SKG works on the Windows platform with .NET and Managed DirectX. The SKG landscape is a virtual sphere similar to a terrestrial globe with latitude and longitude lines, landmarks for the north and south poles, and the equator (Fig. 3). A screenshot of the implemented knowledge landscape is shown in Fig. 4. Just after starting the SKG, the user-defined home position is shown, then a user can explore the landscape by using a wheel mouse. The sphere has three degrees of freedom (latitudinal, longitudinal, and depth directions) because a standard mouse device appears to be unsuitable to operate a sphere with many degrees of freedom. A user can also change the view angles of a sphere. The normal view is an oblique view that has depth, while a bird's eye view is an overhead view that has less occlusion problems.

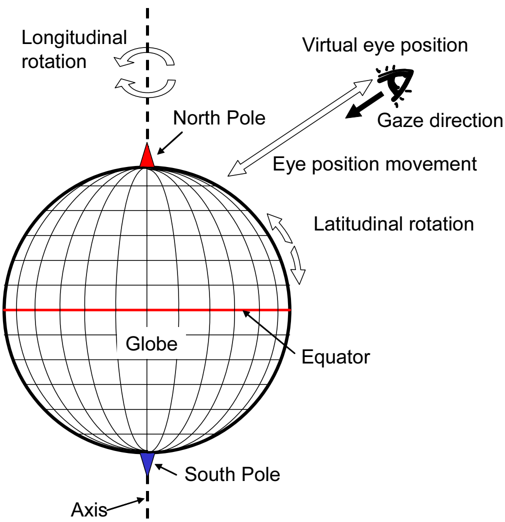


Fig. 3. Zoomable globe

The reason why we adopt a globe shape is its good arbitrariness for contents arrangement. There could be many types of topologies in the space: a finite plane, an infinite plane, 2-dimensional torus, and so on. A finite plane appears to be unsuitable for a user to expand the contents on the edges of the plane. An infinite plane does not pose this problem; however, it is difficult for people to grasp infinite space. The 2-dimensional torus also has no edge; however, such a topology may be unfamiliar for people. A globe shape would be the appropriate solution where people can arrange contents more freely than in a finite plane because a globe has no edge. Moreover, a sphere is more familiar shape than a 2-dimensional torus since it is similar to a terrestrial globe.

A zooming graphical interface [12] is effective to manage large content on a finite screen size. It enables a user to view the entire content by changing the scale, or to observe specific content by changing the focus. In general, zooming interfaces can be classified as linear and nonlinear zooming interfaces. Linear zooming [12] magnifies and shrinks the entire information in a manner similar to a multi-scale map. Nonlinear zooming [13] [14] distorts an arrangement of information to focus on specific information. We adopt the linear zooming interface

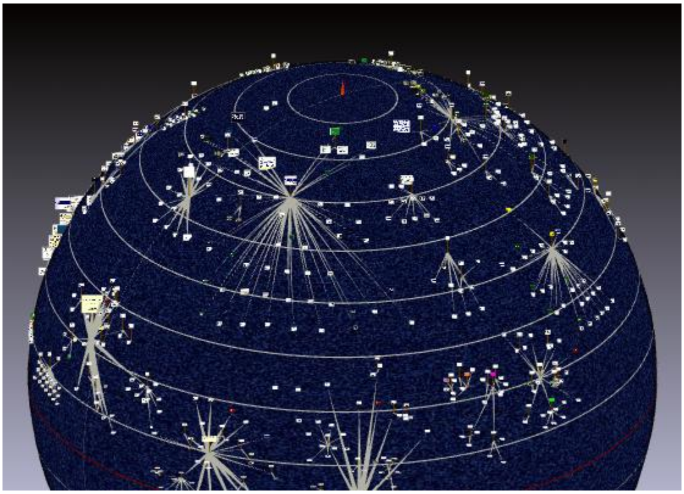


Fig. 4. A screenshot of SKG and tree structured contents

to manage content because we aim to capitalize on the memory of the contents arrangement, which should not be distorted.

What we can manage as a content is any electronic file with natural language annotation. Files are imported as content cards anywhere onto the SKG by dragging and dropping from the PC desktop or downloading from a network server for SKG. Cards are also created by an inline editor on SKG. A card on the SKG is attached with a thumbnail image. An annotation is automatically attached with a quantum if the source content has its annotation (e.g., speaker's note on a PowerPoint ® slide). An annotation is also attached later.

Most parts of contents management on the SKG are now carried out by humans. The SKG can work as a whiteboard or a set of transcription notes in meeting situations. Utterances are written down on the SKG by participants in the conversation. Related handouts, photos, or web pages are also imported onto the SKG. They are placed and connected on the SKG along with the discussion. SKG can be also used for personal contents management. Research slides, movies, leisure photos, bookmarks, and memos have been arranged on the SKG [11].

4 Contour Map Representation

People can categorize contents cards by using tree data structures. Multiple trees are allowed to be added on a surface of SKG. Branches of a tree spread to all directions on the surface. An example of tree data structures added onto the SKG is shown in Fig. 4. The graphical representation of the spatially arranged tree increases in complexity with the number of nodes. Tree nodes get entangled in a web of trees where several branches cross each other. Previous works on arrangement design do not support the management of 1000 or more contents. Our approach is the contour map representation that simplifies huge complex tree structures by using chorographic metaphor.

There are the following four requirements for simplification that emphasize the tree data structure and the arrangement of its nodes.

a) preservation of structures
b) preservation of arrangements
c) design of shape and texture
d) space efficiency

Simplification means the reduction of some information except the focal structure and arrangement. The design of the shape and texture is another point that reinforces human memory for arrangement. Space efficiency must be high for utilizing the limited display space.

Contour map representation is inspired by the expressive power of topographic maps. Contour lines show the arrangement of peaks, and besides, the enclosures can be regarded as hierarchical structures. For people, a contour map has a familiar design that can be memorized. The space efficiency of contour map representation is relatively high because the shape of the contour lines is freer than the shape of primitives like circles or rectangles.

We propose three kinds of contour map representations in this section. The nesting contour visualizes a parent and its children by using enclosures. The dendroidal contour visualizes the shape of trees. The island-like contour visualizes the arrangement of contents. Each contour has different requirements with regard to (a),(b),(c), and (d) as described above.

4.1 Nesting Contour

The nesting contour simplifies hierarchical structures. A contour that encloses a parent and its children shows the region of a subtree clearly. Here, a contour encloses other contours recursively according to a tree structure. One tree structure is represented by one mountainous region that consists of layered contours. We adopt a contour that traces an outline of a tree because primitive shapes of enclosures, such as circles [15], rectangles [12] [16], and cubes [17] lack spatial clues that help people to remember their contents. The primitive shapes are also so inflexible that the enclosures of trees tend to get entangled. A model of the nesting contour and a 3D view on SKG are shown in Fig. 5. The algorithm for drawing the nesting contour of a depth N tree can be described with the following pseudocode:

```
1: Initialize the variable D = 1.
2: Let S be the set of all nodes from depth D to N.
3: Outline the edges of every subtree among S by a contour line of
 width W. If there are no edges in a subtree (that is, D equals N),
 outline each node point among S by a circle of radius W.
 W is inversely proportional to D.
```

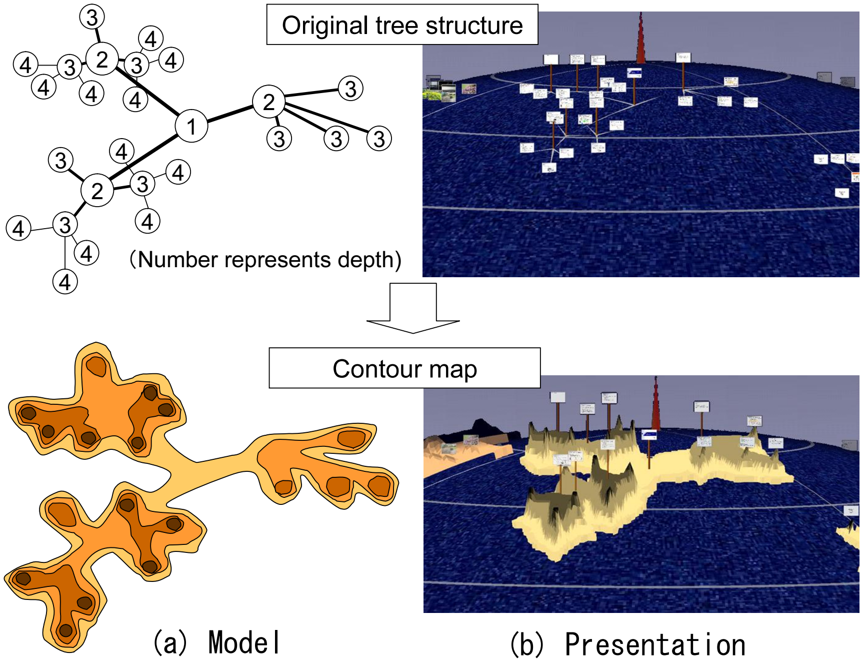


Fig. 5. Nesting contour

```
4: Fill color into the contour.
   The brightness of the color is inversely proportional to D.
5: If D equals N, stop. Else, increase D by 1 and go to step 2.
```

The contour is calculated in real time whenever content cards are moved. Children are higher in altitude and darker than their parent. The location of a leaf node is represented by a circle around the node. Meanwhile, the location of an inner node (including the root node) is ambiguous as a result of the above-mentioned algorithm. We show the location of the inner nodes by fixing a pole in the ground.

We have also worked out a method for emphasizing the leaf nodes in step 3. A contour line near a leaf node follows a randomized zigzag course. The zigzag is suitable for characterizing a leaf node. Moreover, it is a natural expression of the landscape.

4.2 Dendroidal Contour

The spread and shape of branches are good clues to grasp the trends in the growth of tree data structures. A leafy subtree shows that the subcategory represented by the subtree is detailed. A sparse subtree shows that there is room for future development.

The dendroidal contour visualizes the spread and shape of branches from directly overhead. A model of the dendroidal contour and the 3D view on SKG are shown in Fig. 6. The algorithm for drawing a dendroidal contour of a depth N tree can be described with the following pseudocode:

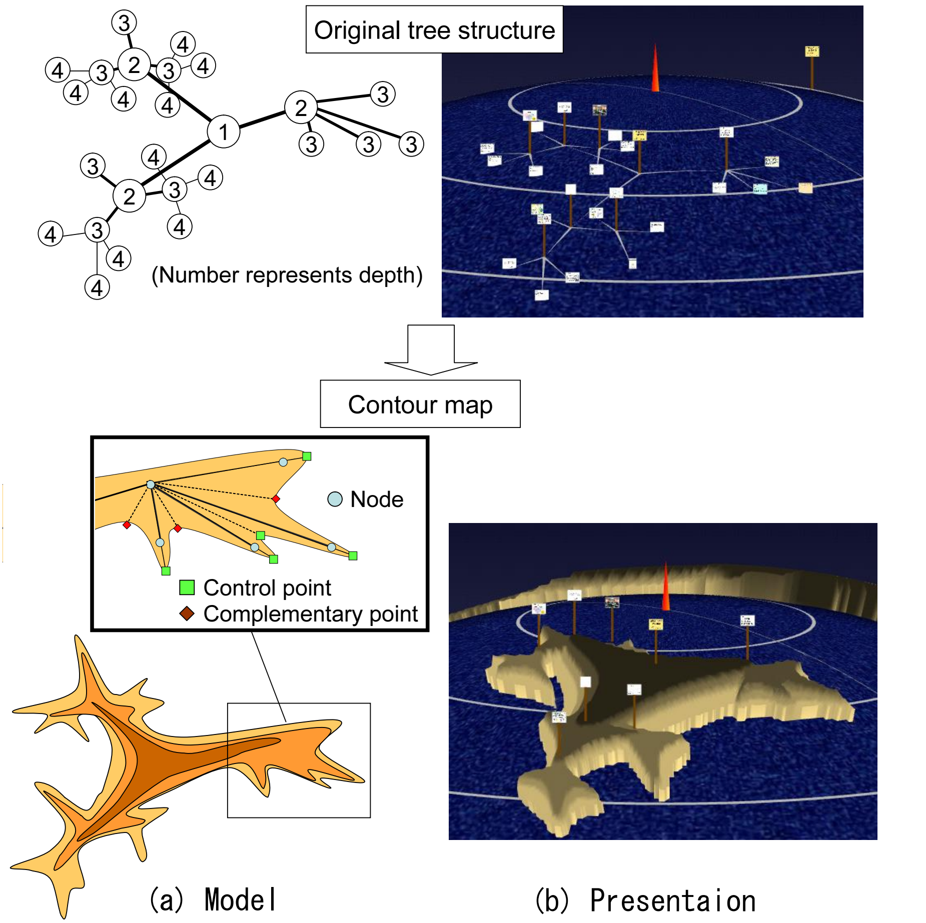


Fig. 6. Dendroidal contour

```
1: Initialize the variable D = N.
2: Let S be the set of all nodes from depth 1 to D.
3: Let T be a tree structured by S.
4: Outline edges of T by a cardinal spline curve. The control point
   of the curve is placed on the extrapolation of the branch that
   connects the leaf node and its parent node in T. Let this parent
   node be named the parent of the control point. The length of the
   extension is W. W is proportional to D.
```

```
 Each of neighbor control points has one complementary point that
 is placed in the middle of these control points. The distance to
 the parent from the complementary point is set shorter than the
 distance from these control points.
5: The overwrapped contour lines are merged.
6: Fill color into the contour. The brightness of the color is
 proportional to D.
7: If D equals 2, stop. Else, decrease D by 1 and go to step 2.
```

The altitude of a root node is the highest while that of the leaf nodes on the periphery of a dendroidal shape is the lowest. The trends in growth are represented by ridges. Leaves can be found at the end of the ridgeline. The arrangement of the nodes is not perfectly clear because some nodes inside the periphery are buried under higher contours. We show the location of the buried nodes by fixing a pole in the ground.

4.3 Island-Like Contour

The island-like contour utilizes an algorithm called blob or metaball [18] that draws organic shapes according to the arrangement and connection of nodes by using a simple density function. Blob calculates a density map by summing up multiple density circles each of which are spread out from the center of the node. In our approach, each node has a density function D(r) described in equation (1).

$$D(r) = \begin{cases} (1 - \frac{r}{R})^4 & (r \leq R) \\ 0 & (r \geq R) \end{cases} \tag{1}$$

r is the distance from the center point of a node to the coordinates of a certain point. R is the effective radius of a density circle. R is inversely proportional to the depth of each node. The decreasing function of r is adopted as D(r) in general. Here, we decide D(r) empirically.

The algorithm for drawing an island-like contour of N nodes tree can be described with the following pseudocode:

```
1: Number nodes from 1 to N.
2: Let P be a point in the coordinate plane.
3: Initialize the variable k = 1.
4: Calculate the density value Dk using D(r). Here, r is the
 distance from the center point of node k to the coordinates of P.
5: If k equals N, go to step 6. Else, increase k by 1 and go to
 step 4.
6: Let Dmax be the maximum value in the set {D1, D2, ..., Dk}.
7: Set the brightness of the color at P to the value in proportion
 to Dmax.
8: Repeat from step 2 to 7 for all the points in the coordinate
 plane.
```

The brightness map is drawn as a result of the abovementioned algorithm. The countour is illustrated by using the posterization technique that quantizes a continuous brightness map into levels that is the same as the depth levels of the tree.

A model of the island contour is shown in Fig. 7. This contour map resembling a coral reef clearly illustrates the location of the nodes.

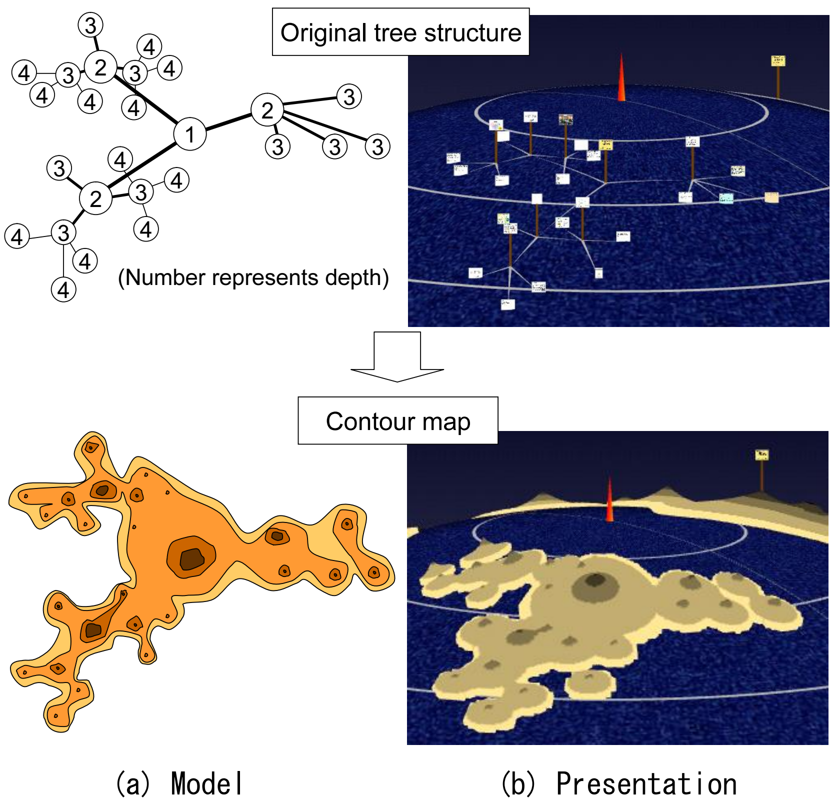


Fig. 7. Island-like contour

5 Discussion

The proposed three contour algorithms are implemented into SKG. The contour map is transformed into 3D terrain surfaces for people to grasp the landscape intuitively. Here, contents cards are placed on the slope. A contour map is always visible regardless of the distance of the eye from the globe. A contour map can show a global shape of contents even if a detailed arrangement of the contents is not shown. An example of the SKG using a nesting map is shown in Fig. 8. Only the contours are visible in the macro scale (Fig. 8(a)). Cards gradually

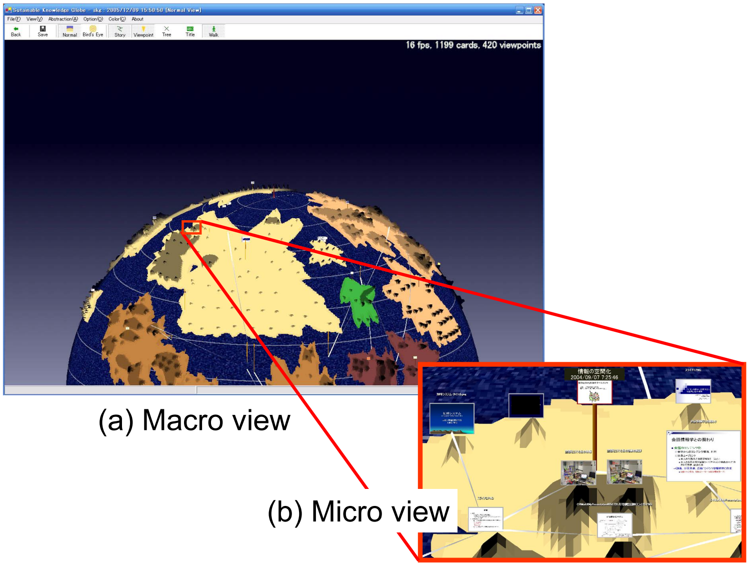


Fig. 8. Overview and detailed view of nesting contour

become visible as the scale becomes larger (Fig. 8(b)). Each tree has a color that illustrates the category and reinforces the design.

The advantage discussed by using the implemented system is shown in Table 1. The hierarchy of the tree structure is clear in the nesting contour. The hierarchy in the dendroidal contour is partially vague because some nodes are buried under higher contours. The island contour has no clear drawing that shows hierarchy.

The branch trends are clear in the dendroidal contour because the line direction of the upper-level branch that suggests rough trends is darker in color than that of the lower-level branch. The shape of the nesting contour weakly suggests the spread of branches; however, the nesting contour has no clear branch lines. The island-like contour has fewer clues for understanding branch lines than the nesting contour.

The arrangement is clear in the island contour because the color at the center position of a node is always darker than those at the surrounding positions. The accurate position of a node is often vague in the nesting and dendroidal contours.

There are many criteria for evaluating the design; however, we adopt natural designs because for knowledge landscapes, we focus on a scenic shape. Dendroidal and island-like shapes are considered to be natural by people. The nesting shape is more mechanical than the other two shapes.

The three algorithms can be adapted according to the user's policy of arrangement design. The nesting contour is suited for representing formally categorized contents such as product catalogs. The dendroidal and island-like contours are suited for presentations where an appealing design and arrangement are important.

Table 1. Comparison of the three contour algorithms

	Hierarchies	Branch trends	Arrangement	Design
Nesting	Clear	Even	Even	Even
Dendroidal	Even	Clear	Even	Popular
Island-like	Vague	Vague	Clear	Popular

6 Conclusion

This article described a model for sustainable contents management, its visualization algorithms, and the implemented system, called sustainable knowledge globe. Three contour map representations were proposed to briefly depict the arrangement and structure of contents. The effectiveness of nesting, dendroidal, and island-like contours were compared from the viewpoint of preservation of the structures and arrangements, in addition to supporting shape and texture design.

As future work, we are planning to investigate the effect of memory support by psychological experiments.

References

1. Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, Curtis Wong: MyLifeBits: Fulfilling the Memex Vision. ACM Multimedia '02 (2002) 235–238
2. Yasuyuki Sumi, Sadanori Ito, Tetsuya Matsuguchi, Sidney Fels, Kenji Mase: Collaborative capturing and interpretation of interactions. Pervasive 2004 Workshop on Memory and Sharing of Experiences (2004) 1–7
3. Takaya Yuizono, Jun Munemori and Yoji Nagasawa: GUNGEN: groupware for a new idea generation consistent support system. Proceedings of Asia Pacific Computer Human Interaction1998, IEEE Press (1998) 357-362
4. Kazuo Misue and Kozo Sugiyama: How Does D-ABDUCTOR Support Human Thinking Processes? Proceedings of CG International '94 – Insight through computer graphics, (1994) 10–21
5. Jiro Kawakita: Hassoho. Chuokoron-sha (1967) (In Japanese)
6. George Robertson, Mary Czerwinski. Kevin Larson, Daniel C. Robbins, David Thiel, Maarten van Dantzich: Data mountain: using spatial memory for document management. In Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology (UIST '98) (1998) 153–162
7. Toyoaki Nishida: Social Intelligence Design and Communicative Intelligence for Knowledgeable Community. Invited Talk, International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society DLKC'04, Kasuga Campus, University of Tsukuba, March 2 - 5, (2004)
8. Toyoaki Nishida: Conversation Quantization for Conversational Knowledge Process. Special Invited Talk, S.Bhalla (Ed.): DNIS2005 (Fourth International Workshop on Databases in Networked Information Systems), LNCS 3433, Springer (2005) 15–33

9. Ken Saito, Hidekazu Kubota, Yasuyuki Sumi, Toyoaki Nishida: Support for Content Creation Using Conversation Quanta. In: T. Washio et al. (Eds.): JSAI 2005 Workshops, LNAI 4012, (2006) 29–40
10. Hidekazu Kubota, Sadao Kurohashi, Toyoaki Nishida: Virtualized-egos using Knowledge Cards. In Proceedings of Seventh Pacific Rim International Conference on Artificial Intelligence (PRICAI-02) WS-5 International Workshop on Intelligent Media Technology for Communicative Reality (IMTCR2002), (2002) 51–54
11. Hidekazu Kubota, Yasuyuki Sumi, Toyoaki Nishida: Sustainable Knowledge Globe: A System for Supporting Content-oriented Conversation. In Proceedings of AISB 2005 Symposium Conversational Informatics for Supporting Social Intelligence & Interaction, (2005) 80–86
12. Benjamin B. Bederson, James D. Hollan: Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. Proceedings of ACM UIST '94, (1994) 17–26
13. George W. Furnas: Generalized fisheye views. Human Factors in Computing Systems CHI '86 Conference Proceedings (1986) 16–23
14. John Lamping, Ramana Rao, Peter Pirolli: A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems (1995)
15. Jin Young Hong, Jonathan D' Andries, Mark Richman, Maryann Westfall: Zoomology: Comparing Two Large Hierarchical Trees. First Prize in Ninth Annual IEEE Symposium on Information Visualization (InfoVis) contest, October (2003) 19–21
16. Keith Andrews, Josef Wolte, Michael Pichler: Information pyramids: A new approach to visualising large hierarchies. In Proceedings of IEEE Visualization'97, Late Breaking Hot Topics, (1997) 49–52
17. Jun Rekimoto, Mark Green: The Information Cube: Using Transparency in 3D Information Visualization. In Proceedings of the Third Annual Workshop on Information Technologies & Systems (WITS'93) (1993) 125–132
18. James F. Blinn: A Generalization of Algebraic Surface Drawing. ACM Transaction on Graphics, Vol.1, No.3, (1982) 235–256

Discussion Ontology: Knowledge Discovery from Human Activities in Meetings

Hironori Tomobe and Katashi Nagao

Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan
tomobe@nagao.nuie.nagoya-u.ac.jp
http://www.nagao.nuie.nagoya-u.ac.jp/members/tomobe.xml

Abstract. Discussion mining is a preliminary study on gathering knowledge based on the content of face-to-face discussion meetings. To extract knowledge from discussion content, we have to analyze not only the surface arguments, but also semantic information such as a statement's intention and the discussion flow during meetings. We require a discussion ontology for this information. This discussion ontology forms the basis of our discussion methodology and requires semantic relations between elements in meetings. We must clarify these semantic relations to build the discussion ontology. We therefore generated discussion content and analyzed meeting metadata to build the ontology.

1 Introduction

Meetings are a normal part of business and academic settings. For example, companies have meetings, which involve deciding the time and place beforehand. However, they also have casual meetings, which neither the time nor place is decided. The themes of meetings are not only for decision making but also for brainstorming to have participants generate ideas. Meetings are indispensable as a place for exchanging opinions with others and for forming a social network.

Knowledge must be shared efficiently in brainstorming meetings. Meetings in which participants obtain a lot of knowledge over a short period of time tend to be evaluated highly, while meetings in which participants obtain only a little knowledge even though they spend a long time in them tend not to be evaluated highly. Therefore we need to determine the elements that are necessary to hold efficient meetings.

We hypothesize that three elements are important for holding efficient meetings: specifying information for on-going discussion, generating minutes that have high reproduction value, and improving the skill of the participants. To achieve these three things, we require a discussion ontology, and this ontology requires that we clarify the semantic relations between elements in a meeting.

To build a discussion ontology, we generated discussion contents with structure for a meeting and analyzed the metadata in the contents. Using browser-based tools for acquiring information about meetings in the real world, we structurized the discussion contents. We inferred the intentions of statements in a

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 33–41, 2007.

meeting based on classification rules generated by a machine learning algorithm and extracted the discussion flows.

2 Discussion Ontology

We define the description of discussion elements and semantic relations between these elements as a discussion ontology. Each meeting is composed of discussions, and each discussion is composed of statements. We define these discussions and statements as discussion elements. Similarly, participants and meeting materials (such as presentation slides and handouts) are defined as discussion elements. Semantic relations exist between these elements: for example, dependence exists between discussions and statements, similar relations exist between statements, and correspondence relations exist between meeting materials and discussions.

In addition, the semantic relations of discussion elements include some restrictions. For example, restrictions such as "After a participant asks a question (one intention), another one must answer (another intention)" are important for keeping the discussion flowing smoothly in meetings. A system can determine the on-going discussion situation and assist with meetings by specifying this situation by clarifying the semantic relations and restrictions. The discussions are also formalized. We attempted to segment discussion based on discussion elements and to clarify semantic relations and restrictions on discussion elements by accumulating information about meetings and by analyzing these metadata.

3 Gathering Metadata for Generating Discussion Content

3.1 Structuring Discussions

Meetings in the real world (i.e., face-to-face meetings) progress on the basis of a time series. Discussion elements like statements and discussions, however, cannot be expressed in a time series. Therefore, discussions need to be properly structured to understand discussions semantically.

Structurization of discussions means clarifying elements that make up a discussion and the semantic relations between the elements. In other words, structurization of discussions involves segmenting meetings based on discussion elements and adding links between segments. The unit of segmentation for discussions and the semantic relations of each element are different for each meeting. Therefore, this study deals with meetings that have presentations because, in these meetings, the moderator projects her/his presentation material on a screen, making the targets of discussion clearer.

However, extracting discussion elements and linking the elements is difficult. Therefore, we used discussion mining[1] to structure discussions semi-automatically.

3.2 Discussion Mining

Most studies on providing technology for discussions and generating minutes have focused on automatic recognition techniques for audio and visual data,

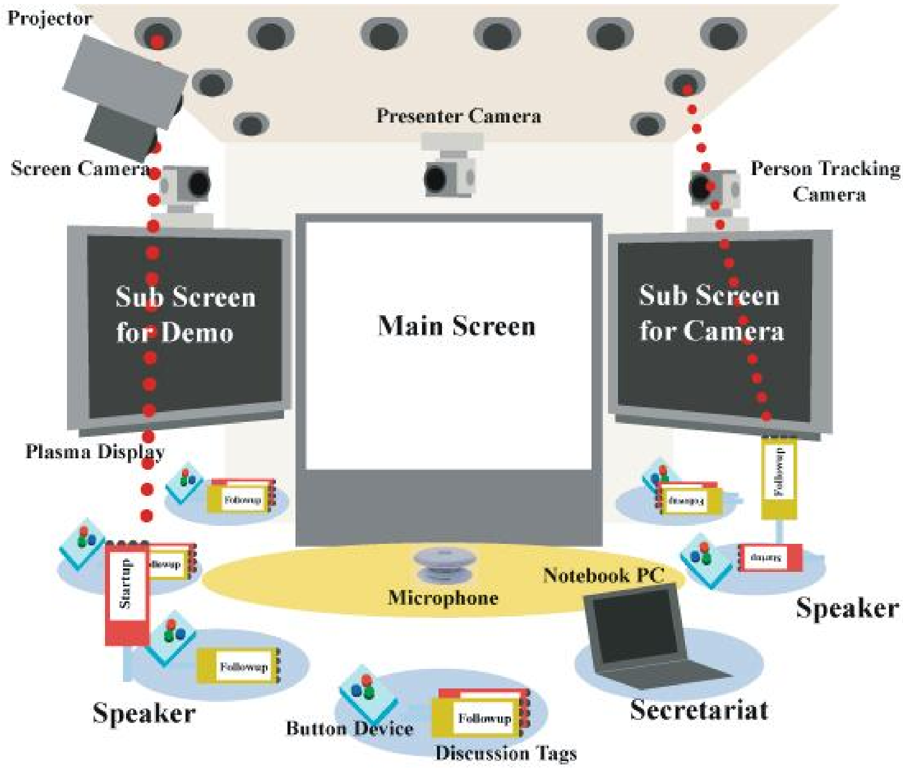


Fig. 1. Discussion Mining System

such as meeting browsers[2]. We used a discussion mining system for generating discussion content. Figure 1 shows an image of the discussion room.

We targeted face-to-face meetings in the discussion mining. Detailed recording is done during the meeting with four cameras and a microphone that are installed in the meeting room. One camera records the main screen, another camera records the presenter's face, and the two remaining cameras record the participants' actions. Audio information is recorded using a microphone installed in the center of the meeting room. Participants in the meeting transmit their IDs and comments using tag devices called discussion tags to enable the discussion to be properly structured. Furthermore, using a button device, the participants can arbitrarily input their stance toward the presentation and any arguments with other participants. This information is added to the minutes in real time and is edited by the secretary. Currently, the secretary inputs text manually, though in the future, this will be done automatically with speech recognition technology. A record of the arguments in XML and MPEG-4 format is saved in an XML database as multimedia minutes.

There are two types of discussion tags: "start-up" and "follow-up." Participants use the startup tag to make remarks that trigger a discussion and use the follow-up tag to make remarks that relate to an ongoing discussion. The system uses these tags to segment the discussion automatically to enable an analysis of the discussion and viewing of the video.

3.3 Usage of Discussion Content

This system enables us to visualize the structure of minutes by creating a graphical display and edit mode for statements with the use of scalable vector graphics (SVG). The graph is semi-automatically structured with pertinent information and keywords of statements and slides, as shown in Fig. 2. This function allows users to edit the content.

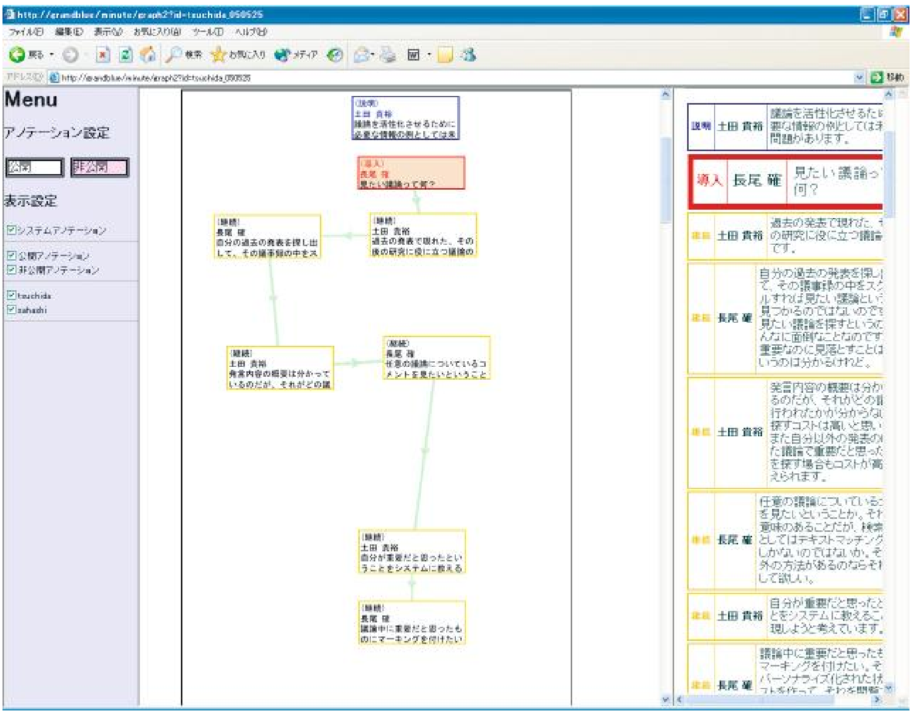


Fig. 2. Graph Viewer of Discussion Content

The detailed discussion can be understood from watching the video recorded by the discussion mining system. However, efficiently inspecting the video is a problem. Our discussion mining system provides a tool for effectively viewing discussion content recorded in the meeting. This system provides a meeting video with a support system for discussion content browsing (shown in Fig. 3).

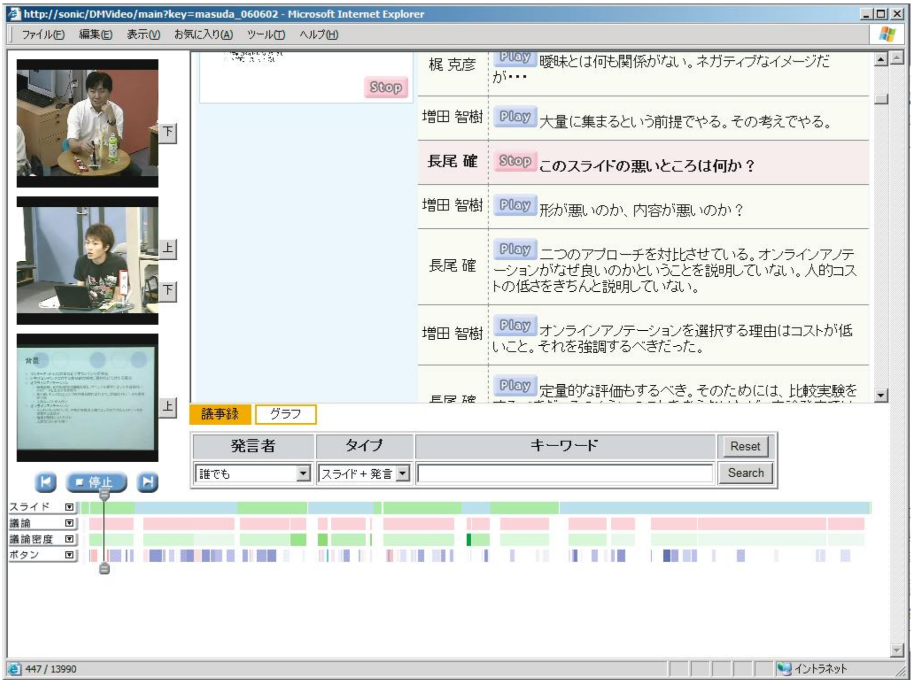


Fig. 3. Support System for Discussion Content Browsing

4 Classification of Statements and Extraction of Discussion Flow

Each statement in a discussion consists of two kinds of information: the topic and the intention. The former is important for understanding the discussion theme and specific content. The latter is important for understanding the discussion flow. In this section, we describe a method for extracting the intention. Knowing the intention of statements should enable understanding the discussion flow. Using this flow and adding the importance of discussion will enable users to find important discussions easily.

The procedure of this method is as follows. First, we determine the intention of each statement. We define an intention tag set for meetings from the information on meetings that we have accumulated over the previous four years. Each intention tag expresses the intention included in each statement, and we limit the number of types of intention tags. Second, we classify each statement based on its intention tag. We infer intention tags for statements by using attributes of the statements such as speaker and type of statement, because adding intention tags to all statements manually is time consuming and difficult. Finally, after we add intention tags to all statements, we find the intention tags that often co-occur in a discussion.

4.1 Intention Tag

The discussion can be analyzed semantically by adding intention tags to each statement in a meeting. We can use intention tags to understand the discussion flow, and the features of face-to-face communication can be clarified by analyzing the local discussion structure, such as "requests" and "replies."

The intention tag set needs to be based on a standard tag set to improve reliability. The tag set naturally varies based on the usage, but we must add tags without any ambiguity. We created intention tag sets based on a standard utterance-unit tagging scheme [3].

We manually added standard utterance-unit tags to all four-minute statements, and we define the tag set to use based on the frequency distribution of the standard utterance-unit tag.

- **Request for unknown information:**
  a statement in which the speaker wants some values as a reply
- **Request for true or false information:**
  a statement in which the speaker wants "yes" or "no" as a reply
- **Reply to unknown information:**
  a statement in which the speaker replies to a request for unknown information
- **Reply to true or false information:**
  a statement in which the speaker replies to a request for true or false information
- **Reply suspension:**
  a statement in which the speaker does not reply directly to the request for information
- **Information presentation:**
  a statement in which the speaker presents her/his knowledge
- **Suggestion/Proposal:**
  a statement in which the speaker suggests or proposes actions to be taken by other participants
- **Opinion/Hope:**
  a statement that presents the speaker's ideas or desired outcome

4.2 Classification of Statements

We stored more than 370 minutes of meetings in the discussion mining system, and about 14,500 statements are included in the minutes. Because classifying these statements into an intention tag set manually is difficult, we inferred which tags were added to each statement by using a machine learning algorithm, C4.5 [4].

We used the following metadata of statements as attributes: "speaker (presenter or participants except for the presenter)," "type of statement (start-up or follow-up)," "speaking time," "occurrence of keywords," etc. In addition, we added an intention tag to three minutes manually to use them as training data and generated classification rules for the intention tag. Examples of the classification rules are shown below.

- **Request for unknown information**
  type of statement = *start-up*
  speaker = *participants except for the presenter*
  occurrence of keywords = *yes*
  speaking time = *short (less than 20 seconds)*

- **Opinion/Hope**
  type of statement = *start-up*
  speaker = *participants except for the presenter*
  occurrence of keywords = *no*
  speaking time = *mid-length (21 - 40 seconds)*

When a statement had more than two types of intention tags as a result of applying the classification rules, we added all of the intention tags to the statement because there are many statements that should have several intention tags. For example, after one participant speaks with the intention "Request for unknown information," another participant may reply with an intention that includes not only the "Reply to unknown information" but also an "opinion/hope." The classification rules' precision was 83%; we tuned the intention tags manually after applying the classification rules.

4.3 Extraction of Discussion Flow Based on Statement Intention

The statement type, either start-up or follow-up, is added to each statement using discussion tags in the discussion mining system. All discussions must start

Table 1. Discussion Flows That Appeared Frequently

intention tag1	intention tag2	frequency
Request for unknown information	Reply to unknown information	307
Request for unknown information	Information presentation	215
Request for true or false information	Reply to true or false information	209
Suggestion/Proposal	Information presentation	207
Reply to unknown information	Request for unknown information	167
Information presentation	Suggestion/Proposal	152
Suggestion/Proposal	Suggestion/Proposal	120
Reply to unknown information	Request for true or false information	78
Information presentation	Opinion/Hope	65
Request for true or false information	Information presentation	41

with a start-up statement and continue with a follow-up statement when participants discuss topics continuously. We define a discussion that has one start-up statement and related follow-up statements as a unit of discussion, and we call this a discussion segment.

We extracted discussion flows from 20 minutes and investigated the flows that appeared frequently and the co-occurrences of statement intention. Table 1 shows the ten flows that appeared frequently.

We believe that a discussion flow that appears frequently is a model of efficient discussion in meetings. This kind of discussion flow is part of the grammar of the discussion, and the discussion according to this grammar will obtain better results for participants in meetings. In contrast, a discussion flow that deviates from this grammar could become a diverging discussion, thereby preventing the meeting from being efficient.

Discussion flows have several potential applications. They can be used to detect the similarities and differences between discussions. A discussion flow can be considered a feature of a discussion, and a discussion grammar can be defined by using this feature. The discussion grammar can be used to summarize the discussion and prepare the minutes. For example, when a discussion in which questions and answers are repeated has the flow "answer - question - answer - question -...," it can be summarized by omitting this repetition pattern.

5 Conclusion and Future Work

We have described a method for classifying statements in discussions and extracting discussion flows. Analysis of the metadata of a structured discussion clarified the importance of the discussion elements and revealed information necessary for understanding discussions. Moreover, the semantic relations between discussion elements, which consist of discussion content, can be clarified, and a discussion ontology can be built. We believe that natural language processing can be used to discover knowledge efficiently based on the topics of discussions.

The minutes of a meeting and any audio and visual data recorded in the meeting play an important role because from them we can record the details of the discussions. We therefore plan to clarify the role of audio and visual data in discussion ontology.

Acknowledgements

This work was supported by the 21st Century COE Program on Intelligent Media Integration at Nagoya University. The authors thank Shigeki Ohira, Takahiro Tsuchida, Meguru Ito, Issei Naruta and Ryosuke Hayashi for their useful suggestions and discussions.

References

1. Nagao, K., Kaji, K., Yamamoto, D. and Tomobe, H.: Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities. Proc. of the Fifth Pacific-Rim Conference on Multimedia (PCM 2004) Part 1. (2004) 522–531
2. Schultz, T., Waibel, A., Bett, M., Metze, F., Pan,Y., Ries, K., Schaaf, T., Soltau, H, Martin, W., Yu, H. and Zechner, K.: The ISL Meeting Room System. Proc. of the Workshop on Hands-Free Speech Communication (HSC-2001). (2001)
3. Araki, M., Itoh, T., Kumagai, T., and Ishizaki, M.: Proposal of a standard utterance-unit tagging scheme. Journal of Japanese Society for artificial intelligence, Vol.14 No.2. (1999) 251–260 (In Japanese)
4. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann. (1993)

Predicting Types of Protein-Protein Interactions Using a Multiple-Instance Learning Model

Hiroshi Yamakawa, Koji Maruhashi, and Yoshio Nakao

Fujitsu Laboratories Ltd.
1-1, Kamikodanaka 4-chome, Nakahara-ku, Kawasaki, Kanagawa, 211-8588, Japan
{ymkw, maruhashi.koji, ynakao}@jp.fujitsu.com

Abstract. We propose a method for predicting types of protein-protein interactions using a multiple-instance learning (MIL) model. Given an interaction type to be predicted, the MIL model was trained using interaction data collected from biological pathways, where positive bags were constructed from interactions between protein complexes of that type, and negative bags from those of other types. In an experiment using the KEGG pathways and the Gene Ontology, the method successfully predicted an interaction type (phosphorylation) to an accuracy rate of 86.1%.

1 Introduction

In recent molecular biology and its application fields including drug discovery, analysis of protein-protein interactions (PPIs) is an emerging issue to elucidate the mechanism of biological processes. Since PPIs play a central role in numerous cellular processes, understanding PPIs provides us with clues to determining potential drug targets in cases where drug targets are identified from a known pathway related to a disease. Although a large volume of PPI data has been collected as described later, many other unknown PPIs are believed to exist (Rhodes et al., 2005). Moreover, only a few PPIs have been elucidated at the functional level. In order to resolve this situation, an approach that combines wet and dry technologies is promising, so that one might construct hypotheses on a biological mechanism based on confirmed PPI data produced by a wet technology with plausible PPI data predicted by a dry technology.

As for PPI prediction, Rhodes et al. (2005) proposed a probabilistic method that integrates model organism interactome data, protein domain data, genome-wide gene expression data and functional annotation data; and Lee et al. (2005) presented an assessment scheme for the reliability of PPI candidates based on a neural network algorithm. However, the aim of these studies is discovering novel PPIs or filtering correct PPIs, not identifying the interaction types between proteins (e.g., activation, inhibition, phosphorylation, or the like).

Although PPI types are essential in describing the mechanism of a biological process, only a few existing PPI databases provide interaction types (Ekins, Nikolsky, and Nikolskaya, 2005). For example, the Human Protein Reference

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 42–53, 2007.

Database (HPRD) (Peri et al., 2003), which is one of the most famous databases providing a large volume of PPI data, does not provide interaction types although it stores 33,710 entries of PPI data at the time of September 2005 and its size continues to grow. Although the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) provides pathway data in terms of protein-protein interaction networks with interaction types, the number of its PPI entries is only several thousands.

PPI type prediction is, therefore, an important issue to understand biological processes. In this paper, we propose a method for predicting PPI types based on a machine learning model using known PPI types provided by the KEGG database (Kanehisa and Goto, 2000) as training data.

2 A Model of Interaction Between a Protein Complex Pair

A PPI described in existing pathways, as provided by the KEGG database (Kanehisa and Goto, 2000), often corresponds to a pair of protein complexes (see rounded rectangles in Fig. 1), each of which is composed of several subunits (simple proteins). On the other hand, functional annotations, as provided by the Gene Ontology (Ashburner et al., 2000), have been accumulated mainly for simple proteins.

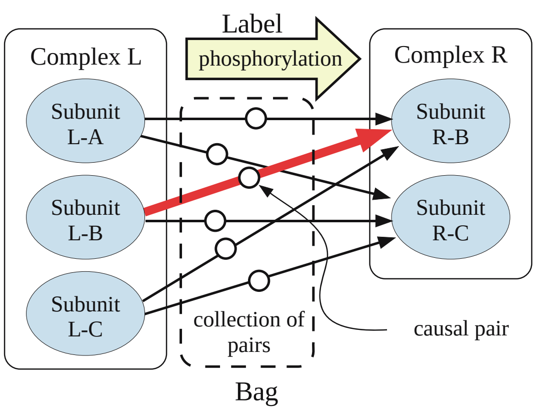


Fig. 1. A MIL Bag Model for PPI Type Prediction: each rounded rectangle depicts a protein complex; the upper block arrow depicts an interaction type (i.e., a label to be learned) between complexes; each simple arrow depicts a possible subunit pairing (i.e., an instance of the MIL model) and the broad arrow indicates the subunit pair that is the only cause of the process that Complex L phosphorylates Complex R

In other words, training data for interaction type predictions are given in terms of a collection of complex pairs labeled with an interaction type (i.e., the target variable); while available features (i.e., the input variables) for machine learning are given for each subunit (simple protein) composing complexes. Figure 1 depicts a case where the target variable is whether the interaction type is phosphorylation and the input variables are given for each subunit of L-A, L-B, ..., and R-B.

It is difficult for a standard supervised learning method to solve a problem of this kind because the relationship between the input variables and the target variable is ambiguous. With regard to this point, we assume that the interaction type between complexes can be determined by a subunit pair across those complexes. In other words, we ignore those cases, for example, where two or more subunits of a complex work cooperatively, or where the active site of a subunit of a complex is hidden by other subunits of that complex. Hereafter, this assumption is referred to as the *subunit reaction assumption.*

This assumption can be validated based on the following discussion. From the protein structure viewpoint, an interaction between a protein pair is often explained by one or a small number of domain[1]-level interactions across those proteins. This fact suggests that an interaction between a protein complex pair can also be reduced to just a few domain-level interactions even though it cannot always be reduced to those of a protein pair across those complexes. In molecular biology, target validation (e.g., validation of a hypothesis that a particular protein causes a particular disease) is often performed by knocking-out the gene that encodes a target protein. The effectiveness of this approach suggests that many individual biological functions originate from a specific protein. Accordingly, the subunit reaction assumption is almost as valid as the tacit assumption based on which biologists may take the above-mentioned approach to target validation.

The assumption also has another advantage in that prediction methods based on it are expected to predict both the interaction types of complex pairs and those of subunit (protein) pairs, the former of which may describe the behavior of a biological system while the latter may indicate the molecular function that could be controlled with chemical substances (i.e., potential drugs).

3 A Method for Predicting Interaction Types Based on Multiple Instance Learning

In this section, we propose a method for predicting PPI types based on a multiple-instance learning (MIL) scheme. As shown in Fig. 1, the PPI-type prediction task based on the subunit reaction assumption can be formalized as a problem of MIL as follows: a complex pair (two rounded rectangles) with an interaction type (the upper block arrow) is formulated as a labeled bag, and a possible subunit pair across a complex pair (the six simple arrows) as an instance.

[1] A portion of a protein thought to have specific molecular functions and often corresponding to a characteristic sequence of amino acids.

3.1 Multiple-Instance Learning Scheme

Multiple-instance learning (MIL) is a scheme of semi-supervised learning for problems with incomplete knowledge concerning the labels of the training data. In the MIL training data, labels to be learned are only assigned to bags of instances but not to individual instances, while every training instance is labeled for supervised learning. For example, in a binary classification problem, a bag is labeled positive if at least one instance in that bag is positive, while the bag is labeled negative only if all the instances in it are negative. The goal of MIL is to predict labels of unseen bags and/or instances based on those labeled bags.

In the pioneering work of (Dietterich, Lathrop, and Lozano-Perez, 1997), MIL was applied to drug activity estimation. Following this, many MIL methods have been proposed and applied to various fields including image classification (Maron and Lozano-Pérez, 1998), stock selection (Maron and Lozano-Pérez, 1998), text classification, face labeling in broadcasting news video (Yang, Yan, and Hauptmann, 2005), Web mining (Zhou, Jiang, and Li, 2005), etc.

3.2 Diverse Density

In this research, we solve the problem using a modified version of the diverse density (DD) framework (Maron and Lozano-Pérez, 1998), which is one of the well-known MIL solutions. The main idea of the DD framework is to find a desired concept point (i.e., the point desired for positive identification) in the feature space that is close to at least one instance from every positive bag and far away from any instances in negative bags. Desired concept points are found according to a score called diverse density (dd), which is a measure of how many different positive bags have instances near the point, and how far the negative instances are away from that point. The optimal concept point is defined as the one with the maximum dd.

A probabilistic version of dd at a point x is calculated using positive bags B_i^+ and negative bags B_i^- by the following formula:

$$dd\,(x) \propto \prod\nolimits_i \Pr\left(x|B_i^{\pm}\right) \tag{1}$$

where $B_{ij}^{\pm}$ denotes the j-th instance in the bag $B_i^{\pm}$; $Pr(x|B_i^{\pm})$ denotes a contribution score of an instance in the bag $B_i^{\pm}$, which is calculated as a likelihood score whether an instance in the bag $B_i^{\pm}$ is near the location x (described later).

Each bag B_x (i.e., a case whose interaction type is predicted) is evaluated by a score at the optimal concept point in that bag, which can be calculated by the following formula:

$$dd(B_x) = \max_m dd(x^{(m)})$$

where $x^{(m)}$ is the point at which instance m is located. It is judged positive if the score is larger than a threshold value(≥ 0).

Every contribution score is calculated using a *noisy-or* model as follows:

$$\Pr\left(x|B_i^+\right) = 1 - \prod_j \left(1 - \Pr\left(x|B_{ij}^+\right)\right)$$
$$\Pr\left(x|B_i^-\right) = \prod_j \left(1 - \Pr\left(x|B_{ij}^-\right)\right) \tag{2}$$

where $\Pr\left(x|B_{ij}^\pm\right)$ denotes the contribution score of an instance $B_{ij}^\pm$, which is calculated using a Gaussian-like distribution as follows:

$$\Pr\left(x|B_{ij}^\pm\right) = \exp\left[-\sum\nolimits_k s\left(B_{ijk} - x_k\right)^2\right] \tag{3}$$

where k denotes the index of the axes of the feature space, and s denotes the scale factor.

3.3 Modified Diverse Density as a Weighted Voting System

In a preliminary experiment, we found that the diverse density score (*dd*) did not work well for our problem. We found many false negative cases related to a few instances from negative bags. This was because the *dd* score calculated by (1) is very sensitive to the contribution score of a negative bag: i.e., the *dd* score is calculated as very low even in the case where only one negative instance is near the point x.

Relating this to our particular problem, this could represent a case where a subunit pair with an interacting potential is inhibited from interaction in such a condition that the active domain of a subunit of a complex is hidden by other subunits of that complex. Accordingly, a point around which many positive instances are near should be given a higher score even though there are a few negative instances near to that point. In this respect, we introduced another diverse density score, *vdd*, based on the following formula:

$$vdd(x) \propto \sum_i \mathrm{sign}_i \left[1 - \prod_j \left(1 - \Pr\left(x|B_{ij}^\pm\right)\right)\right] \tag{4}$$

where sign_i is $+1$ for positive bags and -1 for negative bags. The *vdd* score can be interpreted as a weighted voting system where the absolute value of a voting weight is a likelihood score of whether any of the instances in a bag are at the point. Hereafter, this score is referred to as the *voting diverse density* (*vdd*).

Each bag B_x is evaluated by a score $vdd(B_x) = \max_m vdd(x^{(m)})$ ($x^{(m)}$ is the point at which instance m locates). It is judged positive if the score is larger than a threshold value (≥ 0). The experiments described later used a threshold value of 0.

4 PPI Dataset and Feature Space

In this study, we use the KEGG pathways (Kanehisa and Goto, 2000) for dataset construction, and the Gene Ontology (Ashburner et al., 2000) for feature space construction.

4.1 PPI Dataset Obtained from the KEGG Pathways

The PPI dataset for the experiment was constructed from the KEGG pathways as of March 2006. Firstly, we obtained human pathway data in XML format from the KEGG site[2], and then extracted PPI data, i.e., the *relation* elements whose value for the *type* attribute was *PPrel*. Each record in the PPI data comprises two groups of proteins, either or both of which may correspond to a protein complex, a protein family, or a simple protein, and one or more interaction types, which is described in the *subtype* element in the XML file. As a result, 1,279 different PPI records were obtained.

Table 1 summarizes the distribution of interaction types. The rows from *state* to *compound* individually correspond to an interaction type; the columns from 1 to 17 show the assignment patterns of the interaction types where the cells with a value of 1 in the same column indicate that all the interaction types corresponding to those rows were assigned to one or more identical PPI records. For example, the 6th column indicates that there were six records labeled with both interaction types of *ubiquination* and *inhibition*.

Table 1. Distribution of Interaction Types in PPI Records from the KEGG Pathways

Interaction Type	Assignment Pattern [Dataset Type]																	# Recs.
	1 [−]	2 [−]	3 [−]	4	5	6	7	8 [+]	9 [+]	10 [+]	11 [−]	12	13 [+]	14	15	16 [+]	17 [−]	
state	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4
ubiquination	.	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	13
dephosphorylation	.	.	1	.	.	.	1	.	.	.	.	.	.	.	1	.	.	25
dissociation	.	.	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	14
inhibition	.	.	.	.	1	1	1	.	.	1	.	.	.	.	.	.	.	198
phosphorylation	.	.	.	.	.	.	.	1	1	1	.	.	1	.	.	1	.	249
binding association	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	181
indirect	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	102
activation	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	.	588
compound	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	27
#Records	4	7	8	13	145	6	6	150	1	41	181	97	5	525	11	52	27	1,279

4.2 Feature Space Encoding the Gene Ontology

Feature Vector: A feature vector of each instance, which comprises two proteins and an interaction type, was constructed using the Gene Ontology (GO) (Ashburner et al., 2000). The GO is a vocabulary that describes the attributes of genes. Each term in the vocabulary, called a GO term, represents a possible attribute value possessed by a gene or a protein encoded by genes. The GO has a hierarchical structure, i.e., GO terms are connected by *is-a* relations and construct a directed acyclic graph. GO currently consists of three standard gene ontologies that describe biological processes, cellular components, and molecular

[2] ftp://ftp.genome.jp/pub/kegg/xml/KGML_v0.5/hsa

functions. Many biological resources use the GO terms to annotate the properties of genes or their products (i.e., proteins or RNAs encoded by genes).

An initial feature vector of each instance consists of two components, each of which represents a list of GO terms annotated to one of the proteins composing that instance. The list of GO terms were obtained from the *gene2go* file[3] provided by the National Center for Biotechnology Information (NCBI). The list of GO terms for each protein was extended by adding all the ancestors in the GO hierarchy for each term in the initial list.

Feature Space Construction by Singular Value Decomposition: It is difficult for MIL to use the initial feature vectors. This is because a GO term in the higher layers of the GO hierarchy often relates to too many instances to discriminate different ones, and ones in the lower layers often relate to too few instances to generalize similar instances. In addition, the large size of the GO term set[4] requires expensive computational cost for MIL.

To overcome these problems, the feature space was compressed by using Singular Value Decomposition (SVD).

Logarithmic Probability Weighting: The SVD process tends to emphasize those properties that appear frequently. On the other hand, a GO term in the higher layers of the GO hierarchy often relates to many instances. If all the GO terms are equally weighted, an ineffective feature space will be obtained that emphasize those GO terms in the higher layers of the GO hierarchy.

In this respect, each GO term, j, is weighted with a logarithmic probability weight, W_j, which highlights a moderate abstraction level (depth) in the GO hierarchy. W_j is calculated using the following formula:

$$W_j = -\log(p_j) = -\log(\frac{m_j}{m}) \tag{5}$$

where m is the total number of genes appearing in the *gene2go* file, and m_j is the number of genes with term j in that file. For instance, the topmost term in the hierarchy, which relates to all the genes, is weighted with zero ($W_j = 0$), and is then ignored. On the other hand, a term in the middle layers, which is expected to have an appropriate specificity, is emphasized.

Singular Value Decomposition (SVD): In preparation for the SVD process, a matrix G that represents the relationship between the instances and the GO terms was constructed. A cell G_{ij} is set as W_j if the first protein of instance i relates to term j, or is set as $W_{j'}$ if the second protein of instance i relates to term j' where $j' = j - n_1$ (n_1 is the number of elements for the first proteins of instances), otherwise it is set as 0.

SVD decomposes matrix G as follows:

$$G = USD^T \tag{6}$$

[3] ftp://ftp.ncbi.nih.gov/gene/DATA/

[4] The molecular function ontology contains 622 different terms.

where U and D are a unitary matrix that satisfies $U^T U = I_m$ and $D^T D = I_n$ respectively. The column vector of U is called the left singular vector. The column vector of D is called the right singular vector. The diagonal entries of S are called singular values.

A compressed feature matrix $\tilde{G}$ can be composed using selected right singular vectors d_x, d_y, $\ldots$ as follows:

$$\tilde{G} = G\tilde{D} \text{ where } \tilde{D} = [d_x, d_y, \ldots] \tag{7}$$

5 Experiments: Binary Classification Task Concerning Phosphorylation

This section reports on an experiment to evaluate the proposed method using a PPI prediction task as binary classification concerning an interaction type of phosphorylation. In the experiment, the proposed method was implemented on MATLAB. This is because MATLAB provides an efficient implementation of SVD for a sparse matrix.

5.1 Binary Classification Task Concerning Phosphorylation

Since the GO provides three kinds of ontologies as described in Sect. 4.2, the input variables (i.e., feature vectors) can be constructed by several different approaches concerning which ontologies should be encoded into the feature vectors. In addition, since there are two proteins comprising an instance, the relation across the properties of those proteins can be encoded into the feature vectors. On the other hand, since some bags (i.e., PPI records) have two or more interaction types (Table 1), we can construct complicated tasks where certain combinations of interaction types are formalized as the target variable.

Among those possible tasks, we chose one of the simplest ones: the task of classifying PPIs into binary classes based on whether the PPI type is phosphorylation or not by using the feature vectors constructed from the molecular function ontology. This is because preliminary experiments suggested that this task is most promising among the simpler ones.

By taking account of the distribution of PPI type assignment patterns shown in Table 1, positive and negative bags were constructed as follows.

A set of positive bags was constructed by selecting those PPI records of which one of the interaction types was phosphorylation. As a result, 249 PPI records, those columns with a '[+]' mark in Table 1, were selected as positive bags.

A set of negative bags was constructed by selecting those PPI records of which neither of the interaction types appeared in any records in the set of positive bags. As a result, 227 PPI records, those columns with a '[-]' mark in Table 1, were selected as negative bags.

The remaining 803 PPI records were those whose interaction types included a type appearing in some positive bags but not phosphorylation. For example, PPI records that had only one interaction type of *activation* or *inhibition*.

5.2 Results

There are two major parameters of the proposed method: the dimension of the feature space and the scale factor of the Gaussian-like distribution (3) concerning the voting diverse density. The optimal values of these two parameters were explored in the range described below.

The compressed feature space based on the first few singular vectors by the SVD process is expected to represent major components of the initial feature space. In the experiment, twelve kinds of feature spaces of different dimensions were constructed by (7) with topmost n right singular vectors where $n \in \{[2..10], 12, 15, 20\}$.

The experiments also used the following twelve values in exploring the scale factor s: {500, 1000, 2500, 5000, 10000, 20000, 40000, 80000, 160000, 320000, 640000, 1280000}.

Figure 2 summarizes the results by the accuracy score in leave-on-out cross-validation for each parameter setting. The best accuracy of 86.1% was obtained at the point of $n = 4$ and $s = 160,000$. A comparable accuracy of 85.6% was obtained at the point of $n = 10$ and $s = 10,000$, and local peaks of the accuracy were found along a line connecting those two points. These observation suggest that a smaller scale factor is better for the feature space of a higher dimension. This could be explained by a tendency that the higher the dimension of the feature space, the longer the distance between instances.

Figure 3 is a scatter graph that plots every instance in the condition when the best accuracy was obtained ($n = 4$ and $s = 160,000$). The x-axis of this graph

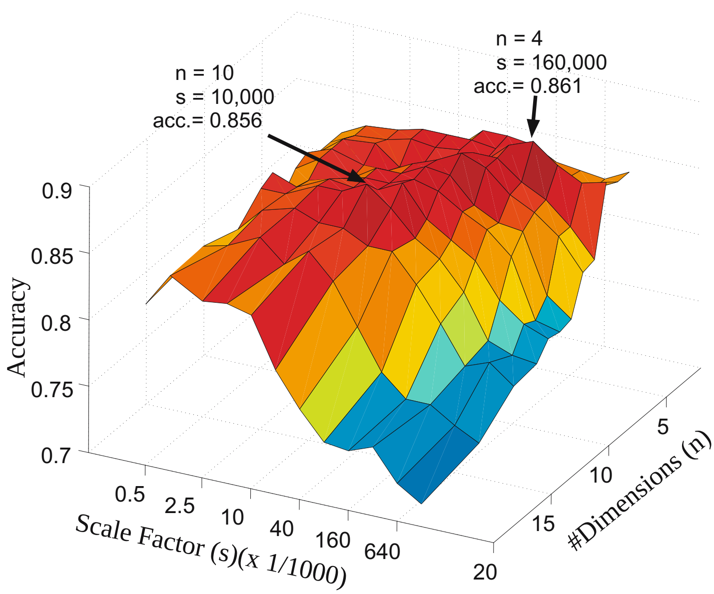


Fig. 2. Prediction Accuracy according to Two Parameters

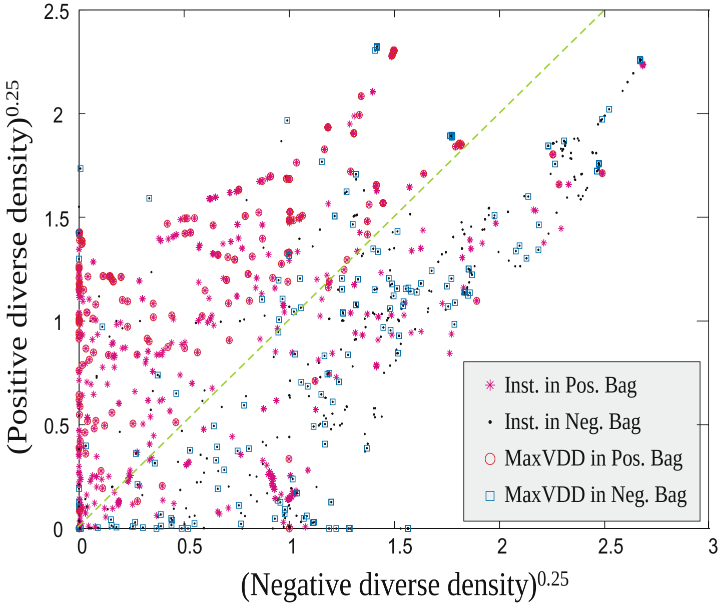


Fig. 3. Contribution of Positive and Negative Bags to Voting Diverse Density Score

indicates the contribution of negative bags by a value of the 0.25th power of the negative components obtained by (4); and the y-axis of this graph indicates the contribution of positive bags by a value of the 0.25th power of the positive components obtained by (4).

Since the voting diverse density at a point equals to the difference of the value indicated by the y-axis and that indicated by the x-axis at that point, an instance located in the area above the dotted line in Fig. 3 is judged positive; while that located in the area below the dotted line is judged negative.

Note the instances in a positive bag (plotted with '*') and that in a negative bag (plotted with '·'). A significant amount (20.5%) of positive instances appear below the dotted line while most of instances at the optimal concept point (plotted with a small circle) in a positive bag appear above the dotted line. This is because the vdd score at the optimal point in a bag, according to its definition, equals to the maximum score among instances in that bag. For the same reason, a relatively large amount of instances at the optimal concept point (plotted with small rectangles) in a negative bag appear above the dotted line. This is the cause of false positive bags.

6 Conclusion

In this paper, we addressed the issue of PPI type prediction, and pointed out the problem that it is difficult for a standard supervised learning method to handle this issue because the relationship between the input variables (annotations for subunits) and the target variable (PPI type) is ambiguous.

We introduced the subunit reaction assumption and proposed that the PPI type prediction task based on this assumption can be formalized as a problem of MIL as follows: a complex pair with an interaction type is formulated as a labeled bag, and a possible subunit pair across a complex pair as an instance.

To solve that MIL problem, we proposed a method, a type of weighted voting system, based on Maron's Diverse Density (Maron and Lozano-Pérez, 1998), and evaluated it based on a binary classification version of that problem.

In the experimental evaluation, we have constructed a dataset, consisting of 1,279 different PPI records, from the KEGG pathways (Kanehisa and Goto, 2000), and a feature space for instances using the Gene Ontology (GO) (Ashburner et al., 2000). We then applied the method to a binary classification task concerning phosphorylation and achieved a highest accuracy of 86.1% in leave-one-out cross-validation when using the four major topmost components of the feature space obtained by SVD.

In the future, we would like to apply this method to the tasks of predicting other interaction types. In addition, we could improve the method by introducing a process to discriminate whether the protein group comprising a PPI record corresponds to a protein complex or a protein family, because the current version of the algorithm ignored those different types of protein groups, which the KEGG pathways may include in the same format. It will also be a future issue to improve the feature space, for example, by constructing an appropriate combination of features, or by encoding the domain or structural information of a protein, or relational information across a protein pair.

Acknowledgement

The authors would like to thank Kentaro Doi of Fujitsu Limited for the very useful discussions on potential applications of this research.

References

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler1, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29. [http://www.geneontology.org/].

Dietterich, Thomas G., Richard H. Lathrop, and Tomas Lozano-Perez. 1997. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1–2):31–71.

Ekins, Sean, Yuri Nikolsky, and Tatiana Nikolskaya. 2005. Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends in Pharmacological Sciences*, 26(4):202–209.

Kanehisa, Minoru and Susumu Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30. [http://www.genome.jp/kegg/].

Lee, Min Su, Seung-Soo Park, and Min Kyung Kim. 2005. A protein interaction verification system based on a neural network algorithm. In *CSB Workshops*, pages 151–154. IEEE Computer Society.

Maron, Oded and Tomás Lozano-Pérez. 1998. A framework for multiple-instance learning. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press.

Peri, Suraj, J. Daniel Navarro, Ramars Amanchy, Troels Z. Kristiansen, Chandra Kiran Jonnalagadda, Vineeth Surendranath, Vidya Niranjan, Babylakshmi Muthusamy, T.K.B. Gandhi, Mads Gronborg, Nieves Ibarrola, Nandan Deshpande, K. Shanker, H.N. Shivashankar, B.P. Rashmi, M.A. Ramya, Zhixing Zhao, K.N. Chandrika, N. Padma, H.C. Harsha, A.J. Yatish, M.P. Kavitha, Minal Menezes, Dipanwita Roy Choudhury, Shubha Suresh, Neelanjana Ghosh, R. Saravana, Sreenath Chandran, Subhalakshmi Krishna, Mary Joy, Sanjeev K. Anand, V. Madavan, Ansamma Joseph, Guang W. Wong, William P. Schiemann, Stefan N. Constantinescu, Lily Huang, Roya Khosravi-Far, Hanno Steen, Muneesh Tewari, Saghi Ghaffari, Gerard C. Blobe, Chi V. Dang, Joe G.N. Garcia, Jonathan Pevsner, Ole N. Jensen, Peter Roepstorff, Krishna S. Deshpande, Arul M. Chinnaiyan, Ada Hamosh, Aravinda Chakravarti, and Akhilesh Pandey. 2003. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Research*, 13(10):2363–2371.

Rhodes, Daniel R, Scott A Tomlins, Sooryanarayana Varambally, Vasudeva Mahavisno, Terrence Barrette, Shanker Kalyana-Sundaram, Debashis Ghosh, Akhilesh Pandey, and Arul M Chinnaiyan1. 2005. Probabilistic model of the human protein-protein interaction network. *Nature Biotechnology*, 23(8):951–959.

Yang, Jun, Rong Yan, and Alexander G. Hauptmann. 2005. Multiple instance learning for labeling faces in broadcasting news video. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 31–40.

Zhou, Zhi-Hua, Kai Jiang, and Ming Li. 2005. Multi-instance learning based web mining. *Applied Intelligence*, 22(2):135–147.

Lattice for Musical Structure and Its Arithmetics

Keiji Hirata[1] and Satoshi Tojo[2]

[1] NTT Communication Science Laboratories,
2-4, Hikaridai, Seika-cho, "Keihanna Science City," Kyoto 619-0237, Japan
hirata@brl.ntt.co.jp
http://www.brl.ntt.co.jp/people/hirata/
[2] Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Abstract. In this paper, we describe and discuss problems in terms of the music formalization framework that we are developing. Currently, we aim at building a contents design framework for ordinary people in a formal approach. Based on the knowledge representation technology, an art piece and contents are represented by a feature structure and the high-level operations upon the feature structure are available. When we apply the framework to implementing a practical musical system, we face some problems related to algebra, knowledge representation, and/or implementation. These are the issues concerned with: a relative pseudo-complement introduced for increasing descriptive power, an abstraction level of the distance between musical fragments, and a formalization of musical operations in addition to that of musical relations.

1 Introduction

We deploy our research in the framework of "design with cases" so that ordinary people who are not design experts can smoothly engage in design activities. We aim that in the framework, people will be able to start with already existing design cases, easily issue instructions about design, and revise them.

To give design with cases a theoretical basis, we first attempted to formalize operations and relations for revising a musical fragment as a starting point. We proposed a methodology for extracting musical structures from a musical fragment and representing the musical fragment as a tree structure[2], based on music theory GTTM[5]. Every note in a tree structure is assigned the degree of importance, according to GTTM. Then, for a tree structure given, we could extract an abstract structure through the tree structure itself, using the degree of importance as a threshold value.

There is a partial order between the two tree structures derived from the same tree structure, called a subsumption relation. From the music point of view, it means that several simpler musical fragments with a fravor similar to the original musical fragment can be produced by changing a threshold value. By introducing the partial order, we constructed a lattice, the elements of which are

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 54–64, 2007.

the tree structures of musical fragments, called lattice for musical structure[3]. Then, a predicate for a subsumption relation, meet (least upper bound), and join (greatest lower bound) were only available.

Moreover, to increase the descriptive power of the framework, we introduced a relative pseudo-complement[3]. We demonstrated that the combination of basic operations, such as meet, joint, and a relative pseud-complement, could realize melody manipulation to some extent. Note that since a musical fragment is represented as an AVM (attribute-value matrices), lattice operations, including a predicate for a subsumption relation, meet, join, and a relative pseud-complement, are implemented as mathematical algorithms.

However, when we build a music system based on our framework we face the following difficulties:

- The value of the relative pseud-complement introduced is often musically meaningless.
- In representing a musical fragment, there are different sensitivities for a distance between musical fragments.
- A procedual aspect of operations on musical fragments is not formalized in the current framework.

In the paper, we examine these difficulties and discuss how to solve them.

2 Relative Pseudo-complement and Heyting Calculus

In this section, for preparation, we provide fundamentals of mathematical notions used in this study.

2.1 Algebra and Lattice

Ordered set. When a binary relation $\sqsubseteq$ in a set satisfies:

- $a \sqsubseteq a$ (*reflexive*), and
- if $a \sqsubseteq b$ and $b \sqsubseteq c$ then $a \sqsubseteq c$ (*transitive*),

the set is called *preordered*. In addition, if

- $a \sqsubseteq b$ and $b \sqsubseteq a$ implies $a = b$ (*anti-symmetric*)

the set is *partially ordered* and is called a *poset*. If the relation is irreflexive ($a \not\sqsubseteq a$) but transitive, the set is called *strictly partially ordered* [4].

For a strictly partially ordered set, if the relation is

- either $a \sqsubseteq b$, $b \sqsubseteq a$, or $a = b$ (*connected*),

then the set is *totally ordered*, or *linear*.

Lattice. The tuple $\langle L, \cap, \cup \rangle$ is a *lattice* if:

- $a \cap a = a, a \cup a = a$,
- $a \cap (b \cap c) = (a \cap b) \cap c,\ a \cup (b \cup c) = (a \cup b) \cup c$,
- $a \cap b = b \cap a,\ a \cup b = b \cup a$, and
- $a \cap (a \cup b) = a,\ a \cup (a \cap b) = a$.

For any two elements in a poset $\langle L, \sqsubseteq \rangle$, if there exist the least upper bound $\sqcup$ the greatest lower bound $\sqcap$, i.e.,

- $a \sqcap b \sqsubseteq a,\ a \sqcap b \sqsubseteq b$,
- if $c \sqsubseteq a$ and $c \sqsubseteq b$ then $c \sqsubseteq a \sqcap b$,
- $a \sqsubseteq a \sqcup b,\ b \sqsubseteq a \sqcup b$, and
- if $a \sqsubseteq c$ and $b \sqsubseteq c$ then $a \sqcup b \sqsubseteq c$,

$\langle L, \sqcap, \sqcup \rangle$ becomes a lattice.

Distributive lattice. If a lattice satisfies

- $a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c)$ and
- $a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c)$,

it is *distributive.*

Those in Figure 1 are not distributive. The lattice in the left-hand side of Figure 1 is called a *pentagon* while the one in the right-hand side a *diamond.* If any part of a lattice is neither a pentagon nor a diamond, the lattice is distributive.

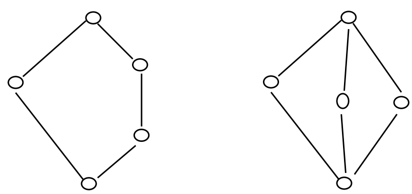


Fig. 1. Non-distributive lattices

Bounded lattice. When a lattice includes the top-most element $\top$ and the bottom-most element $\bot$, the tuple $\langle L, \sqcap, \sqcup, \top, \bot \rangle$ is a *bounded* lattice where $a \sqcup \bot = a$, $a \sqcup \top = \top$, $a \sqcap \bot = \bot$, and $a \sqcap \top = a$.

Complement. An element a^c is called a *complement* of a if $a \sqcap a^c = \bot,\ a \sqcup a^c = \top$.

- $a^{cc} = a$,
- $\bot^c = \top,\ \top^c = \bot$,
- $a \sqsubseteq b$ implies $b^c \sqsubseteq a^c$, and
- $(a \sqcap b)^c = a^c \sqcup b^c,\ (a \sqcup b)^c = a^c \sqcap b^c$.

Boolean algebra. If any element in a bounded lattice has its complement, the lattice is called a *Boolean algebra.*

2.2 Relative Pseudo-complement

For two elements a, b in a lattice L, such a maximal x that satisfies $a \sqcap x \sqsubseteq b$ is called a *relative pseudo-complement* of a with regard to b, denoted by $a \supset b$.

In general, if we define '$\supset$' by $a \supset b \equiv a^c \sqcup b$,

$$a \sqcap a^c \sqsubseteq b \iff a^c \sqsubseteq (a \supset b)$$

and thus the complement a^c also becomes the relative pseudo-complement of a with regard to b.

For any two elements in a lattice L, if there exists a unique relative pseudo-complement then the lattice is called a *relative pseudo-complementary lattice*. If a lattice L is relative pseudo-complementary there exists the maximum element in L. On the contrary, if L has the maximum element and is partially ordered, L is relative pseudo-complementary. In addition, a relative pseudo-complementary lattice is distributive.

2.3 Heyting Calculus

When a relative pseudo-complementary lattice has the minimum element, it is called a *Heyting Algebra*, or *Pseudo-Boolean Algebra* where $\langle L, \sqcap, \sqcup, \supset, \top, \bot \rangle$ satisfies all the prerequisites of Boolean algebra, except $a \sqcup a^c = \top$ where $a^c \equiv a \sqsupset \bot$. In a Heyting algebra, $a \sqcup a^c \sqsubseteq \top$.

Heyting calculus becomes a semantics of *intuitionistic logic*. Suppose a set $\mathcal{W}$ of possible worlds, in which there is the *hereditary* relation '$\preceq$', as

$$\forall \varphi [w \models \varphi \rightarrow w' \models \varphi] \text{ iff } w \preceq w'.$$

A set of such a subset U of $\mathcal{W}$ that satisfies

$$w \in U \text{ and } w \preceq w' \text{ implies } w' \in U$$

form a Heyting calculus. In the left-hand side of Figure 2, given $\{w_1 \preceq w_2, w_1 \preceq w_3, w_3 \preceq w_4\}$,

$$\{\emptyset, \{w_2\}, \{w_4\}, \{w_2, w_4\}, \{w_3, w_4\}, \{w_2, w_3, w_4\}, \{w_1, w_2, w_3, w_4\}\}$$

becomes a Heyting calculus as in the right-hand side of Figure 2.

3 Calculation of Relative Pseudo-complement

A relative pseudo-complement can be considered as a negative number to a positive number in arithmetic, and introducing a relative pseudo-complement is expected to enhance the descriptive power of an algebraic system[3].

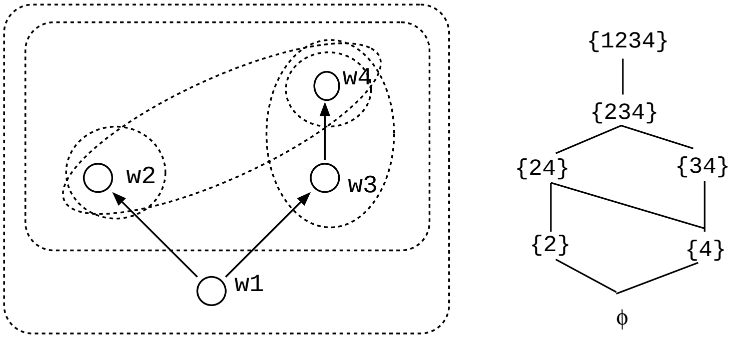


Fig. 2. Semantics of Heyting Calculus

3.1 Efficient Algorithm

Calculating a relative pseudo-complement requires the operations that collect elements satisfying a condition from a relevant domain and identify the maximal element among them (Section 2.2). Therefore, a naive implementation for calculating a relative pseudo-complement is not efficient because of exhaustive search within the domain (in our case, a database of musical fragments). Then, we present an efficient algorithm for calculating a relative pseudo-complement $a \supset b$ in a finite distributive lattice if $a \not\sqsubseteq b$ and $a \not\sqsupseteq b$ hold:

Step 1: Collect all elements x_i $(x_1, \cdots, x_n)$ satisfying $a \sqcap (b \sqcup x_i) \sqsubseteq b$.
Step 2: $a \supset b = b \sqcup x_1 \sqcup \cdots \sqcup x_n$

That is, we can come to the same answer if the search uses limited x (i.e., $b \sqcup x_i$), instead of x itself.

3.2 Value of Relative Pseudo-complement

The above calculation of a relative pseudo-complement $a \supset b$ equivalently includes the step for seeking the maximal element x. In a straightforward implementation of the efficient algorithm, to calculate the maximal element x, a computer first seeks all elements x_i's independent from a and b in a database of music fragments ($a \not\sqsubseteq x_i \wedge a \not\sqsupseteq x_i$, $b \not\sqsubseteq x_i \wedge b \not\sqsupseteq x_i$), repeating the join operataions with x_i's. Accordingly, the value of a relative pseudo-complement calculuated by the efficient algorithm unlimitedly approaches to the top-most element $\top$, which looks like almost all tree structures overlapped to each other. As a result, the calculated value of the maximal x is affected by x_i's independent from a and b in the database. From the music point of view, an element near to $\top$ corresponds to a musical fragment that densely contains almost all notes[1].

For instance, we suppose a musical fragment of 8 bar long with a resolution of 16th note, an element near to $\top$ means a musical fragment, in which 16th

[1] In fact, if the join operation is applied even only once, the resulting value may contain more or less musically unnecessary notes.

notes occur at every 16th beat throughout the 8 bars. At every 16 beat, notes of all pitches are overlapped vertically, and all possible tree structures are folded. Such note sequence is probably not musically meaningful.

Then, we consider two methods for restricting the value of a relative pseudo-complement so that the value is musically meaningful:

(1) restricting the process of seeking x_i's independent from a and b, and
(2) restricting the tree structure of an element near to $\top$.

In restriction (1), the algorithm seeks x_i's that are not only independent from a and b but also musically consistent with b. Thus, it is expected that the result of the join operation $b \sqcup x_1 \sqcup \cdots \sqcup x_n$ increases not so large as the result becomes musically meaningless.

Next, we discuss restriction (2). For instance, we first consider the depth of a tree structure as a restriction. Suppose $\top_n$ is the maximal elements of depth $\leq n$. Here, $\top_0$ means the tree structure of depth 0; that is, it does not contain any note, and is the same as the bottom-most element $\perp$. Then, we have the following ordering:

$$\perp \equiv \top_0 \sqsubseteq \top_1 \sqsubseteq \cdots \top_n \sqsubseteq \cdots \sqsubseteq \top_\infty \equiv \top.$$

Since we can use other measures other than the depth, such as the time span of a musical fragment, and the number of notes, there are various $\top_n$'s. For instance, if we use the number of notes as the measure, $\top_1$ means a maximal element among the musical fragments made of a single note. Thus, it follows that even $\top_1$ contains all notes, but instead, an element near to $\top_1$ is probably musically meaningful. It is viewed that $\top_n$ has the property of multiple-peak function, and we should properly choose $\top_n$ with an appropriate measure, depending on a situation.

4 Abstraction of Distance Between Musical Fragments

4.1 Definitions of Meet and Join Operations

Suppose two notes C3 and C4 are represented in AVM using our framework. Then we can have possible values as results of operations $C3 \sqcap C4$ (meet, common part) and $C3 \sqcup C4$ (join, integration), showed in Table 1.Here, abstraction level means how sensitive to the distance between two musical fragments the results of the meet and join operations are; abstraction level 1 corresponds to the most sensitive. At level 1, C3 and C4 are interpreted totally different from each other. At level 2, C3 is interpreted as a pair of pitch class C and octave position 3. At level 3, we suppose that for instance, relation $3 \sqsubseteq 4$ is added, in terms of octave position. At level 4, special values are introduced, for instance, "3 or 4" defined as $3 \sqcap 4$, denoted as $3|4$, and "3 and 4 at the same time" defined as $3 \sqcup 4$, denoted as [34].

The meet and join operations in Table1 all mathematically work well. In case of the most sensitive to the distance between musical fragments (abstraction

Table 1. Results of Meet and Join Operations of C3 and C4

Abstraction Level	1	2	3	4
meet	$\perp$	C	C3	C3|4
join	$\top$	C$\top$†	C4	C[34]

† Occasionally, a term containing $\top$ as a part is interpreted as $\top$.

level 1), the resulting value of the meet operation of two musical fragments may be the empty $\perp$ or the value close to the empty, as long as the two musical fragments are not quite similar to each other. The meet operation at level 1 is easy to implement, while it is less often to satisfy the applicable conditions of, for instance, the direct proportion algorithm, $b \sqcap x \neq \perp$ (Section 4.2). From a practical point of view, level 1 is in general too specific.

On the other hand, in case of the least sensitive to the distance between musical fragments (abstraction level 4), it is relatively easy to satisfy the above applicable conditions of the direct proportion algorithm, while we face the other problems as follows:

(1) It is difficult to give consistent definitions of the meet and join operations that treat tree structures containing [34] or 3|4.
(2) It is difficult to give consistent transformation from tree structures containing [34] or 3|4 to real sound (called instantiation).

For an example of (1), we do not know wheter or not we can define the value of the meet operation "[34] $\sqcap$ 3|4" consistent with the join operations. If we furthermore take into account the abstraction level in the meet operation, the problem becomes more complicated.

As for (2), the resulting tree structure of the meet operation may contain undefined values or special values (e.g., 3|4 and [34]) at the positions of, for instance, pitch class, octave, and timing. When the resulting tree structure is instantiated to generate real sound, undefined values should be assigned to actual pitch class, octave, duration and onset timing, and special values, such as [34], should be assigned to an actual note extracted from 3 being overlapped with 4. We can think of several methods but do not know which is appropriate when.

4.2 Applicable Domain of Direct Proportion Algorithm

We proposed a convenient operation using a relative pseudo-complement (Section 2), direct proportion algorithm[3]. Given elements a, b and x, the direct proportion algorithm calculates y such that $a : b = x : y$, which is defined by

$$\mathcal{A}_{ab}(x) = a \supset (b \sqcap x).$$

Suppose that elements a and b represent two musical fragments, respectively. Since the direct proportion algorithm can be regarded as the transformation of unknown input x so that a is transformed into b, it can be used for the arrangement algorithm of a unknown musical fragment, imitating an already existing arrangement.

The direct proportion algorithm was constructed as follows. First, we identify the multiplication used in the relation of direct proportion $a : b = x : y$ with the meet operation, and hence, have $a \sqcap y = b \sqcap x$. To obtain the value of y satisfying $a \sqcap y = b \sqcap x$, we give y a small number as an initial value, and then, increase the value of y with satisfying $a \sqcap y \sqsubseteq b \sqcap x$. These steps totally correspond to the calculation of relative pseudo-complement $a \supset (b \sqcap x)$.

However, the applicable domain of $\mathcal{A}_{ab}(x)$ is not broad because we have to satisfy

– $b \sqcap x \neq \bot$
– $a \sqsubseteq b \sqcap x$

to make expression $a \sqcap y = b \sqcap x$ meaningful. When we take into account the divergence of actual musical fragments, even at abstraction level 4, it will be not often that the above conditions are satisfied. In contrast, for musical fragments of small divergence, the above conditions are relatively satisfied, and of course, the direct proportion algorithm will work well.

5 Formal Representation of Media Arrangement

5.1 Feature Structure as Media Representation

It is convenient for us to represent various artistic media in a formal, hierarchical way. We contend that we can represent even such a media by a feature structure, that is a list of feature-value pairs where a value may be replaced for another feature structure recursively. Below is a feature structure where capital F_i is a structure, lower case f_i is a feature label, and v_i is its value.

$$F_1 = \begin{bmatrix} f_1 & \begin{bmatrix} f_3 & v_3 \\ f_4 & v_4 \end{bmatrix} \\ f_2 & F_7 \end{bmatrix}$$

Such a structure is also called a labeled directed acyclic graph (DAG), and needless to say, we can regard the structure is equivalent to a tree, regarding each node of which corresponds to a feature and a leaf to a value.

Two feature structures are in the subsumption relation, or in other words, one is the hyper-structure of the other when one structure properly includes the other from the same top node. For example, F_1 is subsumed both by the following F_2 and F_3, denoted by $F_1 \sqsubseteq F_2$ and $F_1 \sqsubseteq F_3$.

$$F_2 = \begin{bmatrix} f_1 & \begin{bmatrix} f_3 & v_3 \\ f_4 & [f_5 \quad v_5] \end{bmatrix} \\ f_2 & F_7 \end{bmatrix}, \qquad F_3 = \begin{bmatrix} f_1 & \begin{bmatrix} f_3 & v_3 \\ f_4 & v_4 \end{bmatrix} \\ f_2 & F_7 \\ f_6 & v_6 \end{bmatrix}$$

Although both of F_2 and F_3 are elaborations of F_1, the two hyper-structures are differently elaborated. Thus, for the common structure we can define the '$\sqcap$' (meet) operation, as $F_1 = F_2 \sqcap F_3$ between F_2 and F_3. Similarly, we can define '$\sqcup$' (join) operation between F_2 and F_3 that is the unification with consistent feature-values.

$$F_2 \sqcup F_3 = \begin{bmatrix} f_1 \begin{bmatrix} f_3 \; v_3 & \\ f_4 \; [f_5 \quad v_5] \end{bmatrix} \\ f_2 \; F_7 \\ f_6 \; v_6 \end{bmatrix}$$

In order to denote a value v of feature f in structure F, we write $f(F) = v$. Thus, $f_2(F_1) = F_7$ while $f_4(F_1)$ is undefined. To access by a sequence of features, we connect feature labels by '$|$', as $f_1|f_4(F_1) = v_4$. As far as no ambiguity, we omit the long access $f_1|\cdots|f_n$ as $f_1\|f_n$, e.g., $f_1\|f_5(F_2) = v_5$.

5.2 Representation of Time-Span Tree

In this section, we employ *head* [7] notation; the head of a set is the representative among the members of the set and is denoted by '$\bar{\ }$' (bar). Thus, a time-span tree of a sequence of notes $\{e_1, e_2, e_3, e_4\}$ is represented as in Figure 3 where e_3 is most important among $\{e_3, e_4\}$ and $\{\overline{e_3}\}$ is most among $\{\overline{e_1}, \overline{e_3}\}$.

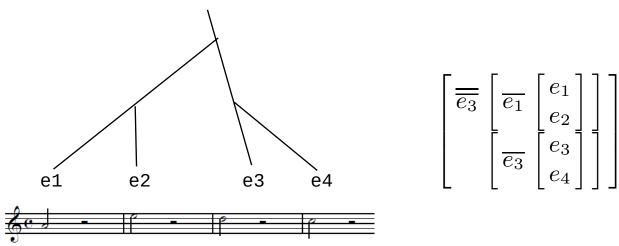


Fig. 3. Time-span tree

5.3 Arrangement by Tensor Product

As a piece of music is a sequence of notes and each of which has the same feature set (though different values), we can reduce the representation of a feature structure by a product of vectors. Suppose that a music piece consists of a sequence of notes $\boldsymbol{e} = (e_1, e_2, \cdots, e_n)$ and each note owns a set of attributes as pitch, time value, possibly accent, and so on, represented by $\boldsymbol{a} = (a_1, a_2, \cdots, a_m)$. Then, the whole structure becomes:

$$\boldsymbol{e} \otimes \boldsymbol{a},$$

where '$\otimes$' is the tensor product that is a matrix of all the combinations of e_i's and a_j's.

Here, let F be an arrangement for the sequence of notes and G for the modification of each note. Thanks to the feature of the tensor product, the combinative arrangement $F \otimes G$ can be applied to $\boldsymbol{e} \otimes \boldsymbol{a}$, and the target of the operation is dissolved, as follows.

$$(F \otimes G)(\boldsymbol{e} \otimes \boldsymbol{a}) = F(\boldsymbol{e}) \otimes G(\boldsymbol{a}).$$

Example 1 (Passacaglia). *Passacaglia is a kind of variation; the base melody of the three-part time is repeated as* basso ostinato*, on which various soprano melodies are added. Historically, BWV582 of J. S. Bach, the fourth movement of the fourth symphony by Brahms, op. 1 of Webern, and so on are known as masterpieces. If we regard* m_0 *the basso ostinato and* m_i *is its variation,* $m_0 \sqsubseteq m_i$ *and various variations form a lattice. The bottom melody of Figure 4 is the basso ostinato of BWV 582, above which two different elaborations are shown. In this case, there is no change in the order of the original notes, but each note is elaborated, as*

$$(I \otimes G)(\boldsymbol{e} \otimes \boldsymbol{a}) = \boldsymbol{e} \otimes G(\boldsymbol{a}).$$

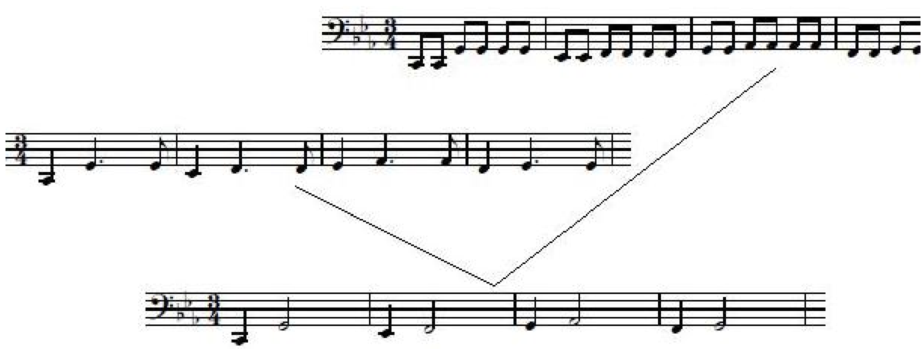


Fig. 4. Basso ostinato of BWV582 and its variations

Example 2 (Dodecaphony). *In dodecaphony, or more generally, a music of serie, the basic melody line can be reversed, inversed, or enlarged. Figure 5 is the inversed melody of the basso ostinato of BWV582. The reversion operation* G_r *and the inversion operation* G_i *form a* group*, together with the* no *operation (identity mapping). On the other hand, the enlargement* G_l *affects time values of notes, as*

$$(G_r \otimes G_l)(\boldsymbol{e} \otimes \boldsymbol{a}) = G_r(\boldsymbol{e}) \otimes G_l(\boldsymbol{a}).$$

Considering the contraction G_c*, as opposed to the enlargement, we can again consider a group of* G_l *and* G_c *with the identity operation. We can define a sequence of operations as*

$$G_i G_j \cdots G_k(\boldsymbol{e} \otimes \boldsymbol{a}),$$

and in addition, the inverse function becomes

$$(G_i G_j \cdots G_k)^{-1} = G_k^{-1} \cdots G_j^{-1} G_i^{-1}.$$

Fig. 5. An inversion of basso ostinato of BWV582

6 Concluding Remarks

We think that it is difficult to find an general answer to the problems described in the paper. Probably, for each design support system treating media, such as music and painting, a different solution should be developed practically. However, the framework we develop may provide common understanding and broad applicability of design with cases. In future, we hope to investigate design with cases across media from both theoretical and practical aspects.

References

1. Barwise, J. and Seligman, J.: Information Flow - The Logic of Distributed Systems -. Cambridge University Press (1997)
2. Hirata, K. and Aoyagi, T.: Computational Music Representation based on the Generative Theory of Tonal Music and the Deductive Object-Oriented Database. *Computer Music Journal* 27:3 The MIT Press (2003) 73–89
3. Hirata, K. and Tojo, S.: Formalization of Media Design Operations Using Relative Pseudo-Complement. 19th Annual Convention of Japan Society of Artificial Intelligence 2B3-08 (2005), In Japanese
4. Landman, F.: Structures for Semantics. Kluwer Acedemic Press (1991)
5. Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*. The MIT Press (1983)
6. Sadie, S. (ed): The New Grove Dictionary of Music and Musicians. Macmillan ublishers Ltd (1980).
7. Pollard, C. and Sag, I.A.: Head-Driven Phrase Structure Grammar. The University of Chicago Press (1994)

Viewlon: Visualizing Information on Semantic Sensor Network

Masayuki Furuyama, Jun Mukai, and Michita Imai

Faculty of Science and Technology, Keio University, Hiyoshi 3-14-1, Kouhokuku, Yokohamashi, Kanagawa, 223-8522 Japan
{furuyama,mukai,michita}@ayu.ics.keio.ac.jp
http://www.ayu.ics.keio.ac.jp

Abstract. The topic of getting information from sensor networks is becoming more important. Middleware will be required to manage various sensor data in future. We propose the Semantic Sensor Network (SS), which applies an ontology to sensor data and enables such data to be treated more easily. However, users have great difficulty to understand the information on the SS because its semantic data are complicatedly related with each other and the data and their relations evolve in time. The requirements for visualizing such information are; 1) the SS must give users the information appropriately, and 2) the SS must answer the users' queries. To meet these requirements, we propose Viewlon, software for visualizing information on the SS. To give the information appropriately, information display levels and a timeline view are introduced. In addition the functionality of processing queries provides the required information to users interactively. We experimented with Viewlon in a simulation environment, and the results show that Viewlon has the ability to present information to users and to answer their queries correctly.

Keywords: Sensor Network, Visualization, Application, Human Interface.

1 Introduction

It has become popular to construct sensor networks using many kinds of sensors and to get various information from the real world. In the future, middleware will be required to manage various sensor data universally. We propose the Semantic Sensor Network(SS), middleware for universally describing and managing sensor data by applying an ontology including class abstractions and inference rules to sensor data. Using class abstractions enables us to manage many kinds of objects in the real world. The SS can provide not only sensor data which are represented by mere numerical values, but also the data which are highly abstracted by applying inference rules, which are defined by a combination of classes and sensors, to sensor data. Users can handle such data more easily than data that are represented by only numerical values.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 65–76, 2007.

However, the relationship between pieces of information on the SS is too complex to provide to users directly because the SS has many kinds of informations such as classes and inference rules, besides sensor data. Therefore, it is necessary to visualize the information on the SS so that users can easily understand it. There are two problems in visualizing the information on the SS. First, we must represent information about sensors, instances, and classes universally, including temporal changes in logical relations among instances. Second, it is necessary not only to give users information but also to answer their queries. To resolve these problems, we propose Viewlon, software for visualizing information on the SS.

2 Semantic Sensor Network

An SS consists of classes, instances, sensors, and inference rules. Figure.1 shows the structure of an SS. Our purpose is visualizing the elements of the SS and relations among them. We describe the individual elements of SS and their relations in the following sections.

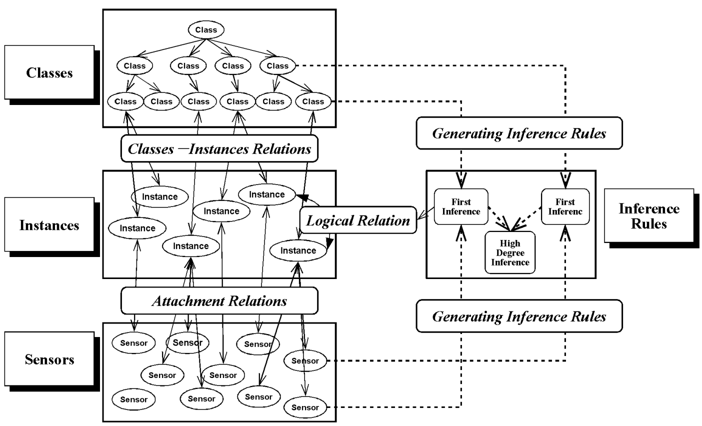


Fig. 1. The structure of the Semantic Sensor Network(SS). The SS consists of classes, instances, sensors, and inference rules.

2.1 Classes

Classes in the SS are defined as abstract templates of objects that exist in the real world. A class defines attributes that describe common information of the objects belonging to a class. By defining inheritance relations among classes, the user of the SS can easily handle attributes and create inference rules. A class tree is a representation of the inheritance relations among classes.

2.2 Instances

An object which exists in the real world is handled as an instance to which sensors are attached and which substantiates a class. The concrete contents of attributes of an instance are called metadata. The metadata are given when the instance is generated. MeT [5] is a system for universally interpretating of sensor data by altarching metadata to this data.

2.3 Sensors

Sensors get the information about the instance (the object in the real world) when they are attached to it. There are many types of sensor, for example, ultrasonic location sensors, which can detect the positions of object [1] (Fig. 2), and Mote, [2] which can detect light, accelerationa and temperature (Fig. 3).

Fig. 2. Ultrasonic location sensors

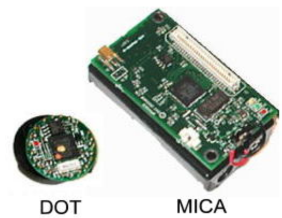


Fig. 3. Mote (Dot and Mica)

2.4 Inference Rules

Inference rules are used to get logical relations among instances from the sensor data. The logical relations of the SS evolve according to the dynamics of the sensor data.

3 Related Work

3.1 Visuali Zing Sensor Data

One of the studies on visualizing data in sensor networks is SpyGlass [4]. SpyGlass is a framework for visualizing sensor data and data traffic in the sensor network by using plugins. By visualizing sensor data, SpyGlass can help development using the sensor network. However, the design of SpyGlass focuses only on the sensor data, not on other information such as classes or inference rules.

SpyGlass simply visualizes sensor data gotten from the sensor network and is not applicable to visualizing information on the SS. Moreover, it does not have a functionality for processing queries.

3.2 Visualizing Information Related to Ontology

Studies on visualizing information based on an ontology are popular in the field of the Semantic Web [3]. Harmelen developed a system for visualizing the ontology on the Semantic Web and the relations among instances and presenting them to users, by representing elements as a set or a graph. The system of Harmelen's study has a function for visualizing answers to users' queries. Logical relations change rapidly because the subject of the SS is the real world. Therefore, temporal information is important for the SS to present. However, most visualizing studies in the field of the Semantic Web focus on not temporal changes but structural ones. Hence, it isn't enough to apply the visualizaion of the Semantic Web to the SS.

4 Viewlon

As we mentioned, visualizing information on the SS requires the following tasks: 1) visualizing the relations among sensors, instances, classes, and logical predicates, and 2) processing queries given by users interactively. However the existing works mentioned in section 3 do not meet these requirements. That task motivated us to develop Viewlon, software to visualize information on the SS.

4.1 System Design

Viewlon works with an SS server. The server manages sensor data, instances, and classes and receives information on the type of sensor and sensor data, information about the attachment between sensors and instances from the server. Viewlon connects to the server, receives the sensor information, and generates logical relations by using the received information and inference rules defined in it. Then, Viewlon visualizes and displays the received information and the generated logical relations.

4.2 Views and Panels Implemented in Viewlon

Various views and panels are implemented in Viewlon. The main view displays information on the SS. The timeline view visualizes temporal changes in the logical relation between two instances. The information panel displays sensor data and inference rules, and it is used for user to input queries. The class tree view displays the structure of the class tree. The inference rule manager panel is used to add and remove inference rules. The real world view simply displays the exact three-dimensional positions of instances calculated from the sensor data.We describe the main functions of Viewlon, i.e., information display levels, timeline view, and the functionality of processing queries in the following sections.

4.3 Information Display Levels

Viewlon visualizes attachment relations between sensors and instances, relations between instances and classes, and logical relations among instances in the main view. Logical relations are generated by applying inference rules to the sensor data. In the main view, an element such as a sensor, an instance, or a class is represented as a node of a disk, and relations among these elements are visualized as graphs or sets of nodes. A graph is drawn with arrows between nodes. However, if these three elements are visualized together on the same screen, the cognitive load increases in proportion to the amount of information. Therefore, it is necessary to separate important information from unimportant information.

Viewlon solves this problem by using the information display levels in the main view. Concretely, information display levels include six levels: Sensors Display Level, Relations among Sensors and Instances Display Level, Relations among Instances Display Level, Relations among Instances and Classes Display Level, Classes Display Level, and Class Tree Display Level. Figure. 4 shows an example of the Relations among Sensors and Instances Display Level. The individual sensor is represented as a node of a disk. Moreover nodes (sensors) that belong to the same instance are displayed as the same subset. The figure shows that arrows are drawn from the Masayuki , an instance of a Human class, to the instances that have the logical relation "Front" from the viewpoint of Masayuki. The changes between the information display levels are controlled by the slider bar on the bottom of the main view (Fig. 4). By using the information display levels, the complexity of information on tne SS can be limited, with only the required information represented.

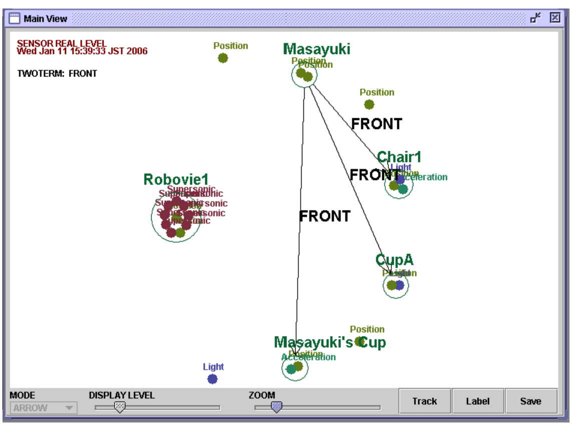


Fig. 4. Example of the main view (Relations among Sensors and Instances Display Level)

4.4 Timeline View

Because the logical relations changes according to change of environment of the real world, a function of visualizing the logical relations among instances in terms of time is necessary in the SS. From the standpoint of cognitive load, however, relations among sensors, instances, and classes and temporal changes of logical relations should not be visualized all together. Viewlon should separately visualize the temporal information from other information.

Viewlon provides a timeline view to solve this problem. Figure. 5 shows an example of a timeline view. A user selects two target instances for display and a logical relation he or she wants to watch. While the selected logical relation between the instances is satisfied, a line is drawn on the timeline view. In Fig. 5, a line is drawn while Chair 1, the instance of the Chair class, exists in front of Masayuki, the instance of the Human class. Visualizing the logical relations between a couple of instances in a temporal range enables users to understatnd the temporal changes in the logical relation.

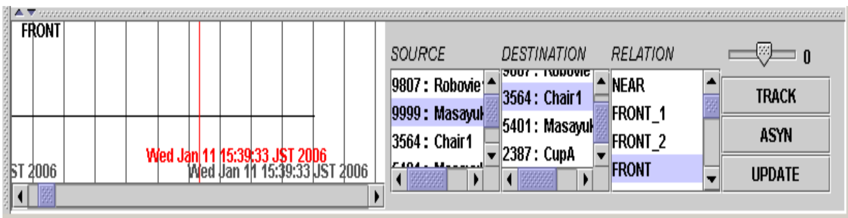


Fig. 5. Timeline view

4.5 Processing Queries

Viewlon has a function to display answers to queries given by users in the Relations among Instances Display Level of the main view. This function enables users to get information on the SS interactively. The user inputs query to the text box in the Information Display Panel, which is sent by the user pushing a button (Fig. 6). The user can input queries only if the information display level is set at the Relations among Instances Display Level. The query is saved, and the answer is displayed dynamically in case that the logical relations among instances change.

Viewlon handles queries about the logical relations among instances. The query is adopted as the form of the first-order predicate logic; supports an unary relation and a binary relation. A query for an unary relation is written in the form of *Relation-Name*(X). A query about the binary relation is written as the form of *Relation-Name*(A, X) or *Relation-Name*(X, A). “A” is the concrete name of the instance the user specifies, “X” represents the subset, which a user wants to know, theat satisfies the logical relation whose name is *Relation-Name*. The answer to the query, i.e., the instances satisfying the query, is displayed as a set of nodes in the main view.

We explain how query processing works with an example. The example is a case in which a user gives the logical relation “Front” which is classified as a binary relation. We assume that there are three instances A, B, and C. We also assume that B and C are in front of A, and A is in front of B. In this situation, when the user inputs a query “Front(A, X)”, the instances B and C are displayed as belonging to the same subset in the main view. When the user inputs “Front(X,A)”, only instance B is displayed as belonging to the subset.

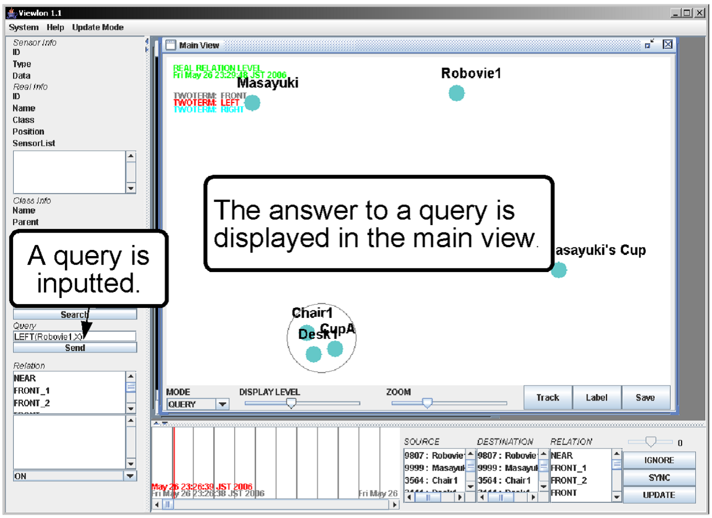


Fig. 6. The way of processing queries. A query is inputted in the text box, and the answer to the query is visualized in the main view.

5 Experiment

This section explains the experiments and their results on Viewlon. We carried out three experiments: a basic test of the functions of visualizing information on the SS, a query test of the functionality of processing queries, and a performance test based on convergence time of nodes in the main view. As for the basic test, we verified whether changes in position in the real world were correctly visualized in the main view and timeline view. At the query test, we verified whether answers to queries given by users were visualized correctly in an environment defined by simulation data. The performance test, comprised an experiment on the convergence time of nodes in terms of number of nodes in the main view and it mesured how many nodes Viewlon was able to visualize properly.

5.1 Experiment of Visualizing Functions

We attached 3D-ultrasonic sensors [1] (position sensors) to the objects in the real world and verified whether logical relations about the positions of the instances were correctly visualized in the main view. We also verified whether the temporal changes in the logical relations were correctly visualized in the timeline view. The objects used in the experiment were a book, a plastic bottle, and a communication robot Robovie [6]. The plastic bottle (BottleA) was an instance of Bottle class, the book (YellowBook) which was an instance of Book class, and Robovie (Robovie1) which was an instance of Robot class. By moving BottleA and YellowBook in front of Robovie1, we verify whether the instances with logical relation "Front" from the viewpoint of Robovie1 were correctly visualized, by representing the set of nodes at the Relations among Instances Display Level in the main view of Viewlon. From Figures. 7 to 9 show the results of the experiment.

When BottleA and YellowBook were in front of Robovie1, BottleA and YellowBook were in the same set representing "Front" in the main view of Viewlon (Fig. 7). After BottleA was taken away from the area in front of Robovie1, the set representing "Front" in the main view included only YellowBook (Fig. 8). Subsequently after YellowBook was taken away from the area in front of Robovie1, the set, which is represented as a circle in the main view, become empty (Fig. 9).

Next we tested the timeline view operations. When BottleA was in front of Robovie1, a line representing the logical relation of "Front" was drawn on the screen (the bottom left window of Fig. 7). After BottleA was taken away from the area in front of Robovie1, the line representing the logical relation "Front" was not drawn near the green bar representing the current time had passed(Fig. 8). After BottleA had been taken away from Robovie1, the line was no longer drawn in the timeline view (Fig. 9). This result shows that representation of the set of the nodes at the Relations among Instances Display Level correctly visualized the logical relations among instances. It also showed that the timeline view properly visualized the temporal changes in logical relations between instances of SS.

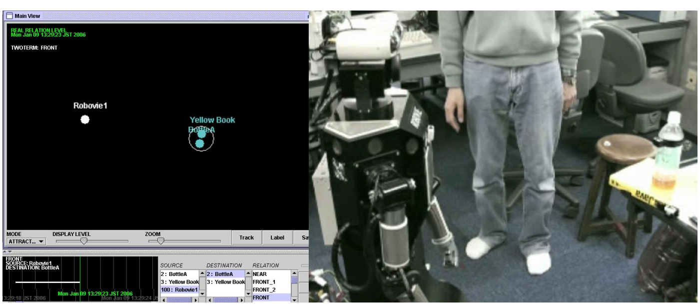


Fig. 7. The case BottleA and YellowBook are in front of Robovie1

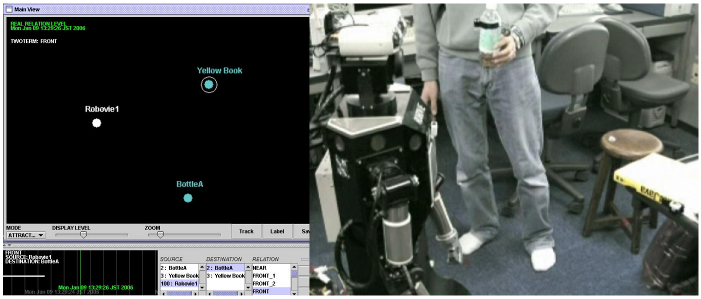


Fig. 8. The case in which BottleA was taken away from the area in the front of Robovie1

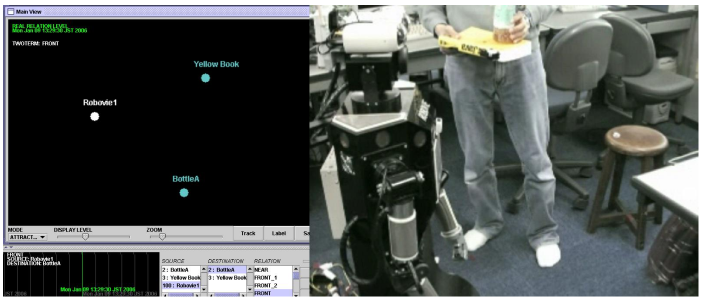


Fig. 9. The case BottleA and YellowBook were taken away

5.2 Experiment on Processing Queries

We tested the functionality of processing queries in an environment defined by simulation data. We assumed that there were six instances in the environment and all of them were attached to position sensors. Table 1 shows what class each instance belongs to. Figure. 10 shows the assumed environment.

In this environment, the user sent two queries to Viewlon, and both queries were about the "Left" relation. This relation refers to the spatial relation between two instances. For example, if "Left(A, B)" is satisfied, instance A is placed on the left of instance B. We checked if Viewlon could correctly visualize the proper information in response to the user's queries about this relation.

The answer to each query is displayed fusing the Instance Relation Level on the main view. Figures. 11 and 12 show the results displayed on the main view with the real world view corresponding to the main view. First, for the query "LEFT(Masayuki, X)", Viewlon displayed the set of nodes to the left of Masayuki in its views (Fig. 11). Next, for the query "LEFT(X, CupA)", Viewlon displayed the set of nodes that had the logical relation "Left" to CupA (Fig. 11).The experimental result showed that Viewlon correctly visualized the answers to the queries given by users.

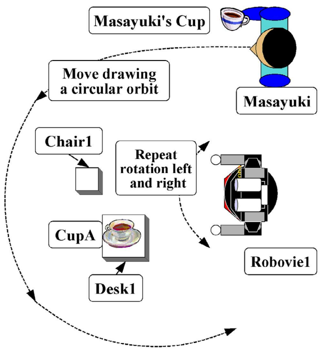


Fig. 10. Simulation environment

Table 1. Relation between class and instance

Instance name	Class name
Masayuki	Human
CupA	Cup
Masayuki's Cup	Cup
Chair1	Chair
Robovie1	Robot
Desk1	Desk

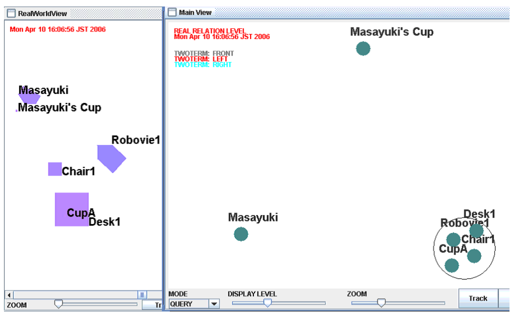


Fig. 11. Experimental result of giving the query LEFT(Masayuki, X)

5.3 Convergence Time of Nodes in Main View

In this section, we describe about the performance of representing information at main view. The performance was measured as the convergence time for switching information display levels from Sensors Display Level to Relations among Sensors and Instances Display Level. We assumed that five sensors were attached to one instance. The specifications of the personal computer used in the experiment were follows: CPU, Mobile Athlon XP 1700+ (clock is 1.46GHz); main memory, DDR 266, 320MB. The operating system was Windows XP.

The measured convergence time increases slowly in Fig. 13. The convergence time is slow while the number of instances is under 25, but increases sharply when the number of instances exceeds 40. The result shows that Viewlon can not visualize over 60 nodes. For practical purposes, the maximum number of instances used in Viewlon seems to be about 25 (125 sensors).

The assumption that five sensors were attached to one instance is a lot. It is enough for Viewlon to be able to visualize dozens of instances considering Viewlon in a room. In light of the above-mentioned points, the result shows that Viewlon can be used in common situations.

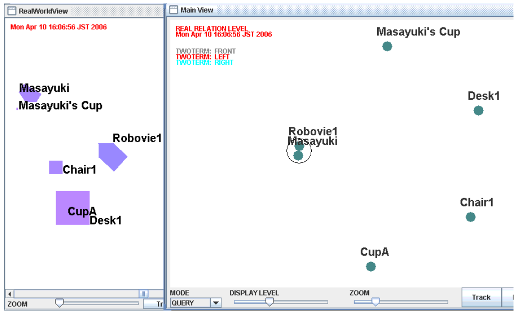


Fig. 12. Experimental result for the query LEFT(X, CupA)

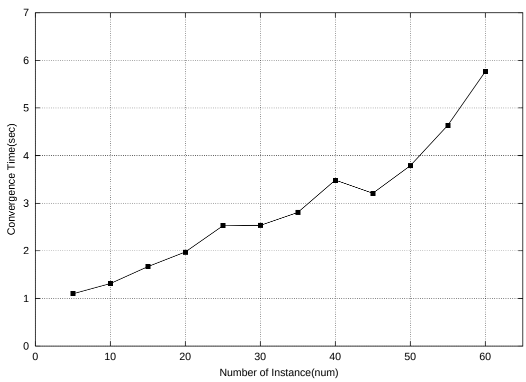


Fig. 13. Number of instances and convergence time

6 Conclusion

We described the design and implementation of Viewlon, which is a visualizing software for Semantic Sensor Network (SS). There are many complex relations in the SS, such as relations among sensors and instances to which they are attached, logical relations between instances, and relations between classes and instances. Therefore, if all such information were fully and directly displayed, the user would have difficulty understand them. Viewlon uses information display levels implemented on its main view and has a timeline view as means of clearly presenting information on the SS. Moreover, Viewlon's functionality of processing queries enabls users to get information on the SS they want or need. The experiment using the real data showed that the main view and timeline view were able to visualize information on the SS. The experiment using simulation data showed that the functionality of processing queries worked properly. Furthermore, the performance evaluation regarding the numbers of sensor nodes and instance nodes showed that Viewlon has enough performance to visualize information in typical situations on the use of SS. In future work, we plan to test

Viewlon by using various sensors in a more complex real world environment. We also plan to enrich the functionality of processing queries.

References

1. Y. Nishida, H. Aizawa, T. Hori, N. Hoffman, T. Kanada and M. Kakikura: 3D ultrasonic tagging system for observing human activity, Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS2003), pp.785-791, 2003.
2. Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong: ACM The Design of an Acquisitional Query Processor For Sensor Networks, SIGMOD Conference, pp.491-502, 2003.
3. Frank van Harmelen, Jeen Broekstra, Christiaan Fluit, Herko ter Horst, Arjohn Kampman, Jos van der Meer, Marta Sabou: Ontology-based Information Visualisation, Proceeding IEEE International Conference Information Visualization, pp.546-554, 2001.
4. Carsten Buschmann, Dennis Pfisterer, Stefan Fischer, Sánndor P. Fekete, Alexander Kröller: SpyGlass: A Wireless Sensor Network Visualizer, SIGBED Review, Vol.2, No.1, pp.1-6, 2005.
5. Yutaka Hirota, Hideyuki Kawashima, Satoru Satake, Takaehi Umezawa, Michita Imai: Design and Implementation of RealWorld Oriented Metadata Management System MeT for Semantic Sensor Network, Research Report of Computer and Education, No.109, Information Processing Society of Japan, 2005-ICS-141, Vol.2005, No.109, pp.29-36,2005(in Japanese).
6. Takayuki Kanda, Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, Ryohei Nakatsu: Development and Evaluation of an Interactive Humanoid RobotRobovie. IEEE International Conference on Robotics and Automation (ICRA2002), pp.1848-1855,2002.

Cooperative Task Achievement System Between Humans and Robots Based on Stochastic Memory Model of Spatial Environment

Tetsunari Inamura[1,2,3], Tomohiro Kawaji[3], Tomoyuki Sonoda[3], Kei Okada[3], and Masayuki Inaba[3]

[1] National Institute of Informatics, Japan
[2] Dept. of Informatics, The Graduate University for Advanced Studies, Japan
[3] Dept. of Mechano-Informatics, The University of Tokyo, Japan

Abstract. Design and acquisition method of environmental model are important issues for robots which act in the real world such as a daily life environment. Robots have to follow up changes of environment and reconstruct the model because of the conditions in the real world environment would be changed constantly. Also the model for unobserved part would be needed and should be inferred when users instruct the robot using symbolic expressions. In this paper, we have focused on stochastic representation of environmental memory to realize smooth communication between humans and robots, and realtime memory management with ambiguities in the real world. We also show that the representation is effective to construct a cooperative task achievement system on intelligent robots.

1 Introduction

Recently, cooperative robots which can communicate with humans and act as co-existing partners in daily-life environments, have been gaining a great deal of attention. These robots are desired to understand users' instructions in various complex situations and environments.

Kobayashi *et al* have proposed a communicable robot that can acquire novel motion pattern via communication between humans, and can join to group conversation with gesture behavior[1]. In several researches [2][3][4][5][6] other human-robot interaction system using natural language communication have been proposed. In these researches, user could request the robot to achieve some tasks such as assembly, bring something to the user and so on. However, these systems could not handle vague user's instructions and uncertain information of environment. As Winograd[7] said, world representation models, especially on using natural language processing systems, have no bounds. Therefore interactive systems have to acquire the environment information in front of its eyes, accumulate them as experiences, and exploit them in vague and uncertain situation.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 77–87, 2007.

We have already proposed an adaptive dialogue management method based on probabilistic information processing[8][9]. In this paper, we aim to let the robot to understand users' instruction with consideration of situation and context in the real-world environments those conditions constantly change. Geometric memory model and probabilistic management method are focused and proposed for the purpose.

2 Stochastic Memory Model of Spatial Environment

2.1 Spatial Memory Model for Environments and Objects

Probabilistic calculation is effective for modeling human computer/robot interaction. Wakamatsu *et al* have also proposed human-robot interaction framework based on probabilistic processing[10]. The work focused on how to control the interaction context based on emotion model by probabilistic calculation. Tanaka*et al*[11] and Stopp*et al*[12] have proposed a probabilistic method to plan behaviors via human-computer interaction. This method could accept spatial and positional expressions such as "Please bring the cup on the right side". probabilistic calculation covered the spatial relationship between agents and objects. However, temporal expression such as "Please bring what you were talking about to me" could not be understood, because following context of the dialogue is one of the untouched issue.

We aim to establish a new approach to probabilistic temporal-spatial model for human-robot interaction, because the potential of the probabilistic approach also effective for both the spatial and temporal expression. It is important that the environment representation model, that is memory model, is able to manage not only objects in eyesight of the robot, but also invisible objects especially referred in the bypast context. It is also needed that the environment model is useful to decide and plan the total behavior of the robot. We therefore define the object memory as a set of following items.

1. Existing probability
   probability distribution which means certainty degree that a target is exist certain place.
2. Probability variable of physical properties distribution
   probability distribution which means physical properties of objects such as color, size, location and so on.
3. 3D Geometric information
   Physical 3D information such as shape and weights, that are used for motion planning of the robot. Self-body information of robots is also includes in this.

2.2 Revision of Spatial Memory by Vision

The first step of the memory revision is acquisition of spatial objects information what the robot seeing by cameras. Applying color segmentation method to input

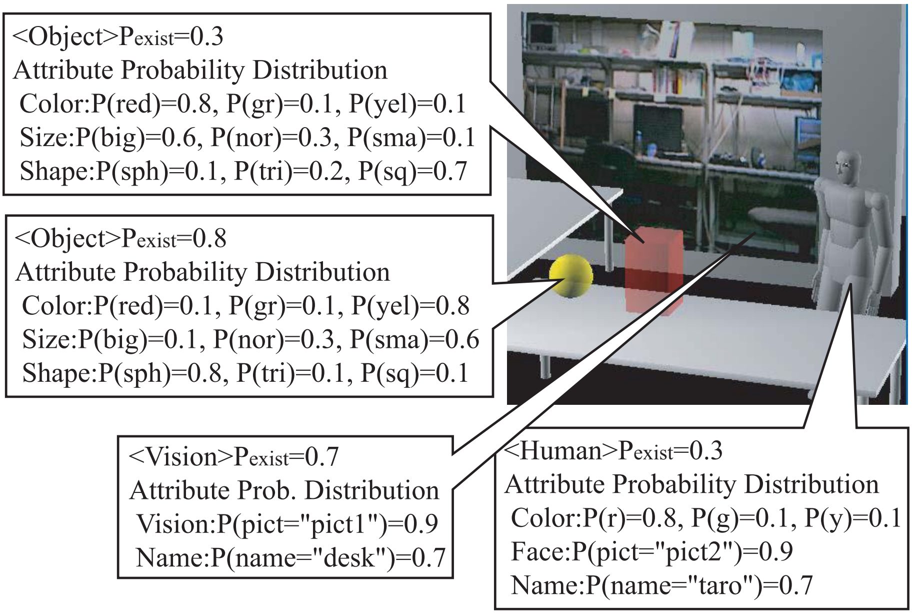


Fig. 1. An example of geometric memory model based on existing probability distribution for the physical properties

images via binocular cameras, several areas are generated. Shape, color, size and location information are acquired by visual processing for each area. A area is considered as an object, then visual features for each area are added to the memory as property of each object. In the measurement of object location by image matching processing for stereo vision, measurement errors always arise. When the error value was high, the system would judge the stereo process is not desirable and give lower 'degree of certainty' for the object. Figure 2 shows an example of memory constructing.

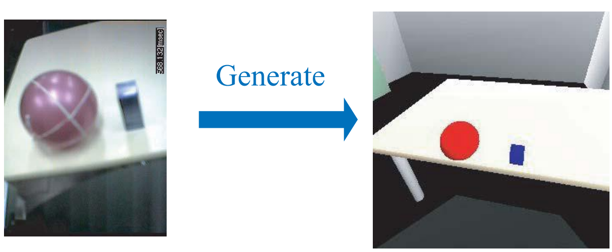


Fig. 2. Object model construction based on real image information

2.3 Reflection of Context for the Memory Model

Next, control of existing probabilities is introduced to consider the time effect of uncertainty. Basically, the existing probability values are reduced with the course of time as following Eq.(1).

$$p_{exist}(t+1) = p_{exist}(t) \cdot (1 - \omega(t) + \alpha \sum_{obj} R(self, obj)) \tag{1}$$

where, $\omega(t)$ is damping coefficient for memory at time $t(0 < \omega < 1)$, α is canonicalize coefficient, and $R(objA, objB)$ is coefficient for memory recalling between $objA$ and $objB$ which satisfies $(0 < R(objA, objB) < 1)$. The $R(objA, objB)$ defines the degree of relationship in context between $objA$ and $objB$. When the $R(objA, objB)$ indicate a high value and $objA$ is referred by user's utterance, the system infers the other object $objB$ would be referred at the same time.

3 Application for Behavior Decision and Dialogue Control

We have proposed an effective dialogue control method[9] which can handle vague and uncertain instructions from users. The method adopted Bayesian Networks to understand and complement the meaning of users' vague instructions. Generally, natural language expressions for ordering to robots have several object words such as name of person, name of place, visual feature of target object, kinds of work and so on, however, humans tend to omit these object words. In such a situation, the robots have to infer the omitted words. 'Task network' that describes the relationship among several names and concepts used in a specific task, was proposed in our previous work[8]. The relationship among several objects could not be decided in advance, therefore Bayesian Network was adopted to describe the loose representation. Figure 3 shows a network model for task achievement.

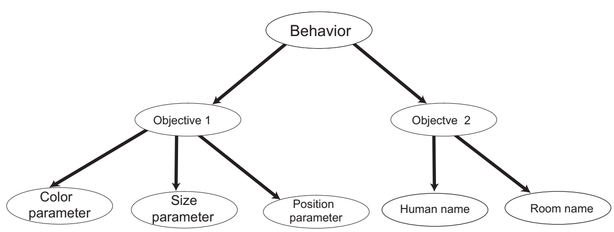


Fig. 3. A Bayesian Network model for task achievement

With the help of the task network, the robot could infer the omitted object words and keep on achieving behaviors even if the user omitted the target objects.

The robot could also infer suitable behavior to be executed against the vague user's request such as only gesture of pointing out a certain object. However, the dialogue control system has an assumption that the properties of objects are given and known, therefore changes of the real world environment could not be reflected to the dialogue control. In this paper, we propose an integration method that the spatial memory model is applied for the dialogue control system.

Figure 4 shows an outlook of integration between dialogue, observation and behavior decision.

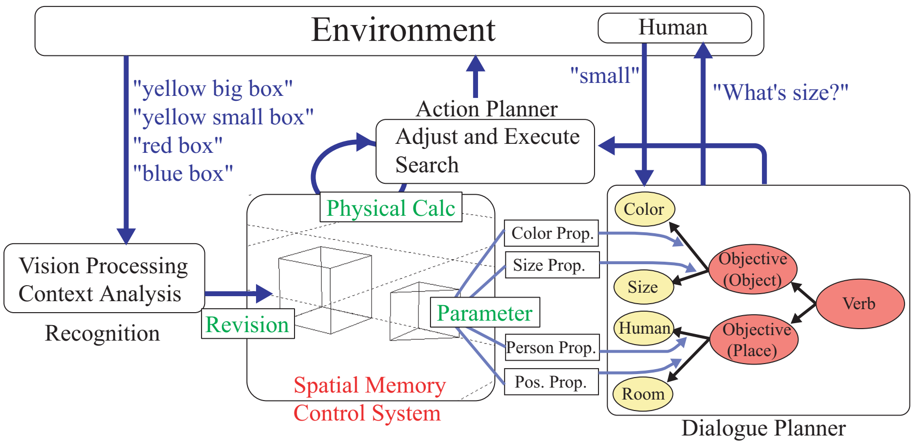


Fig. 4. Application of the spatial memory model to integrate with dialogue controller

4 Cooperative Behavior Using Probabilistic Geometric Memory Model

In this section experiment results of a humanoid robot that tried to achieve tasks through interaction between humans, using the integration method between geometric memory model and behavior planning system.

4.1 An Experiment of Behavior Control and Understanding of Context by Memory

Considers a task in which an user instructs a robot to carry some objects by simple natural language expression. The robot have to judge which is the suitable object as a referred target by the user because many objects are appeared in the eyesight of the robot.

A Bayesian Network shown in Fig.5 is considered for the object reference. Appeared nodes consist of visual sensor node and focus candidate node. The visual sensor nodes correspond to propositions which have several state value of visual feature such as red, blue (in case of color), big, small (in case of size), and so on. The focus candidate nodes correspond to a proposition as “Is current focus candidate as same as user's request?”, not correspond to each existing object. The inference result against the focus candidate node is called as *certainty factor of object identification*. When the certainty factor indicates high value, the focus candidate is suitable.

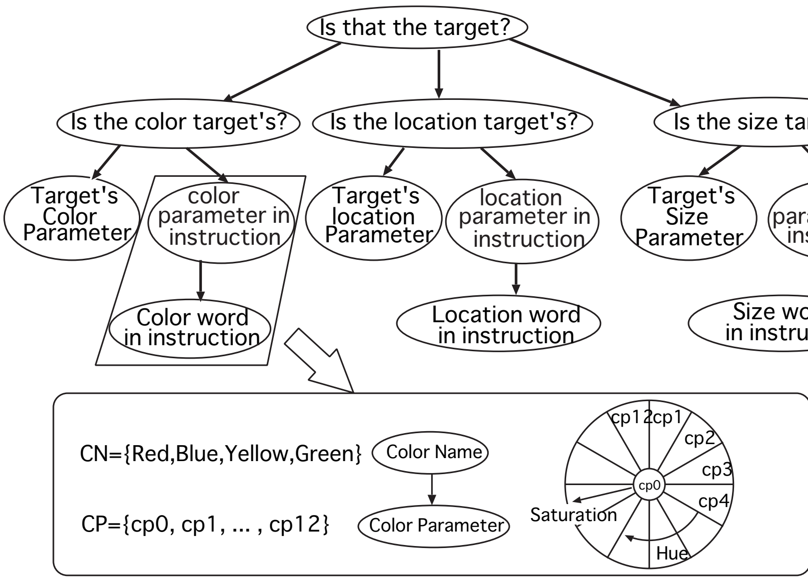


Fig. 5. A Bayesian Network for object reference

Next, consider dialogue management base on the certainty factor of object identification. It is not desirable to focus on a certainty factor of single target object. The distributed value of certainty factor for all candidate objects shows the global situation. We, here, introduce *degree of localization of target object*. The *localization* is defined as “a situation in which the target object is distinguished without any perceptual confusion”. Let CF_i the i-th certainty factor of focus candidate among n targets. It is needed that single certainty factor indicates high value and another factor indicate low value in order for the localization to be established. We define a criterion;

$$L \stackrel{\mathrm{def}}{=} \frac{\max(CF)}{\mathrm{second}(CF)} \tag{2}$$

for the evaluation of localization. $\max(CF)$ returns the highest component value in the vector $\max(CF)$, $\mathrm{second}(CF)$ returns the second highest component value. When the L indicate higher value, the focused candidate would be distinguished

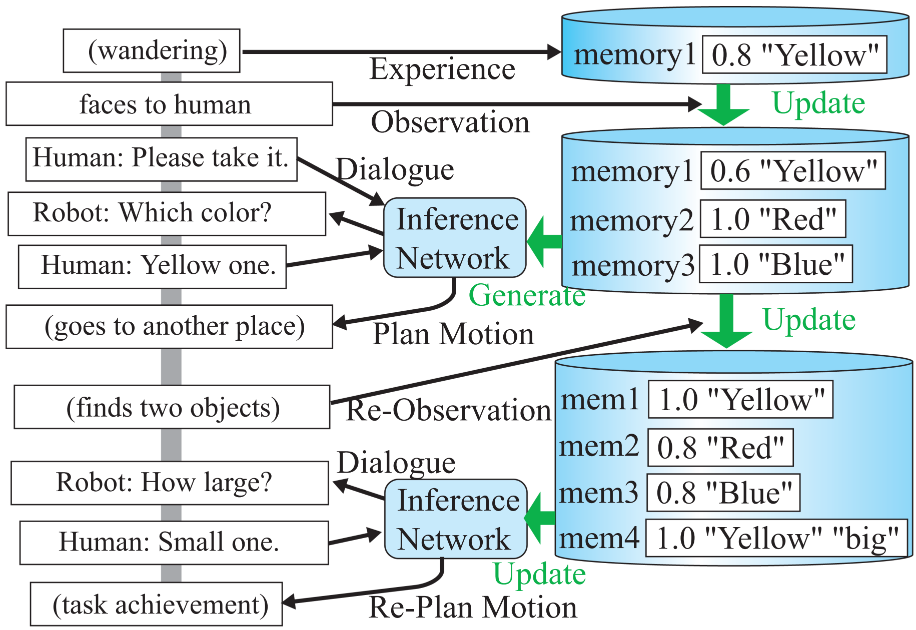


Fig. 6. Referring of the memory model for identifying the object

from other objects. Robots are able to manage the utterance and communication strategy to let the degree of localization be higher.

Cylinders shown in the right side of the Fig.6 show a sequence of memory status. Information about objects which was observed by the robot is stored to the memory, then status of the memory is updated as time advances. The existing probability decreases for an object which has not been observed. The existing probability rises when human's utterance includes the target even if the target have not been observed.

4.2 Representation of Usual Context in Daily-Life Environment

Next, representation of usual context, that is usual spatial relationship between several objects and usual visual properties of target objects, is focused. Considering throwing away cans and PET bottles with segregation as the example behavior. In habituated daily-life environment, we can find the suitable trash boxes with segregation process without wandering nor observation, because the physical properties such as location, color, size and so on, are already known as common sense. A network shown in Fig.8 was used for representation of such a habituated environment, in the situation of segregation of cans and PET bottles. The robot transfer observed image into KGB values using segmentation function, then the visual parameters are input into Red, Blue and Green node. The position and size information are also input into the Position and Size

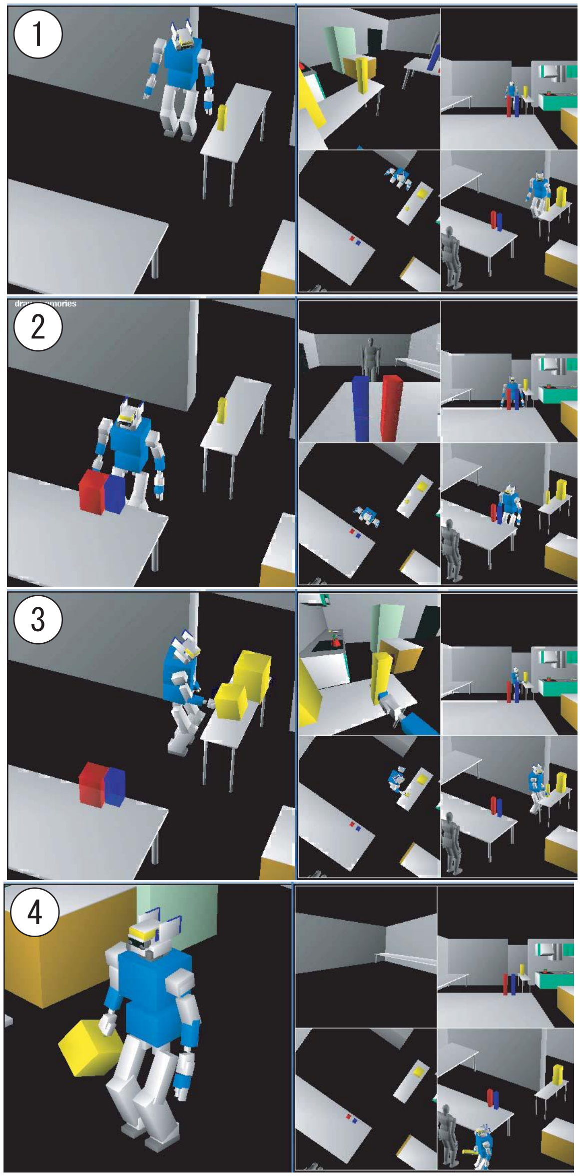


Fig. 7. An example of dialogue based behavior using memory model that manages context of the situation. (left): The memory model, (upper middle): Inner image of the robot, (right lower): outlook of the experiment environment. 1: The humanoid mistook to recognize two yellow objects (only one object was observed). 2: An user said "Take the yellow object" when the position of the humanoid changed. 3: The humanoid observed another yellow object when it moved to the previous place. 4: Achieving the task with making questions for confirmation.

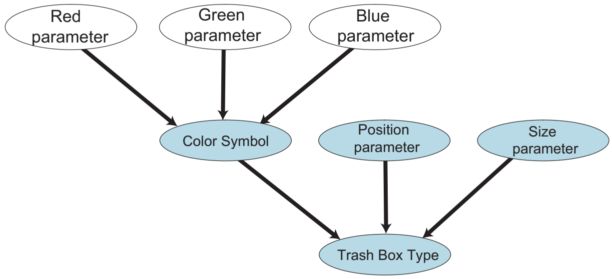


Fig. 8. A Bayesian network model used to identify the type of trash boxes

nodes. The robot distinguish the object by the inference value of the "Trash Box Type" node, with inference calculation between evidence nodes and the inference target node.

Figure 9 shows an experiment result that a humanoid robot tried to cast out PET bottles and cans into trash boxes. First, the robot found a can, then, tried to cast out it into a trash box which is 'CANS ONLY'. Second (scene 1), the robot referred the usual position of the trash box using stochastic memory model, then, the result showed that the trash box located on the left side of the robot. However, the robot found that the blue object in its eyesight was not a trash box for 'CANS', using the Bayesian network shown in Fig.8, then the object is registered as a trash box for PET bottles in memory model. Next (scene 2–3), the humanoid looked around to search a trash box for 'CANS', then succeeded to achieve the disposal task.

5 Conclusion

In this paper, we proposed a stochastic memory management system that can handle uncertain environmental information and vague instructions from users. The probabilities of memory model showed feasibilities not only in the spatial environmental model, but also in the temporal context. Bayesian Network was also adopted to realize a dialogue control system that can complement the vague instructions. Through several experiments, the feasibility of the proposed method was shown. Appropriate questions are generated against vague instructions, and uncertain information are complemented using the stochastic memory model.

In this paper, how to apply the stochastic spatial memory to human-robot interaction, was focused, but several other applications are considered as useful. Especially on personalization based on shared experiences between human and robots, would be an important issue for the communicable robots. We believe that the proposed method is useful to realize such personal robots that can handle uncertain information and vague instructions from users, using personal experiences in daily-life environments.

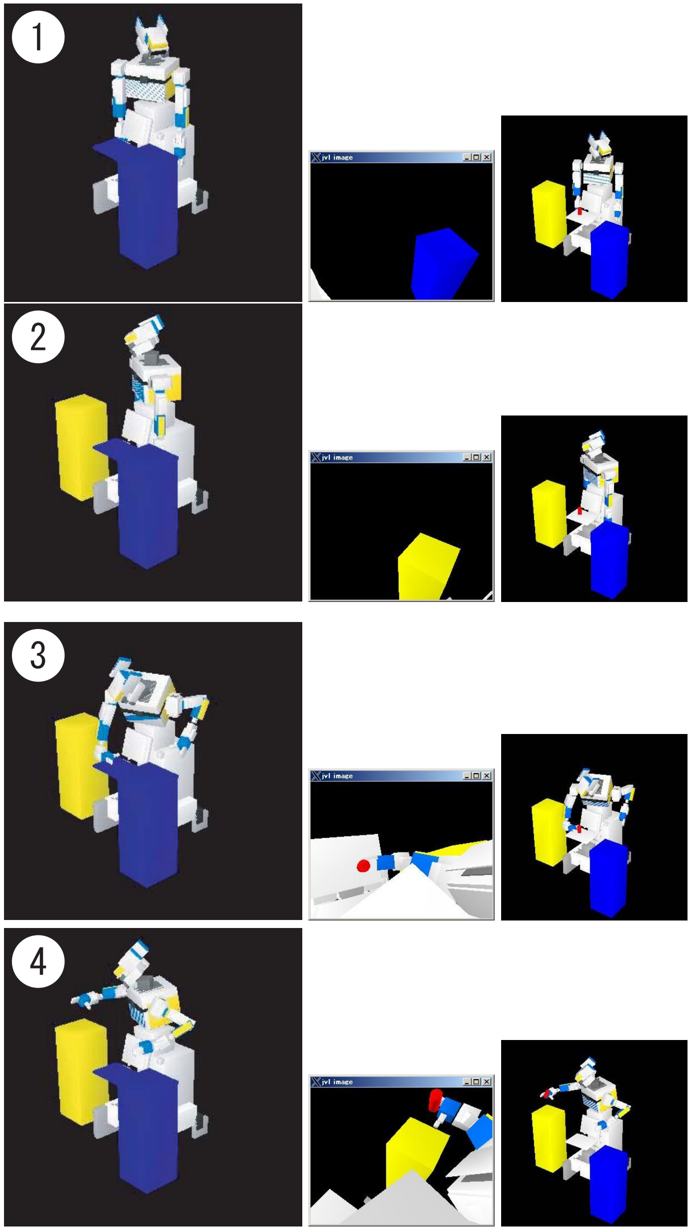


Fig. 9. An example behavior of trash separation; (left) memory model of the robot, (middle) inner image of the robot, (right) an outlook of experiment environment. 1: Deduction of suitable trash box for the can. 2: Trying another direction because of the first target was not suitable. 3-4: Throwing away the can into the target.

References

1. Shinya Fujie, Kenta Fukushima, and Tetsunori Kobayashi. "a conversation robot with back-channel feedback function based on linguistic and nonlinguistic information". In *Proc. Intl. Conf. on Autonomous Robots and Agents (ICARA2004)*, pages 379–384, 2004.
2. Yong K. Hwang, Pang C. Chen, and Peter A. Watterberg. Interactive Task Planning through Natural Language. In *Proceedings of the 1996 IEEE International Conference on Robotics & Automation*, pages 24–29, Apr 1996.
3. R. Bischoff and V. Graefe. Hermes - an intelligent humanoid robot, designed and tested for dependability. *Experimental Robotics VIII, B. Siciliano and P. Dario (eds), springer tracts in advanced robotics 5, Springer*, pages 64–74, 2003.
4. Zaliyana Mohd Hanafiah, Chizu Yamazaki, Akio Nakamura, and Yoshinori Kuno. "human-robot speech interface understanding inexplicit utterances using vision". In *CHI '04 extended abstracts on Human factors in computing systems*, pages 1321–1324, 2004.
5. A. Knoll, B. Hildebrandt, and J. Zhang. Instructing Cooperating Assembly Robots through Situated Dialogues in Natural Language. In *Proceedings of the 1997 IEEE International Conference on Robotics & Automation*, pages 888–894, 1997.
6. Junichi Ido, Yoshio Matsumoto, Tsukasa Ogasawara, and Ryuichi Nisimura. "humanoid with interaction ability using vision and speech information". In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.
7. Terry Winograd and Fernando Flores. *Understanding Computers and Cognition : A New Foundation for Design*. Addison-Wesley, 1987.
8. Tetsunari Inamura, Masayuki Inaba, and Hirochika Inoue. Dialogue control for task achievement based on evaluation of situational vagueness and stochastic representation of experiences. In *In the Proc. of Int'l Conf. on Intelligent Robots and Systems*, pages 2861–2866, 2004.
9. Tetsunari Inamura, Masayuki Inaba, and Hirochika Inoue. Pexis : Probabilistic experience representation based adaptive interaction system for personal robots. *Systems and Computers in Japan*, 35(6):98–109, 2004.
10. Yoshihisa Wakamatsu, Toshiyuki Kondo, and Koji Ito. A proposal of communication design for continual human-agent interaction using natural utterances. In *SICE Annual Conference*, pages 2315–2319, 2003.
11. Yusuke Shinyama, Takenobu Tokunaga, and Hozumi Tanaka. "kairai" - software robots understanding natural language. In *Proc. of 3rd International Workshop on Human Computer Conversation*, pages 158–163, 2000.
12. Eva Stopp, Klaus Peter Gapp, Gerd Herzog, Thomas Laengle, and Tim C. Lueth. Utilizaing Spatial Relations for Natural Language Access to an Autonomous Mobile Robot. In *The 18th German Annual Conference on Artificial Intelligence*, pages 39–50, 1994.

People Who Create Knowledge Sharing Communities

Asako Miura[1], Yasuyuki Kawaura[2], Setsuko Jifuku[3], Naoko Otaki[3], and Makoto Okamoto[3]

[1] Kobe Gakuin University
[2] Tokyo Keizai University
[3] Yahoo Japan Corporation

Abstract. Web-based knowledge sharing communities, which are supported by countless voluntary Internet users, are in widespread use in our lives. In one of these communities, Yahoo! Chiebukuro, one of the most famous web-based knowledge sharing communities in Japan, interpersonal communication among members is evoked as well as information exchange and accumulation. In this research, we conducted a questionnaire survey of Yahoo! Chiebukuro members. Based on 7989 survey samples, we intended to describe some features of community members, patterns of their participation in the community, and the reasons they joined the community. Results suggest that many members join the community and post their own information with the altruistic motive of helping others.

Keywords: knowledge sharing community, Yahoo! Chiebukuro, questionnaire survey, psychology.

1 Web-Based Knowledge Sharing Community

The Internet has become an indispensable tool for human society by accumulating an unprecedentedly large amount of information. Information uploaded to the Internet by an enormous number of users is aggregated and accessed through hyperlinks rather than being scattered around the world. As a result, various information databases are available on the Internet. A large amount of information stored in such databases crystallizes as knowledge over time, and we can get a chance to make some creative emergence on contact with it.

In recent years, certain kinds of web-based information databases are receiving massive interest. These are frequently updated in real time by registered, but anonymous, online users who also constitute a kind of web-based community. A virtual place constructed by people who interact using such information databases will be referred to as a knowledge sharing community in this study. For example, Wikipedia, the biggest multilingual free-content encyclopedia project on the Internet, is a typical example for a knowledge sharing community. Giles[1] compared the accuracy of information posted on Wikipedia with traditional and

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 88–98, 2007.

authoritative encyclopedia, Encyclopædia Britannica, and found that the difference between the two encyclopedias was small and they were comparable in the accuracy in their scientific contents. Wikipedia and other such knowledge sharing communities have gone through an enormous period of growth.

Knowledge sharing communities are supported by the volunteer contribution of millions of Internet users whereas traditional knowledge databases have been supported by the duties of a limited number of authorities. We can find some important clues to explain the details of shared knowledge itself by exploring the reasons why Internet users join such knowledge sharing communities and what they might acquire from them. In this study to examine such research questions, we conducted a questionnaire survey of the Internet users who joined Yahoo! Chiebukuro, one of the major knowledge sharing community services in Japan. Based on the survey findings, we tried to describe some common characteristics of community members, patterns of their participation in the community, and the reasons they joined the community.

1.1 Yahoo! Chiebukuro

In Japan, Yahoo! Chiebukuro is a knowledge sharing community service operated by Yahoo Japan Corporation[1]. "Chie" means knowledge or wisdom and "Bukuro" means bag in Japanese. Therefore, "Chiebukuro" means a bag filled with a great deal of knowledge or wisdom. Anyone who signs up for Yahoo! JAPAN can participate in the Chiebukuro community anonymously and free of charge. Community members can post any questions they have and can post answers to any questions posted by other members. All posted questions and answers are available to the public. The community commenced operations as a beta version from April 4, 2004 and began operating officially on November 7, 2005.

Yahoo! Chiebukuro is a knowledge sharing community that attaches importance to interpersonal communication among members. Registered members can get involved in the community in two different roles as the occasion may demand; these roles are (1) active information-seekers who expresses their wishes for getting information by posting questions, (2) informants who provide information by posting answers. Because the content and genre of questions are not limited, active information-seekers are free to appeal for any information they want. There is also no need to consider whether a question has a single correct answer or not. Ill-defined ambiguous questions that might have multiple alternative answers or no correct answer are allowed also. The ultimate goal of the Yahoo! Chiebukuro community is to be a useful service that accumulates encyclopedic knowledge as much as possible; however, a higher priority is for it to be a comfortable place for interpersonal communication that enables members to exchange and share knowledge on their own terms.

Though both Wikipedia and Yahoo! Chiebukuro can be regarded as knowledge sharing communities, they have distinguishing aspects. In the case of the

[1] URL: http://chiebukuro.yahoo.co.jp/; Yahoo! USA operates the same kind of service named "Yahoo! Answers (http://answers.yahoo.com/)."

Wikipedia project, the biggest goal is to create a free highly reliable encyclopedia and all published information is separately classified and organized, as are the items in an encyclopedia. Accuracy based on objective facts is primarily required in Wikipedia's content. Registered members of Wikipedia get involved in the project in one role, that of informants who provide information by writing items. Though they can add new items or edit items already added, they have no chance to ask questions of other members[2].

One of the unique features of Yahoo! Chiebukuro is that some attributes of members are disclosed accompanied with their postings (both questions and answers). Yahoo! Chiebukuro has a function for specifying the attributes of community members. The estimated attributes are "degree of contribution" and "level". The degree of contribution to the community of each member is measured by the number of questions and answers the member posts. Active information-seekers are allowed to subjectively select the most useful one as the "best answer" from all posted answers to the questions they posted. The "level" of each member is estimated based on two parameters: (1) as an active information-seeker: the ratio of questions for which they select the "best answer" and (2) as an informant: the ratio of their answers that are selected as the "best answer" by active information-seekers who posted relevant questions. As all community members (and even lurkers or non-members) can refer to these two different kinds of attribute information, they can not only browse the content of posted information, but also they can infer how the posters themselves get involved in the community. This feature seems to be one of the important factors that specifies the overall atmosphere of the community, and to have a great impact on member behavior. This is because people generally decide whether or not to have contact with others depending not only on what topics they may share with them but also what kind of people they are.

In this study, our main goal was to describe the common profile of members of Yahoo! Chiebukuro, a typical knowledge sharing community. Especially we focused our attention on general characteristics common or unique to their possible roles as described above, active information seekers who post questions to the community and informants who post answers to others' questions. Members of both groups were active members and actually communicated to others in the community by posting their messages. We conducted a questionnaire survey about members' behaviors in the community and related psychological variables, and we combined those with objective behavioral data obtained from the server.

2 Methods

2.1 Procedure

The questionnaire survey was conducted on the Yahoo! Chiebukuro website from December 6 to 20, 2005. Two versions of the questionnaire were used in this sur-

[2] That is not to say that no interpersonal communication exists in Wikipedia. Members can communicate about the contents of items in a "discussion" section about each item.

vey. One was a “information-seeker” version for members who posted questions, and the other was a “informant” version for members who posted answers. Possible respondents were members who logged in at Yahoo! Chiebukuro and posted a question or an answer during the survey. Just after they had finished posting a question or an answer, the request to participate in the survey was displayed only once for each posting. Participating in the survey was voluntary, and respondents did not receive any compensation.

The total number of respondents was 5916 for the information-seeker version and 2869 for the informant version. If respondents completed both versions of the questionnaire, the data were extracted from the version that they had filled out first. Finally, we considered 5515 responses from information-seekers and the data from 2474 responses from informants as valid for analysis.

2.2 Questionnaire

Cross-version items for all respondents. All respondents were asked to complete the following items (1) demographic traits (gender and age), (2) behaviors in and attitudes towards the community (login frequency in the previous month, interest in one’s estimated degree of contribution, need for communication with other members, intention to continue participation), and (3) psychological traits for information need (twelve items with 4-point Likert scale extracted from Kawaura[2]).

Items for information-seekers only. The unique items on the version which only information-seekers were asked to complete were (1) the kind and content of their question (whether it had a correct answer or not, urgent need, and importance), (2) the reasons they posted their question (nine items with 4-point Likert scale), (3) the reason they selected Yahoo! Chiebukuro for posting their question, and (4) whether they tried to find an answer by themselves before they posted their question, and what were their investigation resources if they did (multiple choice from 6 online and 2 offline alternatives and other).

Items for informants only. The informants were asked to complete the following unique items: (1) the kind and content of the question to which they responded (whether it had a correct answer or not, perceived urgent need, and perceived importance), (2) the reasons they posted their answer (thirteen items with 4-point Likert scale), (3) a self-evaluation of the correctness of their answer, and (4) whether they tried to confirm the accuracy of their answer before posting, and what were the online/offline resources used for investigation if they did (multiple choice from the same alternatives as those in the information-seeker version).

On the questionnaire form, the items were categorized into groups depending on their content, and the order of presenting the items was randomized in each group to counterbalance the order effect.

2.3 Objective Behavioral Index

We collected some objective behavioral indices for all respondents from the server log; they were the number of questions and answers they posted, the number of answers to the questions they posted, the date of their registration to the community, and the date they posted their first question and/or answer. The number of their public nicknames[3] was also counted. If respondents used multiple public nicknames when posting their messages, the data for all of those nicknames were combined into one registered ID.

3 Results

3.1 Demographic Traits

Respondents (both information-seekers and informants) were 4021 males (50.3%) and 3970 females (49.7%) . Their mean age was 31.4 years (SD=11.39). The ratio of males who responded to the informant version was significantly greater than that for the information-seeker version (information-seeker version: male 48.3%, female 51.7%; informant version: male 54.7%, female 45.3%; $\chi^2_{(1)}$=27.88, p<.001). The mean age for those who responded to the informant version was significantly higher than that for the information-seeker version (information-seeker version: 30.1 years (SD=11.23); informant version: 34.3 years (SD=11.20); $t_{(7989)}$=-15.55, p<.001).

3.2 Participation in the Community

To obtain an overview of how respondents behaved in the community, we calculated descriptive statistics for the primary objective indices, and these are shown in Table 1. Though we found a substantive deviation among respondents, the mean number of posted answers was greater than that of posted questions. In addition, though some respondents frequently changed their registered name (up to 33 times), 96.1% of respondents had never changed their registered name.

Table 1. Activities of all respondents in Yahoo! Chiebukuro

	Posted questions	Posted answers	Answers to posted questions
Mean	16.0	91.3	57.1
SD	74.25	702.40	332.19
Maximum	2878	29977	13557

Regarding the number of answers to posted questions, only 4.8% of information-seekers received a smaller number of answers that the number of

[3] In Yahoo! Chiebukuro, community members can register multiple public nicknames for each registered ID and change them as the situation demands.

Table 2. Participation pattern based on posting-behavior

Participation pattern	Posting-behavior	N
Type 1	posted questions only	2968
Type 2	posted more questions than answers	1403
Type 3	posted more answers than questions	2746
Type 4	posted answers only	872

Table 3. Activities in the community by each participation pattern

	Participation pattern	Mean	SD
Posted questions	Type 1	2.7	6.04
	Type 2	36.3	128.16
	Type 3	25.1	83.97
	Type 4	–	–
Posted answers	Type 1	–	–
	Type 2	13.0	32.92
	Type 3	215.2	950.46
	Type 4	137.5	1256.24
Answers to posted questions	Type 1	6.7	15.86
	Type 2	126.7	546.83
	Type 3	93.9	400.65
	Type 4	–	–

questions they posted, that is, had the experience of receiving no feedback to their questions. As can be calculated from the data in Table 1, 3.6 answers per question were posted on average. This tendency was consistent in the patterns of community participation described below. The patterns suggest that Yahoo! Chiebukuro is a knowledge sharing community in which a considerable number of members actively provide information in answer to requests from other members.

Next, we divided the community participation patterns into four types based on the number of questions and answers members posted. The four participation patterns were those in which respondents (1) posted questions only, (2) posted both with more questions than answers, (3) posted both and with more answers than questions[4], and (4) posted answers only. The number of respondents of each type was 2968, 1403, 2746, and 872, respectively, as is summarized in Table 2.

A comparison of the mean number of posted questions and answers among these four participation types is summarized in Table 3. Though respondents who posted questions only (Type 1) accounted for the largest share (37.2%) of the total, the mean number of questions posted by them was very small. This result suggests that they did not get particularly involved in the community. On the other hand, respondents who only posted answers (Type 4) seemed to

[4] If respondents posted the same number of questions and answers, they were classified into Type 3.

get actively involved in the community as the mean number of answers posted by them was 137.5 and was considerably large. Respondents who posted both questions and answers (Types 2 and 3) were the majority in the total, and respondents of Type 3 tended to post their messages, particularly answers, very actively. This result suggests that respondents of Types 2 and 3 played a central role in the knowledge sharing cycle of the community and activated it through posting their own questions as well as answering the questions of others by posting answers.

3.3 Reason for Question-Posting

We examined the reasons respondents posted their questions on Yahoo! Chiebukuro based on the survey of information-seekers. They responded to each of nine items regarding their last question. Table 4 shows the ratio of respondents who answered "agree" or "rather agree" on each item and the mean number of the total of questions they posted. The primary reason for posting questions was to get information aggressively, especially to derive obvious benefits by getting information; whereas, the ratio of respondents who posted answers without any specific need for problem solving but with the intention to boost the Chiebukuro community or gain attention from others was relatively low. However, the mean number of questions posted by respondents who had such intentions as the latter was larger than that of the former. Especially, respondents with the intention to gain attention from others tended to post more questions than others.

Table 4. Simple statistics of question-posting reasons

Items	Ratio(%)	Mean of posted questions
To gain advantages by getting answers	77.8	15.7
To avoid trouble by getting answers	77.7	15.0
To learn something they wonder about	67.5	19.6
To get answers to their tiny questions of life	67.4	18.3
To confirm the validity of their opinions	26.5	21.4
To have fun with answers from others	15.0	27.1
To boost Chiebukuro community	12.2	23.5
To make cyberfriends	4.8	31.7
To gain attention from others	3.6	33.3

We conducted factor analysis with the maximum likelihood method and promax rotation, and extracted the three-factor structure that is shown in Table 5. Judging from the items with high factor loading, these three factors can be respectively labeled as self-presentation and community activation, practical rewards, and psychological rewards.

As are presented in Table 6, we made multiple comparisons of the total score of items that had high factor loading for each of three factors in three of four

Table 5. Factor analysis: Question-posting reasons

Items	Factor Loading		
	Factor 1	Factor 2	Factor 3
To gain attention from others	$\boxed{0.79}$	-0.01	-0.11
To make cyberfriends	$\boxed{0.75}$	0.04	-0.09
To boost Chiebukuro community	$\boxed{0.66}$	0.04	0.06
To have fun with answers from others	$\boxed{0.56}$	-0.10	0.12
To confirm the validity of their opinions	0.31	0.09	0.22
To avoid trouble by getting answers	-0.01	$\boxed{0.93}$	-0.07
To gain advantage by getting answers	0.04	$\boxed{0.36}$	0.24
To get answers to their tiny questions of life	-0.00	0.18	$\boxed{0.64}$
To learn something they wonder about	0.04	-0.10	$\boxed{0.57}$
Int-Fac Corr: F1-F2: -0.01, F2-F3: 0.34, F1-F3: -0.15			

participation patterns[5] to investigate whether the influence of these factors depended on how respondents behaved in the community. A lower score indicates a higher impact for each factor in question-posting behavior. One-way analysis of the variance of all three factors revealed significant differences in participation patterns (p<.01). Respondents who posted both questions and answers (Types 2 and 3) tended to be more motivated by Factor 1, self-presentation and community activation , and Factor 2, practical rewards, than those who posted questions only (Type 1). The impact of Factor 3, psychological rewards, was significantly different in all patterns and was highest for respondents who posted more answers than questions.

Table 6. Multiple comparisons of question-posting reasons among participation patterns

Factors	Participation pattern	Mean
Self presentation/Community activation	Type 1	14.8^{A}
	Type 2	14.5^{B}
	Type 3	14.6^{B}
Practical rewards	Type 1	3.5^{A}
	Type 2	3.8^{B}
	Type 3	3.9^{B}
Psychological rewards	Type 1	4.4^{A}
	Type 2	4.2^{B}
	Type 3	4.0^{C}

3.4 Reason for Answer-Posting

The reasons respondents posted answers on Yahoo! Chiebukuro were examined based on the survey of informants. They answered each of thirteen items re-

[5] Respondents who posted answers only (Type 4) didn't respond to the information-seeker survey.

garding their last answer. Table 7 shows the ratio of respondents who answered "agree" or "rather agree" on each item and the mean number of answers they posted. The altruistic reason "to help information-seekers" was cited most frequently, and subsequently, over half of respondents experienced happiness and joy by joining the community and responding to questions from others. On the other hand, the ratio of respondents who cited the reasons related to their own benefits, such as those based on the norm of reciprocity or self-interest, was relatively low compared to those who cited the other reasons. The mean number of posted answers was not so different among for all reasons.

We conducted factor analysis with the maximum likelihood method and promax rotation and extracted the four-factor structure that is shown in Table 8. Judging from the items with high factor loading, these four factors can be la-

Table 7. Simple statistics of answer-posting reasons

Items	Ratio(%)	Mean
To solve problems of information-seekers	89.7	14.9
To give their own information to information-seekers	86.9	17.3
To just enjoy answering	79.4	15.6
To help information-seekers	78.3	19.2
Due to positive feeling towards Chiebukuro community	77.6	15.8
To feel happy when receiving thanks from information-seekers	66.9	14.4
To receive new information by responding	55.3	16.8
To boost Chiebukuro community	47.4	13.5
To confirm the validity of their opinion	43.1	16.8
To receive support when they are in trouble	41.9	16.4
To open a line of communication	37.4	16.5
To reciprocate answers from others	35.2	14.6
To enhance their evaluations	32.3	14.5

Table 8. Factor analysis: Answer-posting reasons

Items	Factor Loading			
	Factor 1	Factor 2	Factor 3	Factor 4
To solve problems of information-seekers	0.82	-0.05	-0.01	-0.03
To help information-seekers	0.68	0.15	0.02	-0.12
To give their own information to information-seekers	0.60	-0.07	0.02	0.12
To receive support when he/she is in trouble	0.01	0.81	-0.05	0.03
To reciprocate a past courtesy from others	0.01	0.56	0.02	0.00
To receive new information by answering	0.03	0.46	0.13	0.07
Due to positive feeling towards Chiebukuro community	0.11	0.01	0.66	-0.05
To boost Chiebukuro community	-0.05	0.17	0.59	0.04
To just enjoy answering	-0.01	-0.16	0.43	0.33
To open a line of communication	-0.05	0.31	0.40	0.03
To enhance their evaluations	-0.06	0.04	0.03	0.68
To confirm the validity of their opinions	0.03	0.12	0.00	0.50
To feel happy when receiving thanks from information-seekers	0.37	0.06	0.01	0.41
Int-Fac Corr: F1-F2: 0.38, F1-F3: 0.35, F1-F4: 0.36, F2-F3: 0.52, F2-F4: 0.52, F3-F4: 0.56				

Table 9. Multiple comparisons of answer-posting reasons for participation patterns

Factors	Participation pattern	Mean
Altruism	Type 2	5.3^{A}
	Type 3	5.3^{A}
	Type 4	5.1^{A}
Norm of reciprocity	Type 2	7.4^{A}
	Type 3	7.8^{B}
	Type 4	8.7^{C}
Community activation/Entertainment	Type 2	9.4^{AB}
	Type 3	9.1^{A}
	Type 4	9.7^{B}
Self-interest	Type 2	7.9^{AB}
	Type 3	7.6^{A}
	Type 4	8.0^{B}

beled respectively as altruism, norm of reciprocity, community actuvation and entertainment, and self-interest,.

As presented in Table 9, we conducted multiple comparisons of the total score of items that had high factor loading for each of four factors for three of four participation patterns[6] to investigate whether the influence of these factors depended on how respondents behaved in the community. A lower score indicates a higher impact for each factor in answer-posting behavior. One-way analysis of the variance of the three factors (excluding Factor 1, altruism) revealed a significant difference in participation patterns (p<.01). Respondents who posted both questions and answers (Types 2 and 3) tended to be more motivated by Factor 2, the norm of reciprocity, than those who posted answers only (Type 4), and this tendency was more salient for those who posted more questions than answers. Both Factor 3 (activation of community and entertainment) and Factor 4 (self-interest) had a rather great influence on respondents who posted more answers than questions (Type 3). On the other hand, the impact of altruism on posting behavior seemed to be almost equal and high in every participation pattern.

4 Discussion

In this study, we conducted a questionnaire survey to describe the common profile of members of Yahoo! Chiebukuro, a typical knowledge sharing communities. Various kinds of objective behavioral data about survey respondents in the community were obtained from the server and linked to the survey data. Respondents were divided into four of participation pattern based on the number of questions and answers they posted. We categorized the reasons they posted questions or answers with factor analysis, and compared the reasons for the various participation patterns.

[6] Respondents who posted questions only (Type 1) didn't answer the survey of informants.

According to analysis based on objective behavioral indices, that is, the number of posted answers per question and the ratio of participation, our results suggested that Yahoo! Chiebukuro is a knowledge sharing community in which many respondents actively exchange and accumulate information. Further, analysis based on subjective survey data suggested the reasons the respondents posted questions or answers. Our research suggested that communication between information-seekers and informants generally occurs as a combination of the active expression of seeking information and a corresponding supplying of information. Some respondents, though not the majority in the community, had emotional attachments to the community itself, and consequently, they were motivated to post messages actively. Others regarded the community as an easy stage for self-presentation and consequently, they tended to post as many questions as possible.

5 Conclusion

Currently, web-based communities are progressing to the new Web-2.0 phase. As the Web-2.0 worldview is based on the belief that human nature is fundamentally good, everyone is expected to transmit information to many people freely and information will be shared not just with a handful of celebrities but with all people. Though previous studies have pointed out some negative aspects of anonymous web-based communities, for example, flaming[3], this study demonstrates one of the positive aspects of such communities with results suggesting that many respondents join a typical knowledge sharing community and post information with the pure motive of helping others.

As we focus our attention on illustrating the general profile of respondents in this study, there still remain many questionnaire items and objective behavioral indices which have not been investigated yet. In the near future work, we will examine the relationship between their behaviors in the community and their own psychological traits measured by the questionnaire survey. Furthermore, results based on quantitative analysis will link to some kinds of qualitative data, for example, genre and content of question/answer. At that time, we will especially center on a leading community member who posts both questions and answers. In this way, we will be able to determine the dynamic state of web-based knowledge sharing community in more detail.

References

1. Giles, J.: Internet encyclopædias go head to head. Nature, **438** (2005) 900–901
2. Kawaura, Y.: Information needs and behaviors. Asahi AD Monthly (1998, January) 42–47 (In Japanese)
3. Lea, M., O'Shea, T., Fung, P., Spears, R.: 'Flaming' in computer-mediated communication: Observations, explanations, implications. In M. Lea (ed.) Contexts of Computer-Mediated Communication. Harvester-Wheatsheaf. (1992) 89–112

Logic and Engineering of Natural Language Semantics (LENLS) 3

Eric McCready

Department of English
Aoyama Gakuin University

1 The Workshop

This was the third Logic and Engineering of Natural Language Semantics workshop. This year it was held at Funabori Tower Hall in Tokyo, Japan on June 5th and 6th of 2006. There were three invited talks, by Makoto Kanazawa, Chung-Min Lee, and Christopher Potts. In addition, the program committee (see Acknowledgements) selected fifteen abstracts from those submitted, for a total of eighteen talks.

This year's theme was formal pragmatics. In recent years there have been a number of exciting developments in this area. Researchers have applied game-theoretical and utility-theoretic techniques to problems such as Gricean communication and relevance, implicature, and blocking; for instance, new work in multidimensional logic has given insight into conventional implicature; new formal techniques have been applied to discourse structure and coherence. The goal of this year's workshop was to bring researchers in these areas together for discussion. This was an eminent success.

2 Papers

The talks ultimately fell into several distinct categories: questions, information structure, logical and philosophical aspects of dynamic semantics, computational aspects, and quantification. There were also several talks that fit into none of the above categories. Rather than introducing the papers given at the conference, however, I would like to concentrate on the papers in the present volume.

First, there are several papers that consider logical systems directly. First is the contribution of Norry Ogata, 'A Dynamic Semantics of Intentional Identity,' which provides a new dynamic system to model the identity of individuals in intentional contexts, in particular attitudes. Satoru Suzuki considers probabilistic updates in a new solution to the Monty Hall dilemma. Finally, Tomoyuki Yamada, in his contribution 'Logical Dynamics of Commands and Obligations,' works to solve some problems in deontic logic using a dynamic system.

Next comes a group of papers situated at the semantics-pragmatics interface. David Oshima, in his paper 'On Factive Islands: Pragmatic Anomaly versus Pragmatic Infelicity,' uses pragmatic means to account for the unavailability of long-distance dependencies into factive islands. Linton Wang and Eric McCready,

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 101–102, 2007.

in ‘Aspects of the Indefiniteness Effect,’ work to account for a newly noticed phenomenon arising with certain kinds of quantifiers in questions; they claim that infelicity arises with indefinites due to pragmatic preferences on interpretation. Yi Mao and Beihai Zhou provide a new system for understanding metaphors and metonymy in their ‘Interpreting Metaphors in a New Semantic Theory of Concept.’ Finally, Sumiyo Nishiguchi’s contribution ‘Covert Emotive Modality is a Monster’ explores an approach to shifted tenses and other elements in which covert modalities function as monstrous operators.

The final group of papers is substantially pragmatic in nature. Christopher Potts, in his ‘Conversational Implicatures via General Pragmatic Pressures,’ shows a way in which implicature can be derived from standardly assumed pragmatic principles in a formally explicit and sensible way. Next is the approach to exhaustive interpretation with Japanese *dake-wa* ‘only-Top’ of Yurie Hara in her ‘*Dake-wa*: Exhaustifying Assertions,’ where she treats this predicate as quantifying over potential literal acts. Finally, Yukio Furukawa considers the influence of context on quantifier interpretation in his ‘Unembedded ‘Negative’ Quantifiers.’

As the reader will notice, the range of topics addressed by even this subset of the contributed papers is very wide. As a result, the workshop was very stimulating; the participants, with their different perspectives, gave useful and interesting comments on each paper. From the perspective of the organizers at least, the result was very successful. We hope (and believe) that the other participants shared this impression.

Acknowledgements

I would like to thank first the other members of the organizing committee: Yasuo Nakayama Norry Ogata, Katsuhiko Yabushita, and Kei Yoshimoto. Next, I would like to thank the members of the scientific committee: Yurie Hara, Jinung Kim, Nakayama McCready (though it seems a bit odd for me to thank myself), Yasuo Nakayama, Rick Nouwen, Patrizia Paggio, Brian Reese, Rolf Schwitter, Anders Soegaard, and Linton Wang, and the members of the organizing committee. The organizing committee would also like to gratefully acknowledge the financial (and other) support of the Tohoku University 21st Century COE Program in Humanities “Strategic Research and Education Center for an Integrated Approach to Language, Brain and Computation,” and the Japanese Society for Artificial Intelligence.

A Dynamic Semantics of Intentional Identity

Norihiro Ogata

Department of Language and Information, Osaka University
1-8 Machikaneyama, Toyonaka, Osaka, Japan
ogata@lang.osaka-u.ac.jp

Abstract. This paper will propose a new version of *Dynamic Modal Predicate Logic (DMPL)* to treat dynamics of *intentional identity* [1] and other similar notions such as *weak intentional identity* and *multiple intentional identity* by revising the DMPL of modal subordination [2] by adding a kind of counterpart relation. The revised dynamic semantics of DMPL is much simpler than Edelberg's semantics [3] of intentional identity and can treat *weak intentional identity* and *multiple intentional identity* which have not been treated by most semantics of intentional identity as well as Edelberg's semantics.

1 Introduction

Formal semantics of natural language focuses on dynamics of information flow in texts such as anaphora and tense. In particular, [4] formalizes such works as a logic, callled *Dynamic Predicate Logic (DPL)*. This paper will propose a new version of *Dynamic Modal Predicate Logic (DMPL)* to treat dynamics of the *intentional identity* by [1] and other similar notions, of which other version has been already proposed by Ogata [2] to treat modal subordination. The main point of this version of dynamic semantics of $DMPL$ is that the dynamic semantics of $DMPL$ is based on a Kripke semantics of quantified modal logic [5,6,7] with a kind of counterpart relation, which is much simpler than Edelberg's semantics [3], which is a non-Kripke semantics with complicated counterpart relations such as $\approx$ and $\preceq$. Furthermore this version of dynamic semantics of $DMPL$ can treat *weak intentional identity* and *multiple intentional identity* which cannot be treated by most semantics of intentional identity such as [8,9,10,11] as well as Edelberg's semantics.

Section 2 will summarize the notion of intentional idenity. Section 3 will argue Edelberg's semantics of a quantified multi-agent modal language to treat intentional identity, which is slightly reformulated by ordinal notions on multi-modal Kripke models. Section 4 will propose a dynamic semantics of the quantified multi-agent modal language based on multi-modal Kripke models with a kind of counterpart relation.

2 On Intentional Identity and Counterpart Relation

In this paper, I introduce six types of intentional identities, which have no appropriate logic formulas expressing them in the sense of classical quantified modal

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 103–117, 2007.

logic [12], such as *Geach's Hob-Nob sentences*, *Ederberg's Arsky-Barsky sentences*, *cross-speaker anaphora*, and *Weak Intentional Identity*, as follows.

2.1 Geach's Hob-Nob Sentence

Geach [1] found a kind of *identity*, dubbed *intentional identity*, since the identity is similar to the situation where many archers point their arrows at one target (in Latin *intendo arcum in ...*) but we can verify that they are all pointing their arrows the same way, regardless of finding out whether there is any shootable object at the point where the lines of fire meet. He showed the following example:

Example 1. John, a reporter, is decribing an outbreak of witch mania in Gotham village as in (1). He does not believe the existence of any witch.

(1) Hob thinks a witch has blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's sow.

The intended logical form of (1), called *the Hob-Nob sentence*, is not any sentence in (2).

(2) a. There is a witch x such that Hob thinks x has blighted Bob's mare and Nob wonders whether x killed Cob's sow.
b. Hob thinks there is x which is a witch and has blighted Bob's mare and Nob wonders whether x is a witch and killed Cob's sow.
c. Hob thinks a witch has blighted Bob's mare and Nob wonders whether *the witch who blighted Bob's mare* killed Cob's sow.

These are formulated as follows, respectively:

(3) a. $\exists x.\varphi(x) \wedge \Box_{Hob}\psi(x) \wedge \Box_{Nob}\chi(x)$
b. $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x)) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))$
c. $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x)) \wedge \Box_{Nob}(\chi(\iota x.\varphi(x) \wedge \psi(x)))$

(2a) and (3a) imply the real existence of the witch but the reporter does not believe it. Therefore, (2b) and (3a) are not appropriate as the logical forms of (1). In (3b) "$\exists x$" cannot bind x which occurs in $\Box_{Nob}(\varphi(x) \wedge \chi(x))$. Therefore, it is not appropriate as the logical form of (1). In (2c), 'she' is interpreted as an *E-type pronoun.* But it implies that Nob believes that the witch blighted Bob's mare, although the reporter cannot certify it. This point is a basic difference of intentional identity from *modal subordination* which also binds pronouns across the boundary of modalities. For example, (4a) implies (4b), although this kind of inference is not admitted in (1):

(4) Hob thinks a witch has blighted Bob's mare, and wonders whether she killed Cob's sow. Therefore, Hob thinks a witch has blighted Bob's mare, and wonders whether the witch who (Hob thinks) has blighted Bob's mare killed Cob's sow.

(4) is rewritten by formulas as follows:

(5) $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x)) \wedge \Box_{Hob}(\varphi(x) \wedge \chi(x)) \models \Box_{Hob}(\chi(\iota x.\varphi(x) \wedge \psi(x) \wedge \chi(x)))$

That is, in modal subordination as in (3c), the pronoun *she* can be substituted by E-type pronouns, whereas in intentional identity the operation is forbidden.

Furthermore, the interpretation of (6a) as modal subordination leads us to a contradiction, whereas the interpretation of (6a) as intentional identity does not so (compare (6a) with a kind of modal subordination (6b)).

(6) a. Hob thinks a witch has blighted Bob's mare, and Nob thinks she (the same witch) has not blighted Bob's mare.
   b. Hob thinks a witch has blighted Bob's mare, and he thinks she (the same witch) has not blighted Bob's mare.

In modal subordination such as (6b) 'a witch' exists in each mental state w' of Hob which are accessible from the actual world w in the sense of the Kripke semantics and 'she' refers to it in w', therefore, (6b) leads us to a contradiction, whereas in intentional identity such as (6a) 'a witch' exists in each mental state w' of Hob which are accessible from the actual world w and 'she' exists in each mental state w'' of Nob which are accessible from the actual world w, therefore, (6a) does not lead us to a contradiction, but 'she' in w'' must *refer to* 'a witch' in w' *across possible worlds* in some sense. Therefore, in (6a), even though the object denoted by 'she' is the same with the object denoted by 'a witch', the object denoted by 'a witch' exists in each Hob's mental state and the object denoted by 'she' exists in each Nob's mental state, and this fact requires a kind of correspondence relation between them. This is a point which we encounter when we want to handle not modal subordination but intentional identity.

The most popular correspondence relation between objects across possible worlds is David Lewis's [13] *counterpart relation* $\mathcal{C}$, which is axiomatized as follows (omitting mention of 'actual'): let W be a non-empty set of possible worlds, D a non-empty set of possible individuals and $D(\cdot) : W \to \mathscr{P}(D) \setminus \emptyset$

1. $\forall x \forall w.x \in D(w) \to w \in W$
2. $\forall x \forall wu.x \in D(w) \& x \in D(u) \to w = u$
3. $\forall xy.\mathcal{C}(x, y) \to \exists w.x \in D(w)$
4. $\forall xy.\mathcal{C}(x, y) \to \exists w.y \in D(w)$
5. $\forall xy \forall w.(x \in D(w) \& y \in D(w) \& \mathcal{C}(x, y)) \to x = y$
6. $\forall x \forall w.x \in D(w) \to \mathcal{C}(x, x)$

However, $\mathcal{C}$ is too strong to treat Geach's intentional identity, since condition 2 of $\mathcal{C}$ prohibits an object from belonging to two or more possible worlds. So we need another type of counterpart relation, which will be investigated in section 4.

In this paper, I will use (3b), the most literal paraphrase of (1), as the official logical form of (1). Therefore, the basic task of this paper is to show how dynamic semantics of quantified modal formulas such as (3b) reflects intentional identity.

2.2 Edelberg's Arsky-Barsky Sentences

Edelberg progresses the treatment of intentional identity in [14,15,3]. He points out that intentional identity is asymmetric by *Arsky-Barsky sentences*, as follows:

Example 2. Arsky and Barsky investigate the apparent murder of Smith, and *they conclude that Smith was murdered by a single person, though they have no one in mind as a suspect.* A few days later, they investigate the apparent murder of a second person, Jones, and again *they conclude that Jones was murdered by a single person.* At this point, however, a disagreement between the two detectives arises. *Arsky thinks that the two murderers are completely unrelated, and that the person who murdered Smith, but not the one who murdererd Jones, is still in Chicago. Barsky, however, thinks that one and the same person murdered both Smith and Jones.* However, neither Smith nor Jones was really murdered. – [3], p.317; cf. [14]'s example 5 and [15]'s example 3.

In this example, (7a) is true, whereas (7b) is false:

(7) a. Arsky believes someone murdered Smith, and Barsky believes he murdered Jones.
  b. Barsky believes someone murdered Jones, and Arsky believes he murdered Smith.

But the difference between (7a) and (7b) is the precedence of the same conjuncts, i.e., a kind of *asymmetry.* That is, Arsky's suspected murderer of Smith is the same with Barsky's suspected murderer of Jones, whereas Barsky's suspected murderer of Jones is not the same with Arsky's suspected murderer of Jones, since Arsky believes that the murder of Smith and the murderer of Jones are not the same and Barsky believes the two persons are identical.

(7) is formulated by (8):

(8) $\Box_a(\exists x.\varphi(x)\wedge\psi(x))\wedge\Box_b(\varphi(x)\wedge\chi(x)) \not\models \Box_b(\exists x.\varphi(x)\wedge\chi(x))\wedge\Box_a(\varphi(x)\wedge\psi(x))$

Therefore, dynamic semantics of intentional identity must satisfy this asymmetry with respect to $\wedge$.

2.3 Cross Speaker Anaphora

Intentional Identity can be related to communication. Geach [1] saids:

> We very often take ourselves to know, when we hear the discourse of others, that they are meaning to refer to some one person or thing – and that, without ourselves being able to identify this person or thing, without our even being certain that there really is such a person or thing to identify. What we are claiming to know in such cases – let alone, whether the claim is justified – must remain obscure so long as intentional identity is obscure. – p. 151

Although he have not connected (9) and (10) to *intentional identity*, but Stalnaker [16] notices the fact about indefinite descriptions and their anaphoras in dialogues, as follows:

> Consider a famous example of Peter Strawson's (1952: 187):
> (9) X : *A man* jumped off a bridge.
> Y : *He* didn't jump, *he* was pushed.
> In this dialogue, X succeeds in making a certain man available for reference (by X) in the subsequent context, evevn though the statement that accomplished this was rejected by Y. It was the fact that the statement was made, and not the fact that it was acceptedm that did the job.
> It is the speaker introducing the individual into the context, and not necessarily the indicidual using pronoun, whose intentions are relevant to determining the referent of the pronoun. In Strawson's correction dialogue, Y's "he" refres to *the person X had in mind*. The dialogue might take the following return: after Y says "he didn't jump, he was pushed," X corrects the corrrections:
> (10) No, I didn't mean that man. I agree he was pushed, but *the one I was talking about really did jump.*
> Even though Y had *a particular person in mind who he believed to be the referent of his "he," it is the person X had in mind to whom he refers*, and that is why Y can be corrected. (But notice that even though Y's correction is rejected, he succeeded in changing the context by making a different man available for reference by X).
> – p. 110 (italicized by the author.)[1]

Francez and Berg [17] call anaphora which refers an indefinite description by the interlocutor *cross speaker anaphora*.[2]

Ludlow and Neale [19] tentatively analyze "A man" and "He" as a *D-type* term: the x:man x & x is said "x jumped off a bridge," which are rejected by Groenendijk and others [8].

Dekker and van Rooy [20] call a *Hob-Nob situation* such a discourse situation, i.e., a situation where "two or more agents discuss and exchange information about a subject they have agreed upon, when actually there need not be a real thing which they are talking about." The following example also shows that the conversants cannot identify the referent denoted by cross speaker anaphora and its antecedent expressed by an indefinite description "a magistrate of Gotham village":

Example 3.

(11) H: A magistrate from Gotham village has confessed batting youg girls.
N: They say he suspected them of socery. Do you know if more magistrates confessed?

---

[1] Peter Strawson (1952), *Introduction to Logical Theory*, Methuen, London.

[2] For more details about cross speaker ananphora or *cross-utterance anaphora* [8], see Ogata [18].

H: I don't know.
N: Do you know who he is?
H: No idea, he preferred to remain anonymous.

2.4 Weak Intentional Identity

There is a special case of intentional identity and cross speaker anaphora which realizes what I call *weak intentional idenity*, as follows:

Example 4. John, a reporter, is decribing an outbreak of witch mania in Gotham village as in (12a) by interviews to Hob and Nob seperately such as (12b). Hob believes the existence of at least two witchs such that one witch usually kills stock (conventionally let us call her Samantha) and another usually heals people's sick (converntionally let us call her Brigitta), whereas Nob believes the existence of a witch such that sometimes kills stock and sometimes heals people's sick (converntionally let us call her Zelda). John does not believe in the existence of any witch.

(12) a. Hob believes that a witch has blighted Bob's mare and Nob believes that she killed Cob's sow.
b. John: What happened? Hob: A witch has blighted Bob's mare.
John: I heard that a witch has blighted Bob's mare. How do you think?
Nob: Yeah, she killed Cob's cow.

This example is a special case of Geach's intentional identity, i.e., example 1. "A witch" corresponds to *Samantha*, whereas "she" corresponds to *Zelda*, but it is uncertain that Samantha *is identical to* Zelda. Nevertheless John believes that Zelda corresponds to Samantha in some sense and Nob believes that what is denoted by "a witch" is Zelda. I call this example *weak intentional identity.* In weak intentional identity, the property "the same witch" is attributed to some entity by John or Nob, but not semantical. Therefore, even in the semantics, the value of "a witch" can be different from the value of "she" or they can be identical. If the property "the same witch" is attributed to some enitity semantically, i.e., the value of "a witch" is identical to the value of "she," I call the intentional identity *strong intentional identity.* Even in the Arsky-Barsky sentences, Arsky and Barsky only believe there is the murderer of Smith and the murderer of Jones but do not know who are the murderers. Therefore, the Arsky-Barsky sentences are instances of weak intentional identity.

Since example 1 and (12a) are formulated as the same formula, the distinction of weak intentional identity and strong intentional identity cannot represented by formula but we have only to distinguish them in semantics. This suggests some proposals of semantics of intentional identity are inappropriate. For example, solution by *quasi-external anchors* [9], roughly as in (13a), cannot treat weak intentional identity, since the value of 'a witch' and 'she' must be the same in the total model, roughly as in (13b).

(13) a. $\{\Box_{Hob}(\varphi(x) \wedge \psi(x)), \Box_{Hob}\chi(y)\}; [y, x]$, where $[y, x]$ means that y is a *quasi-anchor* of x.

b. $\mathcal{M}, w, g \models [y, x]$ iff for each $h \sqsupseteq g$, if y occurs in φ and $\mathcal{M}, w, g \models \varphi$ and x occurs in ψ then there is $h' \sqsupseteq g$ such that $\mathcal{M}, w, h' \models \psi$ and $h(y) = h'(x)$, where $h \sqsupseteq g$ if $dom(g) \subseteq dom(h)$ and for all $x \in dom(g)$, $g(x) = h(x)$.

2.5 Multiple Intentional Identity

(14a) is an example of what I call *multiple intentional identity*, which is formulated as (14b) which is identical with the formula of the Hob-Nob sentence (3b):

Example 5. Hob saw Bob's mare dead. He thought that a witch has blighted the mare. He also thought that if the mare's heart is injured, the witch is Samantha, if the mare's head is injured, the witch is Brigitta, and if the mare's belly is injured, the witch is Beatrix. On the other hand, Nob saw Cob's sow dead. He thought that a witch killed the sow. Let us calls her Zelda. Then Nob met with Hob and they have started to talk:

(14) a. Hob: A witch has blighted Bob's mare.
Nob: Yeah, she killed Cob's sow.
b. $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x)) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))$

In this example, Hob's possible mental states can be classified into four types:

(a) Bob's mare's heart is injured. Samantha blighted Bob's mare.
(b) Bob's mare's head is injured. Brigitta blighted Bob's mare.
(c) Bob's mare's belly is injured. Beatrix blighted Bob's mare.

Therefore, the value of x occuring in $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x))$ in (a)-type worlds is Samantha, the value of x occuring in $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x))$ in (b)-type worlds is Brigitta, the value of x occuring in $\Box_{Hob}(\exists x.\varphi(x) \wedge \psi(x))$ in (c)-type worlds is Beatrix, the value of x occuring in $\Box_{Nob}(\varphi(x) \wedge \chi(x))$ in Nob's mental states is Zelda, and each of Samantha in (a)-type worlds, Brigitta in (b)-type worlds, Beatrix in (c)-type worlds have a kind of correspondence relation with Zelta in Nob's mental states.

Hence, the required correspondence relation can be multiple.

3 Edelberg's Semantics

Before discussing dynamic semantics of $\mathscr{L}(Ag, L)$, we must mention Edelberg's semantics of intentional identity, which is one of the basic semantics of intentional idenity.[3] Edelberg [3] introduces a formal semantics of $\mathscr{L}(Ag, L)$ to treat the Hob-Nob sentence and the Arsky-Barsky sentences based on a kind of counterpart relation, which is different from Lewis's [13] counterpart relation, and individual concepts called *objects* by Edelberg, i.e., function from possible worlds to possible individuals, as follows[4]:

[3] For a brief survey of semantics of intentional identity, see [18].

[4] This definition is a simplified version of the original, since his notions of *theory* and *context* are omitted and replaced by the concepts used in Kripke models in order to compare with other possible world semantics including my semantics.

Definition 1. *Let $W(\ni w, u)$ be a non-empty set of possible worlds and D a non-empty set of individuals. Then $\mathcal{M} = \langle W, D(\cdot), (R_A)_{A \in Ag}, \Theta, O, \approx, \preceq, I \rangle$ is a model, where*

1. $D(\cdot) : W \to \mathscr{P}(D) \setminus \emptyset$;
2. R_A *is a binary relation on* W *for each* $A \in Ag$;
3. $\Theta \subseteq \{\{u' \in W | (u, u') \in R_A\} | A \in Ag, u \in W\}$,
4. O *is a set of partial functions from* W *to* D *(i.e., a set of objects) such that*
   (a) *if* $o \in O$ *and* $o(w)$ *is defined, then* $o(w) \in D(w)$*, and*
   (b) *for each* $w \in W$ *and* $a \in D(w)$*, there is* $o \in O$ *such that* $o(w) = a$,
5. $\approx$ *is an equivalence relation defined on* O*, called a* counterpart relation *(*à la Ederberg*), and for all* $w \in W$*,* $o \approx o'$ *and* $o(w)$ *and* $o'(w)$ *are both defined, then* $o(w) = o'(w)$;
6. $\preceq$ *is a partial order on* O*, called a subobject relation, such that for each* $o_1, o_2, o_3, o_4 \in O$,
   (a) *if* $o_1 \preceq o_2$ *then* $dom(o_2) \subseteq dom(o_1)$,
   (b) *if* $o_1 \preceq o_2$ *and* $o_1 \preceq o_3$*, then either* $o_2 \preceq o_3$ *or* $o_3 \preceq o_2$ *or* $o_2 = o_3$,
   (c) *every maximal chain under* $\preceq$ *contains an upper and a lower bound, and*
   (d) *if there are* $w, u \in T \in \Theta$ *and* $v, s \in T' \in \Theta$ *such that* $o_1(w), o_2(u), o_3(v), o_4(s)$ *are defined,* $o_1 \approx o_3$*,* $o_2 \approx o_4$*, and* $o_1 \preceq o_2$*, then* $o_3 \preceq o_4$,
7. $I : Rel \to W \to \mathscr{P}(D^+)$ *such that if* $(a_1, \ldots, a_n) \in I(R)(w)$*,* $a_1, \ldots, a_n \in D(w)$.

Let $g : Var \to W \to D$ $(g \in \mathscr{G})$ *be a varible assignment such that* $g(x)(w) \in D(w)$*. Let* $\mathscr{M}$ *be the class of models.*

The interpretation $[\![,]\!] : \mathscr{M} \times W \times \mathscr{G} \times \mathscr{L}(Ag, L) \to \{1, 0\}$ *is defined by recursion* $\phi \in \mathscr{L}(Ag, L)$*, as follows:*

1. $[\![x_1 = x_2]\!](\mathcal{M}, w, g) = 1$ *iff there are* $o_1, o_2, o_3 \in O_{\mathcal{M}}$ *such that* $o_1(w), o_2(w), o_3(w)$ *are defined,* $g(x_1) = o_1$*,* $g(x_2) = o_2$*,* $o_1 \preceq o_3$*, and* $o_2 \preceq o_3$,
2. $[\![R(x_1, \ldots, x_n)]\!](\mathcal{M}, w, g) = 1$ *iff there are* $o_1, \ldots, o_n, o'_1, \ldots, o'_n, o''_1, \ldots, o''_n \in O_{\mathcal{M}}$ *such that* $o_j(w), o'_j(w), o''_i(w)$ *are defined,* $g(x_j) = o_j$*,* $o_j \preceq o'_j$*,* $o'_j \preceq o''_j$*, for each* $1 \leq j \leq n$*, and* $(o'_1(w), \ldots, o'_n(w)) \in I(R)(w)$,
3. $[\![\neg\phi]\!](\mathcal{M}, w, g) = 1$ *iff* $[\![\phi]\!](\mathcal{M}, w, g) = 0$,
4. $[\![\phi_1 \wedge \phi_2]\!](\mathcal{M}, w, g) = 1$ *iff* $[\![\phi_1]\!](\mathcal{M}, w, g) = [\![\phi_2]\!](\mathcal{M}, w, g) = 1$,
5. $[\![\exists x.\phi]\!](\mathcal{M}, w, g) = 1$ *iff there is* $o \in O_{\mathcal{M}}$ *such that* $[\![\phi]\!](\mathcal{M}, w, g[o/x]) = 1$,
6. $[\![\Box_A\phi]\!](\mathcal{M}, w, g, c) = 1$ *iff for all* u *such that* $(w, u) \in R_A$*,* $[\![\phi]\!](\mathcal{M}, w, g, c) = 1$.

In Edelberg's formulation, the Hob-Nob sentence is formulated as follows:

(15) $\exists x[(\Box_{Hob}(\varphi(x) \wedge \psi(x)) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))]$

By definition, (15) is true at world w_0 if and only if:

(16) There is an object o such that for every $w_{Hob} : w_0 R w_{Hob}$, there are objects o_{Hob} and o'_{Hob} such that $o \preceq o_{Hob}$, $o_{Hob} \preceq o'_{Hob}$ and $o_{Hob}(w_{Hob}) \in I(\varphi)(w_{Hob}) \cap I(\psi)(w_{Hob})$, and for every $w_{Nob} : w_0 R w_{Nob}$, there are objects o_{Nob} and o'_{Nob} such that $o \preceq o_{Nob}$, $o_{Nob} \preceq o'_{Nob}$ and $o_{Nob}(w_{Nob}) \in I(\varphi)(w_{Nob}) \cap I(\chi)(w_{Nob})$.

Suppose a 'multiple intentional identity' setting such as example 5. Nob believes Zelda, and Hob believes Samantha, Brigitta and Beatrix. Therefore, in any world u of Nob $o(u) = Zelda$, in any (a)-type world w_a of Hob, $o(w_a) = Samantha$, in any (b)-type world w_b of Hob, $o(w_b) = Brigitta$, in any (c)-type world w_c of Hob, $o(w_c) = Beatrix$. Like this, we can construct object o which satisfies (16). Therefore, even in a multiple intentional identity setting, Edelberg's semantics is not problematic. But then suppose that Nob has said that the witch is Beatrix, i.e., the whole formula in Edelberg's formulation is:

(17) $\exists x[\Box_{Hob}((\varphi(x) \wedge \psi)) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x)) \wedge \Box_{Nob}(x = \mathsf{Beatrix})]$

$\mathsf{Beatrix}$ is interpreted as an object B such that at least for any world w of Hob, $B(w) = Beatrix$. By definition of semantics of $=$, there is an object s such that $o \preceq s$ and $B \preceq s$. By definition of $\preceq$ and $B \preceq s$, at least for each world w_{Hob} of Hob, $s(w_{Hob}) = Beatrix$. By definition of $\preceq$ and $o \preceq s$, at least for each (a)-type world w'_{Hob} of Hob, $s(w'_{Hob}) = Samantha$. This implies contractorily that for each (a)-type world w of Hob, $s(w) = Samantha$ and $s(w) = Beatrix$. Therefore, objects *à la* Edelberg is problematic at least in a multiple intentional identity setting.

Furthermore, formula (15) is completely different from the literal translation of the Hob-Nob sentence to $\mathscr{L}(Ag, L)$ since the quantifiers's scopes are not the total sentences as in (15) but the embedded sentences as in (18), and this is problematic.

(18) $\Box_A(\exists x[\varphi(x) \wedge \psi(x)]) \wedge \Box_B(\varphi(x) \wedge \chi(x))$

To interpret (18) as (15) leads us to *dynamic interpretation* of $\mathscr{L}(Ag, L)$.

I will introduce a dynamic semantics of intentional identity including at least the Hob-Nob sentence and the Arsky-Barsky sentence *without* $\approx$ and $\preceq$, instead, by adding a kind of counterpart relation, and solves this problem in the next section.

4 A Dynamic Semantics

The semantics of $\mathscr{L}(Ag, L)$ is based on the Kripke semantics of *quantified modal logic* (QML) [6,7], which is a quadruple $\langle W, (R_A)_{A \in Ag}, D(\cdot), I\rangle$ (see definition 1), where $D(\cdot)$ requires no restriction such as the the condions which verifies the Barcan Formula or the Converse of the Barcan Formula or David Lewis's condition such that for every $w, u \in W$, $D(w) \neq D(u)$. As in dynamic semantics of modal subordination proposed by Ogata [2], variable assignment must be world-dependent, i.e.:

Definition 2 (World-dependent variable assignments). *Let W be a non-empty set of possible worlds, Var a non-empty set of individual variables, D a non-empty set of individual, $D(\cdot) : W \to \mathscr{P}(D)$ a resident function.*

Then, $g : Var \to W \to D$ is a function such that $g(x)(w) \in D(w)$, called a world-dependent variable assignment.

The class of world-dependent variable assignments over W*,* Var*,* $D(\cdot)$ *is denoted by* $\mathscr{G}(W, Var, D(\cdot))$*.*

$g[w : a/x]$ *is a world-dependent variable assignment defined by*

$$g[w : a/x](y)(u) = \begin{cases} a & \text{if } u = w \text{ and } y = x \\ g(y)(u) & \text{otherwise} \end{cases}$$

Next, I define a kind of counterpart relation, $\mathscr{C}$, which is weaker than Lewis's counterpart relation:

Definition 3. *A counterpart relation* $\mathscr{C} \subseteq D \times W \times D \times W$ *is a relation satisfying the following conditions:*

1. $\forall wu \in W \forall xy \in D.\mathscr{C}(x, w, y, u) \to x \in D(w) \& y \in D(u)$
2. $\forall w \in W \forall xy \in D(w).\mathscr{C}(x, w, y, w) \to x = y$
3. $\forall wu \in W \forall xyz \in D(w).\mathscr{C}(x, w, y, u) \& \mathscr{C}(x, w, z, u) \to y = z$

Now I define the dynamic semantics of $\mathscr{L}(Ag, L)$:

Definition 4. *Let* $\mathcal{M} = \langle W, (R_a)_{a \in Ag}, D(\cdot), I, \mathscr{C} \rangle$ *be a Kripke model with counterpart relation* $\mathscr{C}$ *defined by definition 3,* $A \in Ag$*,* $w \in W$*, and* $g \in \mathscr{G}(W, Var, D(\cdot))$ *and* $\ell \in Link(Var, \mathscr{L}(Ag, L), D)$*.*

Each quadruple $\sigma = \langle \mathcal{M}_\sigma, w_\sigma, g_\sigma \rangle$ *is called a state. The class of states is denoted by* $\mathcal{S}$*.*

Then the dynamic interpretation of each formula $\varphi \in \mathscr{L}(Ag, L)$ $[\![\varphi]\!] : \mathcal{S} \to \mathscr{P}(\mathcal{S})$*, is defined by recursion on the formulation of* φ*, as follows:*

1. $[\![x_1 = x_2]\!](\sigma) = \begin{cases} \{\sigma\} & \text{if } g_\sigma(x_1)(w_\sigma) = g_\sigma(x_2)(w_\sigma), \\ \emptyset & \text{otherwise} \end{cases}$
2. $[\![R(x_1, \ldots, x_n)]\!](\sigma) = \begin{cases} \{\sigma\} & \text{if } (g_\sigma(x_1)(w_\sigma), \ldots, g_\sigma(x_n)(w_\sigma)) \in I(R)(w) \\ \emptyset & \text{otherwise} \end{cases}$
3. $[\![\varphi_1 \wedge \varphi_2]\!](\sigma) = \bigcup_{\sigma' \in \in [\![\varphi_1]\!](\sigma)} [\![\varphi_2]\!](\sigma')$
4. $[\![\exists x.\varphi]\!](\sigma) = \{(\mathcal{M}_\sigma, w_\sigma, g_\sigma[w_\sigma : a/x]) | a \in D(w_\sigma), [\![\varphi]\!](\mathcal{M}_\sigma, w, g[w_\sigma : a/x]) \neq \emptyset, a \in D(w_\sigma)\}$
5. $[\![\Box_A \varphi]\!](\sigma) =$

$$\begin{cases} \{(\mathcal{M}, w, g \circ \varGamma_s^b) | (\mathcal{M}_\sigma, u, g_\sigma \circ \theta_u) \in Z_u, u \in X, (rg(\theta_u), u, b, s) \in \mathscr{C}, s \in W\} \\ \qquad \text{if for all } u \in X. [\![\varphi]\!](\mathcal{M}_\sigma, u, g_\sigma) \neq \emptyset \\ \emptyset \qquad \text{otherwise} \end{cases}$$

where $X = \{u | w_\sigma R_A u\}$, $Z_u = [\![\varphi]\!](\mathcal{M}_\sigma, u, g_\sigma)$, $\varGamma_t^b = \prod_{t \in Y} [t : dm(\theta_u)/b]$, $\theta_u = [u : a/x]$ *for some* $x \in Var$ *and* $a \in D(u)$, $rg(\theta_u) = a$, $dm(\theta_u) = x$, *for all* $u_1, \ldots, u_n \in W$ *such that* $dm(\theta_{u_1}) = \cdots = dm(\theta_{u_n})$, *for some* $u \in W$, $(rg(\theta_{u_1}), u_1, b_1, u), \ldots, (rg(\theta_{u_n}), u_n, b_n, u) \in \mathscr{C}$, $b_1 = \cdots = b_n$, *and* $Y = \{t \in W | (\mathcal{M}_\sigma, u, g_\sigma \circ \theta_u) \in Z_u, u \in X, (rg(\theta_u), u, b, t) \in \mathscr{C}, t \in W\}$

6. $[\![\neg\varphi]\!](\sigma) = \begin{cases} \{\sigma\} & \text{if } [\![\varphi]\!](\sigma) = \emptyset \\ \emptyset & \text{otherwise} \end{cases}$

- *φ is successful at σ if and only if $[\![\varphi]\!](\sigma) \neq \emptyset$.*
- *φ fails at σ if and only if $[\![\varphi]\!](\sigma) = \emptyset$.*
- *φ is semantically equivalent to ψ in $\mathcal{M}$ if and only if for any state σ of $\mathcal{M}$, $[\![\varphi]\!](\sigma) = [\![\psi]\!](\sigma)$.*

Let us see how this semantics works about the Hob-Nob sentence, the Arsky-Barsky sentences, the Hob-Nob sentence in a weak intentional identity setting, the Hob-Nob sentence in a multiple intentional identity setting, and a type of modal subordination.

Proposition 1 (The Hob-Nob sentence). [5] *Let $Var_1 = \{x\}$, $W_1 = \{w_1, w_2, w_3\}$, $R_{Hob} = \{(w_1, w_2)\}$, $R_{Nob} = \{(w_1, w_3)\}$, $D_1 = \{John, Hob, Nob, Sam, Bri\}$, $D_1(w_1) = \{John, Hob, Nob\}$, $D_1(w_2) = \{Hob, Sam, Bri\}$, $D_1(w_3) = \{Nob, Sam\}$, $\mathscr{C}_1 = \{(a, w, a, w) | a \in D_1(w), w \in W_1\} \cup \{(Sam, w_2, Sam, w_3)\} \cup \{(Sam, w_3, Bri, w_2)\}$, $I_1(\varphi)(w_1) = I_1(\psi)(w_1) = I_1(\chi)(w_1) = \emptyset$, $I_1(\varphi)(w_2) = \{Sam, Bri\}$, $I_1(\psi)(w_2) = \{Sam\}$, $I_1(\chi)(w_2) = \emptyset$, $I_1(\varphi)(w_3) = I_1(\chi)(w_3) = \{Sam\}$, $I_1(\psi)(w_3) = \emptyset$, $\mathcal{M}_1 = (W_1, (R_{Hob}, R_{Nob}), D_1(\cdot), I_1)$, $g_1(x)(w_1) = g_1(x)(w_2) = g_1(x)(w_3) = John$, and $\sigma_1 = (\mathcal{M}_1, w_1, g_1)$. Then,*

(1) $\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))$ is successful in σ_1, but
(2) $\exists x(\Box_{Hob}(\varphi(x) \wedge \psi(x)) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x)))$ fails at σ_1.

Verification.

$$\begin{aligned}
(1) \quad & [\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\sigma_1) \\
&= \bigcup_{\sigma' \in [\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x)))]\!](\sigma_1)} [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\sigma') \\
&= [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\mathcal{M}_1, w_1, g_1[w_2 : Sam/x][w_3 : Sam/x]) \\
&\because rg[w_2 : Sam/x] = Sam,\ (rg[w_2 : Sam/x], w_2, Sam, w_2) \in \mathscr{C}_1 \\
&\quad (rg[w_2 : Sam/x], w_2, Sam, w_3) \in \mathscr{C}_1 \\
&= \{(\mathcal{M}_1, w_1, g_1[w_2 : Sam/x][w_3 : Sam/x])\} \neq \emptyset.
\end{aligned}$$

(2) $[\![\exists x(\Box_{Hob}(\varphi(x) \wedge \psi(x)) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x)))]\!](\sigma_1) = \emptyset$ since $g_1(x)(w_1) = John$. □

Next, let us see the case of the Arsky-Barsky sentences by reusing the setting of the Hob-Nob sentence. Only the asymmetry has to be shown.

Proposition 2 (The Arsky-Barsky sentences). *Let $\mathcal{M}_2$ be $\mathcal{M}_1$. Then, (1) is not semantically equivalent to (2) in $\mathcal{M}_2$, where*

(1) $\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))$,
(2) $\Box_{Nob}(\exists x(\varphi(x) \wedge \chi(x))) \wedge \Box_{Hob}(\varphi(x) \wedge \psi(x))$.

[5] *Sam* is short for Samantha, *Bri* for Brigitta, and *Bea* for Beatrix.

Verification. We have only to show that there is a state σ such that $[\![(1)]\!](\sigma) \neq [\![(2)]\!](\sigma)$. Let us choose σ_1 defined in proposition 1. By the verification of proposition 1, $[\![(1)]\!](\sigma_1) = \{(\mathcal{M}_1, w_1, g_1[w_2 : Sam/x][w_3 : Sam/x])\}$. Now let see $[\![(2)]\!](\sigma_1)$.

$$
\begin{aligned}
& [\![\Box_{Nob}(\exists x(\varphi(x) \wedge \chi(x))) \wedge \Box_{Hob}(\varphi(x) \wedge \psi(x))]\!](\sigma_1) \\
= & \bigcup_{\sigma' \in [\![\Box_{Nob}(\exists x(\varphi(x) \wedge \chi(x)))]\!](\sigma_1)} [\![\Box_{Hob}(\varphi(x) \wedge \psi(x))]\!](\sigma') \\
= & [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\mathcal{M}_1, w_1, g_1[w_3 : Sam/x][w_2 : Bri/x]) \\
& \because rg[w_3 : Sam/x] = Sam,\ (rg[w_3 : Sam/x], w_3, Sam, w_3) \in \mathscr{C}_1 \\
& \quad (rg[w_3 : Sam/x], w_3, Bri, w_2) \in \mathscr{C}_1 \\
= & \emptyset \ \because g_1[w_3 : Sam/x][w_2 : Bri/x](x)(w_2) = Bri.
\end{aligned}
$$

Therefore, $[\![(1)]\!](\sigma_1) \neq [\![(2)]\!](\sigma_1)$. □

Proposition 3 (Weak Intentional Identity). *Let* $Var_3 = \{x\}$, $W_3 = \{w_1, w_2, w_3\}$, $R'_{Hob} = \{(w_1, w_2)\}$, $R'_{Nob} = \{(w_1, w_3)\}$, $D_3 = \{John, Hob, Nob, Sam, Bri, Zelda\}$, $D_3(w_1) = \{John, Hob, Nob\}$, $D_3(w_2) = \{Hob, Sam, Bri\}$, $D_3(w_3) = \{Nob, Zelda\}$, $\mathscr{C}_3 = \{(a, w, a, w) | a \in D_3(w), w \in W_3\} \cup \{(Sam, w_2, Zelda, w_3)\}$, $I_3(\varphi)(w_1) = I_3(\psi)(w_1) = I_3(\chi)(w_1) = \emptyset$, $I_3(\varphi)(w_2) = \{Sam, Bri\}$, $I_3(\psi)(w_2) = \{Sam\}$, $I_3(\chi)(w_2) = \emptyset$, $I_3(\varphi)(w_3) = I_3(\chi)(w_3) = I_3(\psi)(w_3) = \{Zelda\}$, $\mathcal{M}_3 = (W_3, (R'_{Hob}, R'_{Nob}), D_3(\cdot), I_3)$, $g_3(x)(w_1) = g_3(x)(w_2) = g_3(x)(w_3) = John$, *and* $\sigma_3 = (\mathcal{M}_3, w_1, g_3)$. *Then,*

$$\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))$$

is successful in σ_3.

Verification.

$$
\begin{aligned}
& [\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\sigma_3) \\
= & \bigcup_{\sigma' \in [\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x)))]\!](\sigma_3)} [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\sigma') \\
= & [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\mathcal{M}_3, w_1, g_3[w_2 : Sam/x][w_3 : Zelda/x]) \\
& \because rg[w_2 : Sam/x] = Sam,\ (rg[w_2 : Sam/x], w_2, Sam, w_2) \in \mathscr{C}_3 \\
& \quad (rg[w_2 : Sam/x], w_2, Zelda, w_3) \in \mathscr{C}_3 \\
= & \{(\mathcal{M}_3, w_1, g[w_2 : Sam/x][w_3 : Zelda/x])\} \neq \emptyset.
\end{aligned}
$$

□

Proposition 4 (Multiple Intentional Identity). *Let* $Var_4 = \{x\}$, $W_4 = \{w_1, w_2, w_3, w_4, w_5\}$, $R''_{Hob} = \{(w_1, w_2), (w_1, w_3), (w_1, w_4)\}$, $R''_{Nob} = \{(w_1, w_5)\}$, $D_4 = \{John, Hob, Nob, Sam, Bri, Bea, Zelda\}$, $D_4(w_1) = \{John, Hob, Nob\}$, $D_4(w_2) = D_4(w_3) = D_4(w_4) = \{Hob, Sam, Bri, Bea\}$, $D_4(w_5) = \{Nob, Zelda\}$, $\mathscr{C}_4 = \{(a, w, a, w) | a \in D(w), w \in W\} \cup \{(Sam, w_2, Zelda, w_5)\} \cup \{(Bri, w_3, Zelda, w_5)\} \cup$

$\{(Bea,w_4,Zelda,w_5)\}$,
$I_4(\varphi)(w_1) = I_4(\psi)(w_1) = I_4(\chi)(w_1) = \emptyset$,
$I_4(\varphi)(w_2) = I_4(\varphi)(w_3) = I_4(\varphi)(w_4) = \{Sam, Bri, Bea\}$,
$I_4(\psi)(w_2) = \{Sam\}$, $I_4(\psi)(w_3) = \{Bri\}$, $I_4(\psi)(w_4) = \{Bea\}$,
$I_4(\chi)(w_2) = I_4(\chi)(w_3) = I_4(\chi)(w_4) = \emptyset$,
$I_4(\varphi)(w_5) = I_4(\chi)(w_5) = I_4(\psi)(w_5) = \{Zelda\}$,
$\mathcal{M}_4 = (W_4, (R''_{Hob}, R''_{Nob}), D_4(\cdot), I_4)$, $g_4(x)(w_1) = g_4(x)(w_2) = g_4(x)(w_3) = g_4(x)(w_4) = g_4(x)(w_5) = John$, *and* $\sigma_4 = (\mathcal{M}_4, w_1, g_4)$. *Then,*

$$\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))$$

is successful in σ_4.

Verification.

$$\begin{aligned}
&[\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\sigma_4)\\
&= \bigcup_{\sigma' \in [\![\Box_{Hob}(\exists x(\varphi(x)\wedge\psi(x)))]\!](\sigma_4)} [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\sigma')\\
&= [\![\Box_{Nob}(\varphi(x) \wedge \chi(x))]\!](\mathcal{M}_4, w_1, g_4[w_2 : Sam/x][w_3 : Bri/x][w_4 : Bea/x][w_5 : Zelda/x])\\
&\because rg[w_2 : Sam/x] = Sam,\ (rg[w_2 : Sam/x], w_2, Sam, w_2) \in \mathscr{C}_4\\
&\quad rg[w_3 : Bri/x] = Bri,\ (rg[w_3 : Bri/x], w_3, Bri, w_3) \in \mathscr{C}_4\\
&\quad rg[w_4 : Bea/x] = Bea,\ (rg[w_4 : Bea/x], w_4, Bea, w_4) \in \mathscr{C}_4\\
&\quad (rg[w_2 : Sam/x], w_2, Zelda, w_5), (rg[w_3 : Bri/x], w_3, Zelda, w_5),\\
&\quad (rg[w_4 : Bea/x], w_4, Zelda, w_5) \in \mathscr{C}_4\\
&= \{(\mathcal{M}_4, w_1, g[w_2 : Sam/x][w_3 : Bri/x][w_4 : Bea/x][w_5 : Zelda/x])\} \neq \emptyset.
\end{aligned}$$

□

The dynamic semantics defined by definition 4 treats a fragment of modal subordination, which includes formulas with □. I call it the □-modal subordination, as a special case of intentional identity, as the following proposition shows.

Proposition 5 (□-Modal Subordination). *Let* $Var_5 = \{x\}$,
$W_5 = \{w_1, w_2, w_3\}$, $R'''_{Hob} = \{(w_1, w_2), (w_1, w_3)\}$,
$D_5 = \{John, Hob, Nob, Sam, Bri\}$, $D_5(w_1) = \{John, Hob, Nob\}$,
$D_5(w_2) = D_5(w_3) = \{Hob, Sam, Bri\}$,
$\mathscr{C}_5 = \{(a, w, a, w) | a \in D_3(w), w \in W_3\}$,
$I_5(\varphi)(w_1) = I_5(\psi)(w_1) = I_5(\chi)(w_1) = \emptyset$,
$I_5(\varphi)(w_2) = \{Sam, Bri\}$, $I_5(\psi)(w_2) = I_5(\chi)(w_2) = \{Bri\}$,
$I_5(\varphi)(w_3) = \{Sam, Bri\}$, $I_5(\psi)(w_3) = I_5(\chi)(w_3) = \{Sam\}$,
$\mathcal{M}_5 = (W_5, R'''_{Hob}, D_5(\cdot), I_5)$, $g_5(x)(w_1) = g_5(x)(w_2) = John$,
and $\sigma_5 = (\mathcal{M}_5, w_1, g_5)$. *Then*

$$\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Hob}(\varphi(x) \wedge \chi(x))$$

is successful in σ_5.

Verification.

$$
\begin{aligned}
& [\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x))) \wedge \Box_{Hob}(\varphi(x) \wedge \chi(x))]\!](\sigma_5) \\
= & \bigcup_{\sigma' \in [\![\Box_{Hob}(\exists x(\varphi(x) \wedge \psi(x)))]\!](\sigma_5)} [\![\Box_{Hob}(\varphi(x) \wedge \chi(x))]\!](\sigma') \\
= & [\![\Box_{Hob}(\varphi(x) \wedge \chi(x))]\!](\mathcal{M}_5, w_1, g_5[w_2 : Bri/x][w_3 : Sam/x]) \\
\because & \; rg[w_2 : Bri/x] = Bri, \; (rg[w_2 : Bri/x], w_2, Bri, w_2) \in \mathscr{C}_3 \\
& rg[w_3 : Sam/x] = Sam, \; (rg[w_3 : Sam/x], w_3, Sam, w_3) \in \mathscr{C}_3 \\
= & \{(\mathcal{M}_5, w_1, g[w_2 : Bri/x][w_3 : Sam/x])\} \neq \emptyset . \Box
\end{aligned}
$$

5 Conclusion

Firslty, this paper has surveyed the concepts of intentional identity. Secondly, this paper has clarified its relevance with cross speaker anaphora whose antecedent is an indefinite description. Thirdly, this paper has characterized weak intentional idenity which avoid some previous proposals of semantics of intentional identity. Fourthly, this paper has characterized multiple intentional identity which avoid Edelberg's semantics of intentional identity. Fifthly, this paper has proposed a dynamic semantics of intentional identity based on standard Kripke models of quantified modal logic, which is simpler than Edelberg's semantics. The semantics can treat not only weak intentional identity and multiple intentional identity but also $\Box$-modal subordination.

References

1. Geach, P.T.: Logic Matters. University of California Press, Barkeley (1972)
2. Ogata, N.: A dynamic semantics of modal subordination. In: Proceedings of the International Workshop of Logic and Engineering of Natural Language Semantics 2005, Tokyo, The Japanese Society of Artificial Intelligence (2005)
3. Edelberg, W.: A perspectivalist semantics for the attitudes. Noûs **29** (1995) 316–342
4. Groenendijk, J., Stokhof, M.: Dynamic predicate logic. Linguistics and Philosophy **14** (1991) 39–100
5. Kripke, S.A.: Semantical considerations on modal logic. Acta Philosophica Fennica **16** (1965) 83–94
6. Garson, J.W.: Quantification in modal logic. In Gabbay, D.M., Guenthner, F., eds.: Handbook of Philosophical Logic, 2nd Edition. Volume 3. Kluwer Academic Publisher, Dordrecht (2001) 267–323
7. Fitting, M., Mendelsohn, R.L.: First-Order Modal Logic. Kluwer Academic Publishers, Dordrecht (1999)
8. Groenendijk, J., Stokhof, M., Veltman, F.: Coreference and modality in the context of multi-speaker dialogue. In Kamp, H., Partee, B., eds.: Proceedings of the 1995 Workshop on Context Dependence, IMS, Stuttgart and UFAL, Pague (1997)
9. Asher, N.: A typology for attitude verbs and their anaphoric properties. Linguistics and Philosophy **10** (1987) 125–197

10. Dekker, P.: Meaning and use of indefinite expressions. Journal of Logic, Language, and Information **11** (2002) 141–192
11. van Rooij, R.: Attitudes and Changing Contexts. Springer Verlag, Berlin (2006)
12. Garson, J.W.: Quantification in modal logic. In Gabbay, D., Guenthner, F., eds.: Handbook of Philosophical Logic. Volume II. D. Reidel Publishing Company, Dordrecht (1984) 249–307
13. Lewis, D.: Counterpart theory and quantified modal logic. Journal of Philosophy **65** (1968) 113–126
14. Edelberg, W.: A new puzzle about intentional identity. Journal of Philosophical Logic **15** (1986) 1–25
15. Edelberg, W.: Intentional identity and the attitudes. Linguistics and Philosophy **15** (1992) 561–596
16. Stalnaker, R.C.: Context and Content. Oxford University Press, Oxford (1999)
17. Francez, N., Berg, J.: A multi-agent extension of DRT. Technical report, Technion - Computer Science Department - Technical Report LCL9409 (1995)
18. Ogata, N.: A dynamic semanics of intentional identity. In McCready, E., ed.: Proceedings of the Third International Workshop on Logic and Engineering of Natural Language Semantics, Tokyo (2006) 81–93
19. Ludlow, P., Neale, S.: Indefinite description: In defense of Russell. Linguistics and Philosophy **14** (1991) 171–202
20. Dekker, P., van Rooy, R.: Intentional identity and information exchange. In an T. Gamkrelidze, R.C., ed.: Proceedings of the Second Tbilisi Symposium on Language, Logic and Computation, Tbilisi State University, Tbilisi. (1998)

Prolegomena to General-Imaging-Based Probabilistic Dynamic Epistemic Logic

Satoru Suzuki

Department of General Education, Komazawa University,
Komazawa 1-23-1, Setagaya-ku, Tokyo, Japan
BXS05253@nifty.com

Abstract. In this paper, I propose a new version of probabilistic dynamic epistemic logic (GIPDEL) that is based on general imaging, and sketch the proof of soundness and completeness of this logic. The Monty Hall dilemma is a common topic in probabilistic dynamic epistemic logic. Using conditionalisation-based probabilistic dynamic epistemic logic (CPDEL), Kooi ([7]) supported the answer that I should switch my choice. However, it is acknowledged that this answer is counterintuitive. Using GIPDEL, I can support the answer that I do not have to switch my choice. Intuition would suggest this answer. Moreover, GIPDEL can give a plausible answer to a modified version of the Monty Hall dilemma to which CPDEL gives an extremely counterintuitive answer.

1 Introduction

Epistemic logic is the logic of knowledge. Dynamic epistemic logic is an extension of epistemic logic that can be used to reason about knowledge changes. Kooi combined probability with dynamic epistemic logic.[1] Because this logic is based on conditionalisation, I call it conditionalisation-based probabilistic dynamic epistemic logic (CPDEL). The Monty Hall dilemma is an open problem which is well-known among linguists, philosophers, psychologists, and logicians. This dilemma has the same structure as the problem of three prisoners. Nowadays this dilemma is a common topic in probabilistic dynamic epistemic logic. In [7], using CPDEL, Kooi supported the answer that I should switch my choice. However, it is acknowledged that this answer is counterintuitive.[2] Imaging is a method of changing probability functions Lewis proposed in [9]. Gärdenfors generalised this method in [3]. In this paper, I propose a new version of probabilistic dynamic epistemic logic that is based on general imaging and sketch the proof of soundness and completeness of this logic. I call this logic general-imaging-based probabilistic dynamic epistemic logic (GIPDEL). Using GIPDEL, I can support

[1] [7].

[2] As for the view that this answer is counterintuitive, refer to [5].

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 118–132, 2007.

the answer that I do not have to switch my choice. Intuition would suggest this answer. Moreover, GIPDEL can give a plausible answer to a modified version of the Monty Hall dilemma to which CPDEL gives an extremely counterintuitive answer.

2 Probabilistic Epistemic Logic PEL

2.1 Language

Fagin and Halpern gave the language of PEL $\mathcal{L}_{\mathsf{PEL}}$.[3]

Definition 1. *$\mathcal{L}_{\mathsf{PEL}}$ is defined in terms of a countable set S of sentential variables, a finite set A of agents, an epistemic operator $\mathbf{K}_a$ and a probability function symbol $\mathbf{P}_a$. The well-formed formulae of $\mathcal{L}_{\mathsf{PEL}}$ are given by the following rule:*

$$\phi ::= s \mid \top \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \mathbf{K}_a(\phi) \mid \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r,$$

where $s \in S, a \in A$ and $r_1, \ldots, r_n, r \in \mathbb{Q}$. $\sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i)$ is called a term of $\mathcal{L}_{\mathsf{PEL}}$, and $\sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r$ is called an a-probability formula of $\mathcal{L}_{\mathsf{PEL}}$. Let Φ_{PEL} denote the set of all well-formed formulae of $\mathcal{L}_{\mathsf{PEL}}$. Let $P_{a,\mathcal{L}_{\mathsf{PEL}}}$ denote the set of all a-probability formulae of $\mathcal{L}_{\mathsf{PEL}}$ and let $T_{\mathcal{L}_{\mathsf{PEL}}}$ denote the set of all terms of $\mathcal{L}_{\mathsf{PEL}}$.

$\bot, \vee, \rightarrow$ and $\leftrightarrow$ are introduced by the standard definitions. We use the usual abbreviations for readability.

2.2 Semantics

Fagin and Halpern defined a multi-agent structured Kripke model for PEL as follows:[4]

Definition 2. *A multi-agent structured Kripke model for* PEL *$\mathcal{M}_{\mathsf{PEL}}$ is an $(n+3)$-tuple $(W, \pi, R_{a_1}, \ldots, R_{a_n}, \mathcal{P})$, where W is a set of possible worlds, π is a truth assignment to each $s \in S$ for each $w \in W$, R_{a_i} is an equivalence relation on $W \times W$ for $i = 1, \ldots, n$, and $\mathcal{P}$ is a probability assignment that assigns to each $a \in A$ and each $w \in W$ a probability space $\mathcal{P}(a, w) := (W_{a,w}, \mathcal{F}_{a,w}, P_{a,w})$, where $W_{a,w} \subset W$ is the sample space, $\mathcal{F}_{a,w}$ is a σ-field of subsets of $W_{a,w}$, and $P_{a,w}$ is a probability measure defined on $\mathcal{F}_{a,w}$. We define $R_a(w_1)$ and $W_{a,w_1}(\phi)$ as follows:*

$$R_a(w_1) := \{w_2 : (w_1, w_2) \in R_a\},$$
$$W_{a,w_1}(\phi) := \{w_2 \in W_{a,w_1} : (\mathcal{M}_{\mathsf{PEL}}, w_2) \models \phi\}.$$

---

[3] [[1]: 343].
[4] [[1]: 343, 346].

Moreover, $\mathcal{M}_{\mathsf{PEL}}$ satisfies the following conditions:[5]

$$
\begin{array}{ll}
(\mathit{CONS}) & \begin{array}{l}\text{For all } a \in A \text{ and } w \in W,\\ \text{if } \mathcal{P}(a,w) = (W_{a,w}, \mathcal{F}_{a,w}, P_{a,w}), \text{ then } W_{a,w} \subset R_a(w),\end{array}\\
(\mathit{OBJ}) & \mathcal{P}(a_1, w) = \mathcal{P}(a_2, w) \text{ for all } a_1, a_2 \in A \text{ and } w \in W,\\
(\mathit{SDP}) & \begin{array}{l}\text{For all } a \in A \text{ and } w_1, w_2 \in W,\\ \text{if } w_2 \in R_a(w_1), \text{ then } \mathcal{P}(a, w_1) = \mathcal{P}(a, w_2),\end{array}\\
(\mathit{UNIF}) & \begin{array}{l}\text{For all } a \in A \text{ and } w_1, w_2 \in W\\ \text{if } \mathcal{P}(a, w_1) = (W_{a,w_1}, \mathcal{F}_{a,w_1}, P_{a,w_1}) \text{ and } w_2 \in W_{a,w_1},\\ \text{then } \mathcal{P}(a, w_2) = \mathcal{P}(a, w_1),\end{array}\\
(\mathit{MEAS}) & \begin{array}{l}\text{For all } a \in A \text{ and } w \in W \text{ and } \phi \in \Phi_{\mathcal{L}_{\mathsf{PEL}}},\\ W_{a,w}(\phi) \in \mathcal{F}_{a,w},\end{array}
\end{array}
$$

Let $\mathbb{M}_{\mathsf{PEL}}$ denote the class of all structured Kripke models for PEL.

(CONS) postulates that the belief system of an agent who places positive probability on an event he knows to be false is inconsistent. (OBJ) postulates the objectivity of probability assignments. (SDP) postulates that the choice of probability space is the same in all worlds the agent considers possible. (UNIF) postulates that we can partition $R_a(w)$ into subsets such that at every world in a given subset, the probability space is the same. (MEAS) postulates that all well-formed formulae define measurable sets.

Fagin and Halpern gave the following truth definition.[6]

Definition 3. *The notion of $\phi \in \Phi_{\mathcal{L}_{\mathsf{PEL}}}$ being true at $w \in W$ in $\mathcal{M}_{\mathsf{PEL}}$, in symbols $(\mathcal{M}_{\mathsf{PEL}}, w) \models \phi$ is inductively defined as follows:*

$$
\begin{array}{l}
(\mathcal{M}_{\mathsf{PEL}}, w) \models s \quad \text{iff} \quad \pi(w)(s) = \mathbf{true},\\
(\mathcal{M}_{\mathsf{PEL}}, w) \models \phi_1 \wedge \phi_2 \quad \text{iff} \quad (\mathcal{M}_{\mathsf{PEL}}, w) \models \phi_1 \quad \text{and} \quad (\mathcal{M}_{\mathsf{PEL}}, w) \models \phi_2,\\
(\mathcal{M}_{\mathsf{PEL}}, w) \models \neg\phi \quad \text{iff} \quad (\mathcal{M}_{\mathsf{PEL}}, w) \not\models \phi,\\
(\mathcal{M}_{\mathsf{PEL}}, w_1) \models K_a(\phi) \quad \text{iff} \quad (\mathcal{M}_{\mathsf{PEL}}, w_2) \models \phi \quad \text{for all } w_2 \in R_a(w_1),\\
(\mathcal{M}_{\mathsf{PEL}}, w) \models \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r \quad \text{iff} \quad \sum_{i=1}^{n} r_i P_{a,w}(W_{a,w}(\phi_i)) \geq r.
\end{array}
$$

If $(\mathcal{M}_{\mathsf{PEL}}, w) \models \phi$ for all $w \in W$, we write $\mathcal{M} \models \phi$ and say that ϕ is valid in $\mathcal{M}$. If ϕ is valid in all models in $\mathbb{M}_{\mathsf{PEL}}$, we write $\mathbb{M}_{\mathsf{PEL}} \models \phi$ and say that ϕ is valid with respect to $\mathbb{M}_{\mathsf{PEL}}$.

We define the probability of the semantic value of $\phi \in \Phi_{\mathcal{L}_{\mathsf{PEL}}}$ as follows:

Definition 4.

$$
P_{a,w_1}(W_{a,w_1}(\phi)) := \begin{cases} \sum_{w_2 \in W_{a,w_1}(\phi)} P_{a,w_1}(\{w_2\}) & \text{if } \vdash_{\mathsf{PEL}} \phi \not\leftrightarrow \bot,\\ 0 & \text{otherwise.} \end{cases}
$$

5 [[1]: 350–352].
6 [[1]: 343, 347].

2.3 Syntax

Fagin and Halpern gave the axiom system of PEL as follows:[7]

Definition 5.

- ***Axioms of*** PEL

(A1) *All tautologies of classical sentential logic,*
(A2) $\mathbf{K}_a(\phi_1 \to \phi_2) \to (\mathbf{K}_a(\phi_1) \to \mathbf{K}_a(\phi_2))$ (*K*),
(A3) $\mathbf{K}_a(\phi) \to \phi$ (*T*),
(A4) $\mathbf{K}_a(\phi) \to \mathbf{K}_a\mathbf{K}_a(\phi)$ (*Positive Introspection*),
(A5) $\neg\mathbf{K}_a(\phi) \to \mathbf{K}_a\neg\mathbf{K}_a(\phi)$ (*Negative Introspection*),
(A6) $\mathbf{P}_a(\phi) \geq 0$ (*Nonnegativity*),
(A7) $\mathbf{P}_a(\top) = 1$ (*Normalisation*),
(A8) $\mathbf{P}_a(\phi_1 \wedge \phi_2) + \mathbf{P}_a(\phi_1 \wedge \neg\phi_2) = \mathbf{P}_a(\phi_1)$ (*Additivity*),

(A9) $(\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \leftrightarrow (\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) + 0\mathbf{P}_a(\phi_{n+1}) \geq r)$
(Adding and Deleting 0 Terms),

(A10) $(\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \to (\sum_{i=1}^{n} r_{j_i}\mathbf{P}_a(\phi_{j_i}) \geq r)$
if $j_1, \ldots, j_n$ *is a permutation of* $1, \ldots, n$ (*Permutation*),

(A11) $(\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \wedge (\sum_{i=1}^{n} r'_i\mathbf{P}_a(\phi_i) \geq r') \to \sum_{i=1}^{n}(r_i + r'_i)\mathbf{P}_a(\phi_i) \geq (r + r')$
(Addition of Coefficients),

(A12) $(\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \leftrightarrow (\sum_{i=1}^{n} r' r_i\mathbf{P}_a(\phi_i) \geq r' r)$
if $r' > 0$ *(Multiplication of Nonzero Coefficients),*

(A13) $(t \geq r) \vee (t \leq r)$ *if* $t \in T_{\mathcal{L}_{\mathsf{PEL}}}$ *(Dichotomy),*
(A14) $(t \geq r_1) \to (t > r_2)$ *if* $t \in T_{\mathcal{L}_{\mathsf{PEL}}}$ *and* $r_1 > r_2$ *(Monotonicity),*

(A15) $\mathbf{K}_a(\phi) \to (\mathbf{P}_a(\phi) = 1)$ *(Correspondent to CONS),*
(A16) $(\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \to (\sum_{i=1}^{n} r_{j_i}\mathbf{P}_b(\phi_{j_i}) \geq r)$ *(Correspondent to OBJ),*
(A17) $\phi_1 \to (\mathbf{P}_a(\phi_1) = 1)$ *if* $\phi_1 \in P_{a,\mathcal{L}_{\mathsf{PEL}}}$ *or* ϕ_1 *is* $\neg\phi_2$ *such that* $\phi_2 \in P_{a,\mathcal{L}_{\mathsf{PEL}}}$
(Correspondent to UNIF),
(A18) $\phi_1 \to \mathbf{K}_a(\phi_1)$ *if* $\phi_1 \in P_{a,\mathcal{L}_{\mathsf{PEL}}}$ *or* ϕ_1 *is* $\neg\phi_2$ *such that* $\phi_2 \in P_{a,\mathcal{L}_{\mathsf{PEL}}}$
(Correspondent to SDP),

- ***Inference Rules of*** PEL

(R1) $\dfrac{\phi_1 \quad \phi_1 \to \phi_2}{\phi_2}$ (*Modus Ponens*),

(R2) $\dfrac{\phi}{\mathbf{K}_a(\phi)}$ (*Generalisation*),

(R3) $\dfrac{\phi_1 \leftrightarrow \phi_2}{\mathbf{P}_a(\phi_1) = \mathbf{P}_a(\phi_2)}$ (*Distributivity*).

[7] [[1]: 344, 353, 357].

If $\phi \in \Phi_{\mathsf{PEL}}$ is provable by (R1),(R2) or (R3) from (A1),(A2),(A3),(A4),(A5),(A6),(A7),(A8),(A9),(A10),(A11),(A12),(A13),(A14),(A15),(A16),(A17) or (A18), we write $\vdash_{\mathsf{PEL}} \phi$. The subsystem consisting of (A1),(A2),(A3),(A4),(A5),(R1) and (R2) is called S5.

2.4 Soundness and Completeness

We can prove the soundness theorem of PEL in the usual way.

Theorem 1. ***(Soundness)***
If $\vdash_{\mathsf{PEL}} \phi$, then $\mathbb{M}_{\mathsf{PEL}} \models \phi$.

Fagin and Halpern proved the completeness theorem of PEL.[8]

Theorem 2. ***(Completeness)***
If $\mathbb{M}_{\mathsf{PEL}} \models \phi$, then $\vdash_{\mathsf{PEL}} \phi$.

3 Conditionalisation-Based Probabilistic Dynamic Epistemic Logic CPDEL

3.1 Language

Kooi gave the language of CPDEL $\mathcal{L}_{\mathsf{CPDEL}}$.[9]

Definition 6. *$\mathcal{L}_{\mathsf{CPDEL}}$ is defined in terms of a countable set S of sentential variables, a finite set A of agents, an epistemic operator $\mathbf{K}_a$, a probability function symbol $\mathbf{P}_a$ and an update operator $[\ ]$. The well-formed formulae of $\mathcal{L}_{\mathsf{CPDEL}}$ are given by the following rule:*

$$\phi ::= s \mid \top \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \mathbf{K}_a(\phi) \mid \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r \mid [\phi_1]\phi_2,$$

where $s \in S, a \in A$ and $r_1, \ldots, r_n, r \in \mathbb{Q}$. $[\phi_1]\phi_2$ is interpreted as "ϕ_2 is the case after everyone simultaneously and commonly learns that ϕ_1 is the case." Let $\Phi_{\mathcal{L}_{\mathsf{CPDEL}}}$ denote the set of all well-formed formulae of $\mathcal{L}_{\mathsf{CPDEL}}$. Let $P_{a,\mathcal{L}_{\mathsf{CPDEL}}}$ denote the set of all a-probability formulae of $\mathcal{L}_{\mathsf{CPDEL}}$ and let $T_{\mathcal{L}_{\mathsf{CPDEL}}}$ denote the set of all terms of $\mathcal{L}_{\mathsf{CPDEL}}$.

3.2 Semantics

Based on Definition 2 and [[7]: 388] , we define an updated multi-agent structured Kripke model for CPDEL as follows:

8 [[1]: 357–359].
9 [[7]: 387].

Definition 7. *When a multi-agent structured Kripke model* $\mathcal{M}_{\mathsf{CPDEL}} := (W, \pi, R_{a_1}, \ldots, R_{a_n}, \mathcal{P})$ *and* $\phi_1 \in \Phi_{\mathcal{L}_{\mathsf{CPDEL}}}$ *are given, an updated multi-agent structured Kripke model for* CPDEL $\mathcal{M}_{\phi_1,\mathsf{CPDEL}}$ *is an* $(n+3)$*-tuple* $(W_{\phi_1}, \pi_{\phi_1}, R_{a_1,\phi_1}, \ldots, R_{a_n,\phi_1}, \mathcal{P}_{\phi_1})$*, where*

$$\begin{aligned}
&W_{\phi_1} = W,\\
&\pi_{\phi_1} = \pi,\\
&R_{a_1,\phi_1} := \{(w_1, w_2) \in R_{a_1} : (\mathcal{M}_{\mathsf{CPDEL}}, w_2) \models \phi_1\},\\
&\qquad\vdots\\
&R_{a_n,\phi_1} := \{(w_1, w_2) \in R_{a_n} : (\mathcal{M}_{\mathsf{CPDEL}}, w_2) \models \phi_1\},\\
&\mathcal{P}_{\phi_1} := (W_{a,w_1,\phi_1}, \mathcal{F}_{a,w_1,\phi_1}, P_{a,w_1,\phi_1}), \textit{where}\\
&\quad W_{a,w_1,\phi_1} := \begin{cases} W_{a,w_1} & \textit{if } P_{a,w_1}(W_{a,w_1}(\phi_1)) = 0,\\ \{w_2 \in W_{a,w_1} : (\mathcal{M}_{\mathsf{CPDEL}}, w_2) \models \phi_1\} & \textit{otherwise,} \end{cases}\\
&\quad \mathcal{F}_{a,w_1,\phi_1} \textit{ is a } \sigma\textit{-field of subsets of } W_{a,w_1,\phi_1},\\
&\quad P_{a,w_1,\phi_1}(W_{a,w_1,\phi_1}(\phi_2)) := \begin{cases} P_{a,w_1}(W_{a,w_1}(\phi_2)) & \textit{if } P_{a,w_1}(W_{a,w_1}(\phi_1)) = 0,\\ \dfrac{P_{a,w_1}(W_{a,w_1}(\phi_1 \wedge \phi_2))}{P_{a,w_1}(W_{a,w_1}(\phi_1))} & \textit{otherwise} \end{cases}\\
&\quad \textit{(Conditionalisation).}
\end{aligned}$$

Moreover, $\mathcal{M}_{\mathsf{CPDEL}}$ *and* $\mathcal{M}_{\phi_1,\mathsf{CPDEL}}$ *satisfies (CONS),(OBJ),(SDP),(UNIF) and (MEAS). Let* $\mathbb{M}_{\mathsf{CPDEL}}$ *denote the class of all structured Kripke models for* CPDEL*.*

Based on [[7]: 388], we give the following truth definition.

Definition 8. *The notion of* $\phi \in \Phi_{\mathcal{L}_{\mathsf{CPDEL}}}$ *being true at* $w \in W$ *in* $\mathcal{M}_{\mathsf{CPDEL}}$*, in symbols* $(\mathcal{M}_{\mathsf{CPDEL}}, w) \models \phi$ *is inductively defined as follows:*

$$\begin{aligned}
&(\mathcal{M}_{\mathsf{CPDEL}}, w) \models s \quad \textit{iff} \quad \pi(w)(s) = \mathbf{true},\\
&(\mathcal{M}_{\mathsf{CPDEL}}, w) \models \phi_1 \wedge \phi_2 \quad \textit{iff} \quad (\mathcal{M}_{\mathsf{CPDEL}}, w) \models \phi_1 \quad \textit{and} \quad (\mathcal{M}_{\mathsf{CPDEL}}, w) \models \phi_2,\\
&(\mathcal{M}_{\mathsf{CPDEL}}, w) \models \neg\phi \quad \textit{iff} \quad (\mathcal{M}_{\mathsf{CPDEL}}, w) \not\models \phi,\\
&(\mathcal{M}_{\mathsf{CPDEL}}, w_1) \models K_a(\phi) \quad \textit{iff} \quad (\mathcal{M}_{\mathsf{CPDEL}}, w_2) \models \phi \quad \textit{for all } w_2 \in R_a(w_1),\\
&(\mathcal{M}_{\mathsf{CPDEL}}, w) \models \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r \quad \textit{iff} \quad \sum_{i=1}^{n} r_i P_{a,w}(W_{a,w}(\phi_i)) \geq r,\\
&(\mathcal{M}_{\mathsf{CPDEL}}, w) \models [\phi_1]\phi_2 \quad \textit{iff} \quad (\mathcal{M}_{\phi_1,\mathsf{CPDEL}}, w) \models \phi_2.
\end{aligned}$$

If $(\mathcal{M}_{\mathsf{CPDEL}}, w) \models \phi$ *for all* $w \in W$*, we write* $\mathcal{M}_{\mathsf{CPDEL}} \models \phi$ *and say that* ϕ *is valid in* $\mathcal{M}$*. If* ϕ *is valid in all models in* $\mathbb{M}_{\mathsf{CPDEL}}$*, we write* $\mathbb{M}_{\mathsf{CPDEL}} \models \phi$ *and say that* ϕ *is valid with respect to* $\mathbb{M}_{\mathsf{CPDEL}}$*.*

3.3 Syntax

Besides (A1),(A2),(A3),(A4),(A5),(A6),(A7),(A8),(A9),(A10),(A11),(A12),(A13), (A14),(A15),(A16),(A17) and (A18), the axiom system of CPDEL has the following axioms based on [[7]: 395]:

(A19) $[\phi_1](\phi_2 \to \phi_3) \to ([\phi_1](\phi_2) \to [\phi_1](\phi_3))$ (K),
(A20) $\neg[\phi_1]\phi_2 \leftrightarrow [\phi_1]\neg\phi_2$ (Functionality),
(A21) $s \leftrightarrow [\phi]s$ (Atomic Permanence),
(A22) $[\phi_1]\mathbf{K}_a(\phi_2) \leftrightarrow \mathbf{K}_a(\phi_1 \to [\phi_1]\phi_2)$ (Knowledge Update),

(A23) $\mathbf{P}_a(\phi) > 0 \to (([\phi]\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \leftrightarrow (\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi \land [\phi]\phi_i) \geq r\mathbf{P}_a(\phi)))$ (Probability Update 1),

(A24) $\mathbf{P}_a(\phi) = 0 \to (([\phi]\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) \leftrightarrow (\sum_{i=1}^{n} r_i\mathbf{P}_a([\phi]\phi_i) \geq r))$ (Probability Update 2).

Besides (R1),(R2) and (R3), the axiom system of CPDEL has the following inference rule based on [[7]: 395]:

$$\text{(R4)} \quad \frac{\phi_2}{[\phi_1]\phi_2} \quad \text{(Generalisation).}$$

If $\phi \in \Phi_{\mathsf{CPDEL}}$ is provable by (R1),(R2),(R3) or (R4) from (A1),(A2),(A3),(A4), (A5),(A6),(A7),(A8),(A9), (A10),(A11), (A12),(A13), (A14),(A15),(A16), (A17), (A18),(A19),(A20),(A21),(A22),(A23) or (A24), we write $\vdash_{\mathsf{CPDEL}} \phi$.

3.4 Monty Hall Dilemma

The *Monty Hall dilemma* is an open problem which is well-known among linguists, philosophers, psychologists, and logicians. Nowadays this dilemma is a common topic in probabilistic dynamic epistemic logic. Kooi gave an answer to this dilemma in terms of CPDEL. This dilemma is stated as follows:[10]

Example 1. Suppose you're on a game show, and you're given the choice of three doors. Behind the door is a car, behind the others, goats. You pick a door, say number 1, and the host (Monty Hall), who knows what's behind the door, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?

3.5 Semantic Analysis of Monty Hall Dilemma in Terms of CPDEL

Assume that $\mathcal{M}_{\mathsf{CPDEL}} := (W, \pi, R_I, R_{MH}, \mathcal{P})$ is given. Let W be $\{w_1, w_2, w_3, w_4\}$ where w_1 is a world where there is a car behind the door 1 and MH opens the door 2, w_2 is a world where there is a car behind the door 1 and MH opens the door 3, w_3 is a world where there is a car behind the door 2 and MH opens the door 3, and w_4 is a world where there is a car behind the door 3 and MH opens the door 2.

Because $w_1 \in R_I(w_2)$ for all $w_1, w_2 \in W$, from (SDP) we have

$$\mathcal{P}(I, w_1) = \mathcal{P}(I, w_2) = \mathcal{P}(I, w_3) = \mathcal{P}(I, w_4).$$

[10] [[14]: 6].

Then we have, for all $w \in W$,

$$P_{I,w}(\{w_1\}) = P_{I,w}(\{w_2\}) = \frac{1}{6}, \quad P_{I,w}(\{w_3\}) = P_{I,w}(\{w_4\}) = \frac{1}{3}.$$

Let ϕ_3, ψ_1 and ψ_2 each denote the following sentence.

ϕ_3 := MH opens the door 3,
ψ_1 := there is a car behind the door 1,
ψ_2 := there is a car behind the door 2.

Because

$$P_{I,w}(W_{I,w}(\psi_1)) = P_{I,w}(\{w_1\}) + P_{I,w}(\{w_2\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

we have, for all $w \in W$,

$$(\mathcal{M}_{\mathsf{CPDEL}}, w') \models \mathbf{P}_I(\psi_1) = \tfrac{1}{3} \quad \text{for all } w' \in R_I(w).$$

So we have

$$\mathcal{M}_{\mathsf{CPDEL}} \models \mathbf{K}_I(\mathbf{P}_I(\psi_1) = \frac{1}{3}).$$

Moreover, because, for all $w \in W$,

$$\begin{aligned} P_{I,w,\phi_3}(W_{I,w,\phi_3}(\psi_1)) &= \frac{P_{I,w}(W_{I,w}(\phi_3 \wedge \psi_1))}{P_{I,w}(W_{I,w}(\phi_3))} = \frac{P_{I,w}(\{w_2\})}{P_{I,w}(\{w_2\}) + P_{I,w}(\{w_3\})} = \frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}} = \frac{1}{3}, \\ P_{I,w,\phi_3}(W_{I,w,\phi_3}(\psi_2)) &= \frac{P_{I,w}(W_{I,w}(\phi_3 \wedge \psi_2))}{P_{I,w}(W_{I,w}(\phi_3))} = \frac{P_{I,w}(\{w_3\})}{P_{I,w}(\{w_2\}) + P_{I,w}(\{w_3\})} = \frac{\frac{1}{3}}{\frac{1}{6}+\frac{1}{3}} = \frac{2}{3}, \end{aligned}$$

we have, for all $w \in W$,

$$\begin{aligned} (\mathcal{M}_{\phi_3,\mathsf{CPDEL}}, w') &\models \mathbf{P}_I(\psi_1) = \tfrac{1}{3} \quad \text{for all } w' \in R_{I,\phi_3}(w), \\ (\mathcal{M}_{\phi_3,\mathsf{CPDEL}}, w') &\models \mathbf{P}_I(\psi_2) = \tfrac{2}{3} \quad \text{for all } w' \in R_{I,\phi_3}(w). \end{aligned}$$

So we have the following results:

$$\mathcal{M}_{\mathsf{CPDEL}} \models [\phi_3]\mathbf{K}_I(\mathbf{P}_I(\psi_1) = \frac{1}{3}), \quad \mathcal{M}_{\mathsf{CPDEL}} \models [\phi_3]\mathbf{K}_I(\mathbf{P}_I(\psi_2) = \frac{2}{3}).$$

Therefore I should switch my choice. However, Ichikawa comments that there are overwhelmingly many examinees that support the answer that they do not have to switch their choices.[11] Isn't there a method of representing the belief change that supports this answer? There exists such a method. General imaging is a prime candidate for this task. I will propose a new version of probabilistic dynamic epistemic logic (GIPDEL) that is based on general imaging.

[11] [[5]: 27].

4 General-Imaging-Based Probabilistic Dynamic Epistemic Logic GIPDEL

4.1 Language

I give the language of GIPDEL $\mathcal{L}_{\mathsf{GIPDEL}}$.

Definition 9. *$\mathcal{L}_{\mathsf{GIPDEL}}$ is defined in terms of a finite set S of sentential variables, a finite set A of agents, an epistemic operator $\mathbf{K}_a$, a characteristic function symbol $\mathbf{G}_i$, a probability function symbol $\mathbf{P}_a$ and an update operator $[\ ]$. The well-formed formulae of $\mathcal{L}_{\mathsf{GIPDEL}}$ are given by the following rule:*

$$\phi ::= s \mid \top \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \mathbf{K}_a(\phi) \mid \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r \mid \sum_{i=1}^{n} r_i \mathbf{G}_i(\phi) = r \mid [\phi_1]\phi_2,$$

where $s \in S, a \in A$ and $r_1, \ldots, r_n, r \in \mathbb{Q}$. Let $\Phi_{\mathcal{L}_{\mathsf{GIPDEL}}}$ denote the set of all well-formed formulae of $\mathcal{L}_{\mathsf{GIPDEL}}$. Let $P_{a,\mathcal{L}_{\mathsf{GIPDEL}}}$ denote the set of all a-probability formulae of $\mathcal{L}_{\mathsf{GIPDEL}}$ and let $T_{\mathcal{L}_{\mathsf{GIPDEL}}}$ denote the set of all terms of $\mathcal{L}_{\mathsf{GIPDEL}}$.

4.2 Semantics

We prepare some concepts for the definition of general imaging. Let $w_2 \preceq_{w_1} w_3$ denote that $w_2 \in W$ is at least similar to $w_1 \in W$ as $w_3 \in W$ is. Let $w_2 \prec_{w_1} w_3$ denote that $w_2 \in W$ is more similar to $w_1 \in W$ than $w_3 \in W$ is. The comparative similarity system is defined as follows:[12]

Definition 10. *We posit an assignment of $\preceq_w$ and $R_a(w)$ to $w \in W$. Let us call such an assignment a comparative similarity system iff, for each $w_1, w_2, w_3, w_4 \in W$, the following six conditions hold.*

1. *$\preceq_{w_1}$ is transitive; that is, if $w_2 \preceq_{w_1} w_3$ and $w_3 \preceq_{w_1} w_4$, then $w_2 \preceq_{w_1} w_4$.*
2. *$\preceq_{w_1}$ is strongly connected; that is, for any w_2 and w_3, $w_2 \preceq_{w_1} w_3$ or $w_3 \preceq_{w_1} w_2$.*
3. *w_1 is self-accessible; that is, $w_1 \in R_a(w_1)$.*
4. *w_1 is strictly $\preceq_{w_1}$-minimal; that is, if any w_2 different from w_1, $w_1 \prec_{w_1} w_2$.*
5. *Inaccessible worlds are strictly $\preceq_{w_1}$-maximal; that is, if $w_3 \notin R_a(w_1)$, then for any w_2, $w_2 \preceq_{w_1} w_3$.*
6. *Accessible worlds are more similar to w_1 than inaccessible worlds; if $w_2 \in R_a(w_1)$ and $w_3 \notin R_a(w_1)$, $w_2 \prec_{w_1} w_3$.*

Imaging is a method of changing probability functions Lewis proposed in [9]. It produces minimal disturbance in the following sense:

> Imaging P on A gives a minimal revision in this sense: unlike all other revisions of P to make A certain, it involves no gratuitous movement of probability from worlds to dissimilar worlds.[13]

[12] This definition is based on [[8]: 48–49].
[13] [[9]: 148].

In order to define imaging, it is necessary to assume that, in a comparative similarity system, for any $w \in W$, there is a unique $w' \in W$ that is the most similar to w among the worlds where ϕ is true. General imaging is a version of imaging Gärdenfors proposed in [3]. In order to define general imaging, we have only to assume that, in a comparative similarity system, for any $w \in W$, there is at least one world that is the most similar to w among the worlds where ϕ is true. Let W_ϕ^w denote the set of all worlds that are the most similar to w among the worlds where ϕ is true. We define $g_{W_\phi^{w_1}} : W \longrightarrow \mathbb{R}$ as follows:

Definition 11.

$$g_{W_\phi^{w_1}}(w_2) := \begin{cases} \dfrac{1}{|W_\phi^{w_1}|} & \text{if } w_2 \in W_\phi^{w_1}, \\ 0 & \text{otherwise}, \end{cases} \quad \text{where} \sum_{w_2 \in W_\phi^{w_1}} g_{W_\phi^{w_1}}(w_2) = 1.$$

By means of Definition 4 and Definition 11, we define general imaging as follows:

Definition 12.

$$P^{\odot}_{a,w_1,\phi_1}(W_{a,w_1,\phi_1}(\phi_2)) := \begin{cases} \displaystyle\sum_{w_3 \in W_{a,w_1}(\phi_2)} \sum_{w_2 \in W_{a,w_1}} \left(g_{W^{w_2}_{a,w_1,\phi_1}}(w_3) \cdot P_{a,w_1}(\{w_2\}) \right) & \text{if } \vdash_{\mathsf{GIPDEL}} \phi_1 \not\leftrightarrow \bot \not\leftrightarrow \phi_2, \\ 0 & \text{if } \vdash_{\mathsf{GIPDEL}} \phi_1 \not\leftrightarrow \bot \leftrightarrow \phi_2, \\ 1 & \text{if } \vdash_{\mathsf{GIPDEL}} \phi_1 \leftrightarrow \bot. \end{cases}$$

We define an updated multi-agent structured Kripke model for GIPDEL as follows:

Definition 13. *When a multi-agent structured Kripke model $\mathcal{M}_{\mathsf{GIPDEL}} := (W, \preceq, \pi, R_{a_1}, \ldots, R_{a_n}, \tilde{\mathcal{P}})$, where W is a finite set of possible worlds and $\tilde{\mathcal{P}}$ is an extended probability assignment that assigns to each $a \in A$ and each $w \in W$ an extended probability space $\tilde{\mathcal{P}}(a, w) := (W_{a,w}, \mathcal{F}_{a,w}, 1, 0, P_{a,w})$, where $\mathcal{F}_{a,w}$ is a field of subsets of $W_{a,w}$, and $\phi_1 \in \Phi_{\mathcal{L}_{\mathsf{GIPDEL}}}$ are given, an updated multi-agent structured Kripke model for GIPDEL $\mathcal{M}_{\phi_1,\mathsf{GIPDEL}}$ is an $(n+4)$-tuple $(W_{\phi_1}, \preceq_{\phi_1}, \pi_{\phi_1}, R_{a_1,\phi_1}, \ldots, R_{a_n,\phi_1}, \tilde{\mathcal{P}}_{\phi_1})$, where*

$$\begin{aligned}
&W_{\phi_1} = W, \\
&\preceq_{\phi_1} = \preceq \quad (\textit{defined by Definition 10}), \\
&\pi_{\phi_1} = \pi, \\
&R_{a_1,\phi_1} := \{(w_1, w_2) \in R_{a_1} : (\mathcal{M}_{\mathsf{GIPDEL}}, w_2) \models \phi_1\}, \\
&\quad\vdots \\
&R_{a_n,\phi_1} := \{(w_1, w_2) \in R_{a_n} : (\mathcal{M}_{\mathsf{GIPDEL}}, w_2) \models \phi_1\}, \\
&\tilde{\mathcal{P}}_{\phi_1} := (W_{a,w_1,\phi_1}, \mathcal{F}_{a,w_1,\phi_1}, g_{W^{w_2}_{a,w_1,\phi_1}}, P_{a,w_1,\phi_1}), \textit{ where} \\
&\qquad W_{a,w_1,\phi_1} := \begin{cases} W_{a,w_1} & \textit{if } P_{a,w_1}(W_{a,w_1}(\phi_1)) = 0, \\ \{w_2 \in W_{a,w_1} : (\mathcal{M}_{\mathsf{GIPDEL}}, w_2) \models \phi_1\} & \textit{otherwise}, \end{cases} \\
&\qquad \mathcal{F}_{a,w_1,\phi_1} \textit{ is a field of subsets of } W_{a,w_1,\phi_1}, \\
&g_{W^{w_2}_{a,w_1,\phi_1}} \textit{ was defined by Definition 11}, \\
&\qquad P_{a,w_1,\phi_1}(W_{a,w_1,\phi_1}(\phi_2)) := \\
&\begin{cases} P_{a,w_1}(W_{a,w_1}(\phi_2)) & \textit{if } P_{a,w_1}(W_{a,w_1}(\phi_1)) = 0, \\ P^{\odot}_{a,w_1,\phi_1}(W_{a,w_1,\phi_1}(\phi_2)) \quad (\textit{defined by Definition 12}) & \textit{otherwise}, \end{cases}
\end{aligned}$$

Moreover, $\mathcal{M}_{\mathsf{GIPDEL}}$ *and* $\mathcal{M}_{\phi_1,\mathsf{GIPDEL}}$ *satisfies (CONS),(OBJ),(SDP),(UNIF) and (MEAS). Let* $\mathbb{M}_{\mathsf{GIPDEL}}$ *denote the class of all structured Kripke models for* GIPDEL*.*

I give the following truth definition.

Definition 14. *The notion of* $\phi \in \Phi_{\mathcal{L}_{\mathsf{GIPDEL}}}$ *being true at* $w \in W$ *in* $\mathcal{M}_{\mathsf{GIPDEL}}$*, in symbols* $(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \phi$ *is inductively defined as follows:*

$(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models s$ *iff* $\pi(w)(s) = \mathbf{true}$,
$(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \phi_1 \wedge \phi_2$ *iff* $(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \phi_1$ *and* $(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \phi_2$,
$(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \neg\phi$ *iff* $(\mathcal{M}_{\mathsf{GIPDEL}}, w) \not\models \phi$,
$(\mathcal{M}_{\mathsf{GIPDEL}}, w_1) \models K_a(\phi)$ *iff* $(\mathcal{M}_{\mathsf{GIPDEL}}, w_2) \models \phi$ *for all* $w_2 \in R_a(w_1)$,

$$(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \sum_{i=1}^{2^n} r_i \mathbf{G}_i(\phi) = r \quad \textit{iff} \quad \sum_{i=1}^{2^n} r_i \cdot \left\{ \begin{matrix} 1 \textit{ if } (\mathcal{M}_{\mathsf{GIPDEL}}, w_i) \models \phi \\ 0 \qquad \textit{otherwise} \end{matrix} \right\} = r,$$

(where $S := \{s_1, \ldots, s_n\}$ *and* $(\mathcal{M}_{\mathsf{GIPDEL}}, w_1) \models s_1 \& \ldots \& s_n$ *and* $(\mathcal{M}_{\mathsf{GIPDEL}}, w_2) \models \neg s_1 \& \ldots \& s_n$ *and* $\ldots$ *and* $(\mathcal{M}_{\mathsf{GIPDEL}}, w_{2^n-1}) \models \neg s_1 \& \ldots \neg s_{n-1} \& s_n$ *and* $(\mathcal{M}_{\mathsf{GIPDEL}}, w_{2^n}) \models \neg s_1 \& \ldots \& \neg s_{n-1} \& \neg s_n$*,)*

$$(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r \quad \textit{iff} \quad \sum_{i=1}^{n} r_i P_{a,w}(W_{a,w}(\phi_i)) \geq r,$$

$(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models [\phi_1]\phi_2$ *iff* $(\mathcal{M}_{\phi_1,\mathsf{GIPDEL}}, w) \models \phi_2$.

If $(\mathcal{M}_{\mathsf{GIPDEL}}, w) \models \phi$ *for all* $w \in W$*, we write* $\mathcal{M}_{\mathsf{GIPDEL}} \models \phi$ *and say that* ϕ *is valid in* $\mathcal{M}$*. If* ϕ *is valid in all models in* $\mathbb{M}_{\mathsf{GIPDEL}}$*, we write* $\mathbb{M}_{\mathsf{GIPDEL}} \models \phi$ *and say that* ϕ *is valid with respect to* $\mathbb{M}_{\mathsf{GIPDEL}}$*.*

4.3 Syntax

The axiom system of GIPDEL is the same as that of CPDEL, except that the former has the following axioms and inference rule, instead of (A23) and (A24).

(A25) $(\sum_{i=1}^{n} r_i \mathbf{G}_i(\phi) = r) \leftrightarrow (\sum_{i=1}^{n} r_i \mathbf{G}_i(\phi) + 0\mathbf{G}_{n+1}(\phi) = r)$
(Adding and Deleting 0 Terms),

(A26) $(\sum_{i=1}^{n} r_i \mathbf{G}_i(\phi) = r) \rightarrow (\sum_{i=1}^{n} r_{j_i} \mathbf{G}_{j_i}(\phi) = r)$
if $j_1, \ldots, j_n$ is a permutation of $1, \ldots, n$ (Permutation),

(A27) $(\sum_{i=1}^{n} r_i \mathbf{G}_i(\phi) = r) \wedge (\sum_{i=1}^{n} r'_i \mathbf{G}_i(\phi) = r') \rightarrow \sum_{i=1}^{n} (r_i + r'_i) \mathbf{G}_i(\phi) = (r + r')$
(Addition of Coefficients),

(A28) $(\sum_{i=1}^{n} r_i \mathbf{G}_i(\phi) = r) \leftrightarrow (\sum_{i=1}^{n} r' r_i \mathbf{G}_i(\phi) = r' r)$
if $r' > 0$ (Multiplication of Nonzero Coefficients),

(A29) $([\phi_1] \sum_{i=1}^{n} r_i \mathbf{G}_i(\phi_2) = r) \leftrightarrow (\sum_{i=1}^{n} r_i \mathbf{G}_i([\phi_1]\phi_2) = r)$
(Characteristic Function Update),

(A30) $([\phi] \sum_{i=1}^{n} r_i \mathbf{P}_a(\phi_i) \geq r) \leftrightarrow (\sum_{i=1}^{n} r_i \mathbf{P}_a([\phi]\phi_i) \geq r)$
(Probability Update),

$$\text{(A31)}\quad \begin{array}{l}(\mathbf{P}_a(s_1\&\ldots\&s_n) = r_1\&\mathbf{P}_a(\neg s_1\&\ldots\&s_n) = r_2\&\ldots\\ \&\mathbf{P}_a(\neg s_1\&\ldots\&\neg s_{n-1}\&s_n) = r_{2^n-1}\&\mathbf{P}_a(\neg s_1\&\ldots\&\neg s_{n-1}\&\neg s_n) = r_{2^n}\\ \&\mathbf{P}_a(\phi_1) = \sum_{i=1}^{2^n} r_i\mathbf{G}_i(\phi_1)) \to ([\phi_2]\mathbf{P}_a(\phi_1) = \sum_{i=1}^{2^n} r_i\mathbf{G}_i(\phi_1)),\\ \text{where } S := \{s_1,\ldots,s_n\} \quad \text{(Linearity of Probability Update)},\end{array}$$

$$\text{(R5)}\quad \frac{\phi_1 \leftrightarrow \phi_2}{\mathbf{G}_i(\phi_1) = \mathbf{G}_i(\phi_2)} \quad \text{(Distributivity)}.$$

If $\phi \in \Phi_{\mathsf{GIPDEL}}$ is provable by (R1),(R2),(R3),(R4) or (R5) from (A1),(A2),(A3), (A4),(A5), (A6),(A7), (A8), (A9),(A10), (A11), (A12),(A13), (A14),(A15), (A16), (A17),(A18),(A19),(A20),(A21),(A22),(A25),(A26),(A27),(A28),(A29),(A30) or (A31), we write $\vdash_{\mathsf{GIPDEL}} \phi$.

4.4 Soundness and Completeness

We can prove the soundness theorem of GIPDEL in the usual way.

Theorem 3. ***(Soundness)***
If $\vdash_{\mathsf{GIPDEL}} \phi$*, then* $\mathbb{M}_{\mathsf{GIPDEL}} \models \phi$.

In order to prove completeness of GIPDEL, we give a translation function $\tau : \mathcal{L}_{\mathsf{GIPDEL}} \to \mathcal{L}_{\mathsf{PEL}}$. Because completeness of PEL is proved, it suffices to show that every well-formed formula is equivalent to its translation in GIPDEL. This method is usual in the literature of dynamic epistemic logics.[14]

Definition 15. *A translation function* $\tau : \mathcal{L}_{\mathsf{GIPDEL}} \to \mathcal{L}_{\mathsf{PEL}}$ *is defined as follows:*

1. $\tau(s) = s$,
2. $\tau(\top) = \top$,
3. $\tau(\neg\phi) = \neg\tau(\phi)$,
4. $\tau(\phi_1 \wedge \phi_1) = \tau(\phi_1) \wedge \tau(\phi_2)$,
5. $\tau(\mathbf{K}_a(\phi)) = \mathbf{K}_a(\tau(\phi))$,
6. $\tau(\sum_{i=1}^{n} r_i\mathbf{G}_i(\phi) = r) = (\sum_{i=1}^{n} r_i\mathbf{G}_i(\tau(\phi)) = r)$,
7. $\tau(\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) = (\sum_{i=1}^{n} r_i\mathbf{P}_a(\tau(\phi_i)) \geq r)$,
8. $\tau([\phi]s) = s$,
9. $\tau([\phi_1]\neg\phi_2) = \neg\tau([\phi_1]\phi_2)$,
10. $\tau([\phi_1](\phi_2 \wedge \phi_2)) = \tau([\phi_1]\phi_2) \wedge \tau([\phi_1]\phi_3)$,
11. $\tau([\phi_1]\mathbf{K}_a(\phi_2)) = \mathbf{K}_a(\tau(\phi_1) \to \tau([\phi_1]\phi_2))$,
12. $\tau([\phi_1]\sum_{i=1}^{n} r_i\mathbf{G}_i(\phi_2) = r) = (\sum_{i=1}^{n} r_i\mathbf{G}_i(\tau([\phi_1]\phi_2)) = r)$,
13. $\tau([\phi]\sum_{i=1}^{n} r_i\mathbf{P}_a(\phi_i) \geq r) = (\sum_{i=1}^{n} r_i\mathbf{P}_a(\tau([\phi]\phi_i)) \geq r)$.

We can prove the following lemma.

[14] As for this method, refer to [[4]: 95-97], [[6]: 110-113] and [[7]: 396-397].

Lemma 1. *For every $\phi \in \Phi_{\mathsf{GIPDEL}}$, $\vdash_{\mathsf{GIPDEL}} \tau(\phi) \leftrightarrow \phi$.*

From Theorem 2 and Lemma 1, we can prove the completeness theorem of GIPDEL.

Theorem 4. ***(Completeness)***
If $\mathbb{M}_{\mathsf{GIPDEL}} \models \phi$, then $\vdash_{\mathsf{GIPDEL}} \phi$.

4.5 Semantic Analysis of Monty Hall Dilemma in Terms of GIPDEL

Assume that $\mathcal{M}_{\mathsf{GIPDEL}} := (W, \preceq, \pi, R_I, R_{MH}, \tilde{\mathcal{P}})$ is given. Then $\preceq$ enables us to assume that w_2 is the most similar to w_1 among the worlds where ϕ_3 is true, and to assume that w_2 and w_3 are the most similar to w_4 among the worlds where ϕ_3 is true.

Because, for all $w \in W$,

$$P_{I,w,\phi_3}(W_{I,w,\phi_3}(\psi_1)) = P^{\odot}_{I,w,\phi_3}(W_{I,w,\phi_3}(\psi_1)) = \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{2} \cdot \tfrac{1}{3} = \tfrac{1}{2},$$
$$P_{I,w,\phi_3}(W_{I,w,\phi_3}(\psi_2)) = P^{\odot}_{I,w,\phi_3}(W_{I,w,\phi_3}(\psi_2)) = \tfrac{1}{3} + \tfrac{1}{2} \cdot \tfrac{1}{3} = \tfrac{1}{2},$$

we have, for all $w \in W$,

$$(\mathcal{M}_{\phi_3,\mathsf{GIPDEL}}, w') \models \mathbf{P}_I(\psi_1) = \tfrac{1}{2} \quad \text{for all } w' \in R_{I,\phi_3}(w),$$
$$(\mathcal{M}_{\phi_3,\mathsf{GIPDEL}}, w') \models \mathbf{P}_I(\psi_2) = \tfrac{1}{2} \quad \text{for all } w' \in R_{I,\phi_3}(w).$$

So we have the following results:

$$\mathcal{M}_{\mathsf{GIPDEL}} \models [\phi_3]\mathbf{K}_I(\mathbf{P}_I(\psi_1) = \frac{1}{2}), \quad \mathcal{M}_{\mathsf{GIPDEL}} \models [\phi_3]\mathbf{K}_I(\mathbf{P}_I(\psi_2) = \frac{1}{2}).$$

Therefore I do not have to switch my choice. In this way, GIPDEL can represent the belief change that supports this answer.

5 Semantic Analyses of Modified Version of Monty Hall Dilemma

5.1 Modified Version of Monty Hall Dilemma

Ichikawa presented a modified version of the problem of three prisoners.[15] Likewise, we can state a modified version of the Monty Hall dilemma as follows:

Example 2. Suppose you're on a game show, and you're given the choice of three doors. Behind the door is a car, behind the others, goats. Somehow, you know the probability of there being a car behind the door 1 is $\frac{1}{4}$, the probability of there being one behind the door 2 is $\frac{1}{4}$ and the probability of there being one behind the door 3 is $\frac{1}{2}$. You only tentatively pick a door, say number 1, and the host (Monty Hall), who knows what's behind the door, opens another door, say number 2, which has a goat. Then what do you think the probability of there being a car behind the door 1?

15 [[5]: 29].

5.2 Semantic Analysis in Terms of CPDEL

Assume that $\mathcal{M}_{\mathsf{CPDEL}} := (W, \pi, R_I, R_{MH}, \mathcal{P})$ is given. Then we have, for all $w \in W$,

$$P_{I,w}(\{w_1\}) = P_{I,w}(\{w_2\}) = \frac{1}{8}, \quad P_{I,w}(\{w_3\}) = \frac{1}{4}, \quad P_{I,w}(\{w_4\}) = \frac{1}{2}.$$

Let ϕ_2 denote the sentence that MH opens the door 2.
Because, for all $w \in W$,

$$P_{I,w,\phi_2}(W_{I,w,\phi_2}(\psi_1)) = \frac{P_{I,w}(W_{I,w}(\phi_2 \wedge \psi_1))}{P_{I,w}(W_{I,w}(\phi_2))} = \frac{P_{I,w}(\{w_1\})}{P_{I,w}(\{w_1\}) + P_{I,w}(\{w_4\})} = \frac{\frac{1}{8}}{\frac{1}{8} + \frac{1}{2}} = \frac{1}{5},$$

we have, for all $w \in W$,

$$(\mathcal{M}_{\phi_2,\mathsf{CPDEL}}, w') \models \mathbf{P}_I(\psi_1) = \tfrac{1}{5} \quad \text{for all } w' \in R_{I,\phi_2}(w).$$

So we have the following result:

$$\mathcal{M}_{\mathsf{CPDEL}} \models [\phi_2]\mathbf{K}_I(\mathbf{P}_I(\psi_1) = \frac{1}{5}).$$

However, Ichikawa comments that there is no reason to believe that the probability that there is a car behind the door 1 decreases even though there may be reason to believe that the probability remains.[16]

5.3 Semantic Analysis in Terms of GIPDEL

Assume that $\mathcal{M}_{\mathsf{GIPDEL}} := (W, \preceq, \pi, R_I, R_{MH}, \tilde{\mathcal{P}})$ is given. Then $\preceq$ enables us to assume that w_1 is the most similar to w_2 among the worlds where ϕ_2 is true, and to assume that w_1 and w_4 are the most similar to w_3 among the worlds where ϕ_2 is true.
Because, for all $w \in W$,

$$P_{I,w,\phi_2}(W_{I,w,\phi_2}(\psi_1)) = P^{\odot}_{I,w,\phi_2}(W_{I,w,\phi_2}(\psi_1)) = \frac{1}{8} + \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8},$$

we have, for all $w \in W$,

$$(\mathcal{M}_{\phi_2,\mathsf{GIPDEL}}, w') \models \mathbf{P}_I(\psi_1) = \tfrac{3}{8} \quad \text{for all } w' \in R_{I,\phi_2}(w).$$

So we have the following result:

$$\mathcal{M}_{\mathsf{GIPDEL}} \models [\phi_2]\mathbf{K}_I(\mathbf{P}_I(\psi_1) = \frac{3}{8}).$$

GIPDEL can give a plausible answer also in Example 2.

[16] [[5]: 30].

6 Concluding Remarks

Kooi says:

> The Monty Hall dilemma is a puzzle for which intuitions fail many people. The best way to show that the counterintuitive results are correct is to use some formal method. CPDEL provides such a method.[17]

I do not agree with him. The Monty Hall dilemma does not illustrate that intuitive reasoning is sometimes incoherent with mathematical rules. But it shows that there are probability changes that cannot be represented in CPDEL.[18] In this paper, I proposed GIPDEL and showed that these probability changes can be represented in GIPDEL. The modified version of this dilemma supported this opinion.

References

1. Fagin, R. and Halpern, J.Y.: Reasoning about Knowledge and Probability. Journal of the Association for Computing Machinery **41** (1994) 340-367.
2. Fagin, R. and Halpern, J.Y.: Corrigendum: Reasoning about Knowledge and Probability. Journal of the Association for Computing Machinery **45** (1998) 214.
3. Gärdenfors, P.: Knowledge in Flux: Modeling the Dynamics of Epistemic States. The MIT Press, Cambridge, Mass. (1988)
4. Gerbrandy, J.: Bisimulation on Planet Kripke. ILLC Dissertation Series, ILLC, Amsterdam (1999).
5. Ichikawa, S.: Studies in Understanding Probability. (in Japanese). Kyoritsu, Tokyo (1998).
6. Kooi, B.P.: Knowledge, Chance, and Change. ILLC Dissertation Series, ILLC, Amsterdam (2003).
7. Kooi, B.P.: Probabilistic Dynamic Epistemic Logic. Journal of Logic, Language and Information **12** (2003) 381–408.
8. Lewis, D.: Counterfactuals. Basil Blackwell, Oxford (1973).
9. Lewis, D.: Probabilities of Conditionals and Conditional Probabilities. Philosophical Review **85** (1976) 297–315, rep. in Lewis, D.: *Philosophical Papers II.* Oxford UP, Oxford (1986) 133–152.
10. Suzuki, S.: On Imaging. (in Japanese). Philosophical Studies (University of Tokyo) **17** (1999) 24–40.
11. Suzuki, S.: Imaging Reconsidered. (in Japanese). Philosophical Studies (University of Tokyo) **18** (2000) 79–102.
12. Suzuki, S.: Belief Changes Represented by Imaging. (in Japanese). the Autumn Regular Meeting, the Japan Association for Philosophy of Science (2004).
13. van Benthem, J.: Conditional Probability Meets Update Logic. Journal of Logic, Language and Information **12** (2003) 409–421.
14. Vos Savant, M.: The Power of Logical Thinking. St. Martin's Press, New York (1996).

---

[17] [[7]: 403].

[18] I argued in [10] and [11] on the relation between imaging and AGM and argued in [12] on the relation between imaging and the diachronic Dutch book argument.

Logical Dynamics of Commands and Obligations

Tomoyuki Yamada

Division of Philosophy and Cultural Sciences,
Graduate School of Letters, Hokkaido University
Nishi 7, Kita 10, Kita-ku, Sapporo, Hokkaido, 060-0810, Japan
yamada@LET.hokudai.ac.jp

Abstract. In this paper, an eliminative command logic ECL introduced in Yamada [21] will be slightly refined into ECL II by allowing command terms and deontic operators to be indexed by a Cartesian Product of a given finite set of agents and a given finite set of command issuing authorities. Complete axiomatization and interesting validities will be presented, and a concrete example of a situation in which conflicting commands are given to the same agent by different authorities will be discussed extensively.

1 Introduction

Suppose your political guru commanded you to join an important political demonstration to be held in Tokyo next year. Unfortunately, it was to be held on the very same day on which an international one-day conference on logic was to be held in São Paulo, and the boss of your department had commanded you to attend that conference. It is possible for you to obey either command, but it is not possible for you to obey both as no available means of transportation is fast enough to enable you to join the demonstration in Tokyo and attend the conference in São Paulo on the same day. You have to decide which command to obey. But you are sure whichever command you may choose, you will regret not being able to obey the other.

In this paper, ECL (eliminative command logic) developed in Yamada [21] will be slightly refined in order to analyze adequately the situation you are supposed to be in after the issuance of your guru's command in the above example and the two commands that jointly brought about that situation. ECL is a variant of update logic, inspired by the development of DEL (dynamic epistemic logic) in Plaza [15], Groeneveld [9], Gerbrandy and Groeneveld [6], Gerbrandy [5], Baltag, Moss, & Solecki [2], and Kooi & van Benthem [11] among others. In DEL, the language of standard epistemic logic is utilized as the static base language on which its dynamic extensions are based. In the case of the logic of public announcements, for example, formulas of the static language of epistemic logic are used to describe situations before and after the announcements. Each situation is represented by an epistemic model, and public announcements are analyzed as events that update epistemic models. Thus in DEL, the dynamic extension of the epistemic logic, we have formulas of the form $[\phi!]K_i\psi$, which means that after every truthful public announcement to the effect that ϕ, an agent i knows that ψ.

The basic idea of ECL is to capture the workings of acts of commanding in terms of changes they bring about in the deontic status of possible courses of actions in the

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 133–146, 2007.

form of update logic by using multi-agent deontic logic instead of epistemic logic as a vehicle. Thus in ECL, we have formulas of the form $[!_i\phi]O_i\psi$, which means that after every successful acts of commanding an agent i to see to it that ϕ, it is obligatory upon i to see to it that ψ. Although ECL inherits various inadequacies from monadic deontic logic, some interesting principles are captured and seen to be valid nonetheless. [1]

Moreover, since in ECL, effects of acts of commanding are captured in terms of changes in deontic aspects of the situation, it enables us to isolate the effects of illocutionary acts of commanding from the perlocutionary consequences utterances may have upon actions and attitudes of addressees. Since Grice [8], lots of philosophers, linguists and computer scientists have tried to characterize uses of sentences in terms of utterers' intentions to produce various effects in addressees. But utterers' intentions usually go beyond illocutionary acts by involving reference to perlocutionary consequences, while illocutionary acts can be effective even if they failed to produce intended perlocutionary consequences. Thus, in the above example, even if you refuse to go to Brazil, it will not make your boss's command void. Your refusal would not constitute disobedience if it could make her command void. Her command is effective even if she failed to get you to form the intention to attend the conference; she has changed the deontic status of your possible alternative courses of actions.[2] Thus, ECL can be seen as an interesting case study from the point of view of speech act theory.[3]

In this paper, ECL will be slightly refined by allowing deontic operators and command type terms to be indexed by the Cartesian product of a given set of agents and a given set of command issuing authorities. In the resulting logic ECL II, the situation you are supposed to be in after the issuance of your guru's command in the above example will be represented as an obligational dilemma, so to speak.

2 The Static Base Logic MDL$^+$II

In Yamada [21], the language of command logic $\mathcal{L}_{\mathsf{CL}}$ was defined by dynamifying a static base language, the language of monadic deontic logic with an alethic modality $\mathcal{L}_{\mathsf{MDL}^+}$. $\mathcal{L}_{\mathsf{CL}}$ then was given a truth definition that incorporates an eliminative notion of acts of commanding, which leads to a version of eliminative command logic ECL.[4] In $\mathcal{L}_{\mathsf{MDL}^+}$, a formula of the form $O_i\varphi$ is used to represent the proposition that it is obligatory upon an agent i to see to it that φ. We refine this base language as follows:

[1] As is noted in [21], the use of monadic deontic language in ECL does not reflect any substantial theoretical commitment. It is used just to keep things as simple as possible at this early stage of the development of dynamic deontic logic. By dynamifying richer deontic languages interesting possibilities for further exploration into the logical dynamics of communicational acts will be opened up.

[2] There seems to be a growing recognition of the importance of such kind of institutional effects of speech acts among logicians and computer scientists recently. For example, more than one third of the authors explicitly talk about the *'count-as'* relation and/or institutional facts in their papers in DEON 2006 workshop. See Goble & Meyer [7].

[3] For more on the distinction between illocutionary and perlocutionary acts, see Lectures VIII–X of Austin [1].

[4] I refer to the language not as $\mathcal{L}_{\mathsf{ECL}}$ but as $\mathcal{L}_{\mathsf{CL}}$ as it might be given a truth definition that incorporates a non-eliminative notion of acts of commanding. For more on this, see Section 5.

Definition 1. *Take a countably infinite set* Aprop *of proposition letters, a finite set I of agents, and a finite set J of command issuing authorities, with p ranging over* Aprop, *i over I, and j over J. The refined multi-agent monadic deontic language $\mathcal{L}_{\mathsf{MDL^{+}II}}$ is given by:*

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid O_{(i,j)}\varphi$$

The set of all well formed formulas (sentences) of $\mathcal{L}_{\mathsf{MDL^{+}II}}$ is denoted by $S_{\mathsf{MDL^{+}II}}$ and operators of the form $O_{(i,j)}$ are called deontic operators. For each $i \in I$ and $j \in J$, we call a sentence (i, j)-free if no operators of the form $O_{(i,j)}$ occur in it. We call sentence alethic if no deontic operators occur in it, and boolean if no modal operators occur in it. For each $i \in I$ and $j \in J$, the set of all (i, j)-free sentences is denoted by $S_{(i,j)\text{-free}}$. The set of all alethic sentences and the set of all boolean sentences are denoted by S_{Aleth} and S_{Boole} respectively.

$\perp$, $\vee$, $\rightarrow$, $\leftrightarrow$, and $\Diamond$ are assumed to be introduced by standard definitions. We also abbreviate $\neg O_{(i,j)}\neg\varphi$ as $P_{(i,j)}\varphi$, and $O_{(i,j)}\neg\varphi$ as $F_{(i,j)}\varphi$. Note that Aprop $\subset S_{\text{Boole}} \subset S_{\text{Aleth}} \subset S_{(i,j)\text{-free}} \subset S_{\mathsf{MDL^{+}II}}$ for each $i \in I$ and $j \in J$.

A formula of the form $O_{(i,j)}\varphi$ is to be understood as meaning that it is obligatory upon an agent i with respect to an authority j to see to it that φ. In order to accommodate the distinction of authorities, we allow deontic accessibility relations to be indexed by $I \times J$. Thus we define:

Definition 2. *By an $\mathcal{L}_{\mathsf{MDL^{+}II}}$-model, I mean a quadruple $M = (W^M, R^M_A, R^M_D, V^M)$ where:*

(i) *W^M is a non-empty set (heuristically, of 'possible worlds')*

(ii) *$R^M_A \subseteq W^M \times W^M$*

(iii) *R^M_D is a function that assigns a subset $R^M_D(i, j)$ of R^M_A to each pair (i, j) of an agent $i \in I$ and an authority $j \in J$*

(iv) *V^M is a function that assigns a subset $V^M(p)$ of W^M to each proposition letter $p \in$ Aprop .*

We usually abbreviate $R^M_D(i, j)$ as $R^M_{(i,j)}$. Note that for each $i \in I$ and $j \in J$, $R^M_{(i,j)}$ is required to be a subset of R^M_A. Thus we assume that whatever is permitted is possible.

The truth definition for the formulas of $\mathcal{L}_{\mathsf{MDL^{+}II}}$ can be given in a standard way by associating the alethic modal operator $\Box$ with R^M_A and each deontic operator $O_{(i,j)}$ with $R^M_{(i,j)}$ as follows:

Definition 3. *Let M be an $\mathcal{L}_{\mathsf{MDL^{+}II}}$-model and w a point in M. If $p \in$ Aprop, $\varphi, \psi \in S_{\mathsf{MDL^{+}II}}$, and $(i, j) \in I \times J$, then:*

(a) *$M, w \models_{\mathsf{MDL^{+}II}} p$ iff $w \in V^M(p)$*

(b) *$M, w \models_{\mathsf{MDL^{+}II}} \top$*

(c) *$M, w \models_{\mathsf{MDL^{+}II}} \neg\varphi$ iff $M, w \not\models_{\mathsf{MDL^{+}II}} \varphi$ (i.e. it is not the case that $M, w \models_{\mathsf{MDL^{+}II}} \varphi$)*

(d) *$M, w \models_{\mathsf{MDL^{+}II}} (\varphi \wedge \psi)$ iff $M, w \models_{\mathsf{MDL^{+}II}} \varphi$ and $M, w \models_{\mathsf{MDL^{+}II}} \psi$*

(e) *$M, w \models_{\mathsf{MDL^{+}II}} \Box\varphi$ iff for every v such that $(w, v) \in R^M_A$, $M, v \models_{\mathsf{MDL^{+}II}} \varphi$*

(f) *$M, w \models_{\mathsf{MDL^{+}II}} O_{(i,j)}\varphi$ iff for every v such that $(w, v) \in R^M_{(i,j)}$, $M, v \models_{\mathsf{MDL^{+}II}} \varphi$.*

A formula φ is true in an $\mathcal{L}_{\mathsf{MDL^+II}}$-model M at a point w of M if $M, w \models_{\mathsf{MDL^+II}} \varphi$. We say that a set Σ of formulas of $\mathcal{L}_{\mathsf{MDL^+II}}$ is true in M at w, and write $M, w \models_{\mathsf{MDL^+II}} \Sigma$, if $M, w \models_{\mathsf{MDL^+II}} \psi$ for every $\psi \in \Sigma$. If $\Sigma \cup \{\varphi\}$ is a set of formulas of $\mathcal{L}_{\mathsf{MDL^+II}}$, we say that φ is a semantic consequence of Σ, and write $\Sigma \models_{\mathsf{MDL^+II}} \varphi$, if for every $\mathcal{L}_{\mathsf{MDL^+II}}$-model M and every point w such that $M, w \models_{\mathsf{MDL^+II}} \Sigma$, $M, w \models_{\mathsf{MDL^+II}} \varphi$. We say that a formula φ is valid, and write $\models_{\mathsf{MDL^+II}} \varphi$, if $\emptyset \models_{\mathsf{MDL^+II}} \varphi$.

The formulas of $\mathcal{L}_{\mathsf{MDL^+II}}$ can be used to talk about the situations before and after the issuance of a command. Consider again the previous example. Before the issuance of your boss's command, it was not obligatory upon you to attend the workshop in São Paulo, but after the issuance it became obligatory. Let p express the proposition that you will attend that workshop, and a and b represent you and your boss respectively. Furthermore, let (M, s) and (N, s) be the model world pairs that represent the situations before and after the issuance respectively. Then we have:

$$M, s \models_{\mathsf{MDL^+II}} \neg O_{(a,b)} p \tag{1}$$

$$N, s \models_{\mathsf{MDL^+II}} O_{(a,b)} p \ . \tag{2}$$

Thus the change brought about by your boss's command is captured in a sense by using formulas of $\mathcal{L}_{\mathsf{MDL^+II}}$.[5]

Now we define proof system for $\mathsf{MDL^+II}$:

Definition 4. *The proof system for* $\mathsf{MDL^+II}$ *contains the following axioms and rules:*

(Taut)	*all instantiations of propositional tautologies over the present language*
($\Box$-Dist)	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
($O_{(i,j)}$-Dist)	$O_{(i,j)}(\varphi \to \psi) \to (O_{(i,j)}\varphi \to O_{(i,j)}\psi)$ *for each* $(i, j) \in I \times J$
(Mix)	$P_{(i,j)}\varphi \to \Diamond\varphi$ *for each* $(i, j) \in I \times J$
(MP)	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$
($\Box$-Nec)	$\dfrac{\varphi}{\Box\varphi}$
($O_{(i,j)}$-Nec)	$\dfrac{\varphi}{O_{(i,j)}\varphi}$ *for each* $(i, j) \in I \times J$.

An $\mathsf{MDL^+II}$*-proof of a formula φ is a finite sequence of $\mathcal{L}_{\mathsf{MDL^+II}}$-formulas having φ as the last formula such that each formula is either an instance of an axiom, or it can be obtained from formulas that appear earlier in the sequence by applying a rule. If there is a proof of φ, we write $\vdash_{\mathsf{MDL^+II}} \varphi$. If $\Sigma \cup \{\varphi\}$ is a set of $\mathcal{L}_{\mathsf{MDL^+II}}$-formulas, we say that φ is deducible in* $\mathsf{MDL^+II}$ *from Σ and write $\Sigma \vdash_{\mathsf{MDL^+II}} \varphi$ if $\vdash_{\mathsf{MDL^+II}} \varphi$ or there are formulas $\psi_1, \ldots, \psi_n \in \Sigma$ such that $\vdash_{\mathsf{MDL^+II}} (\psi_1 \wedge \cdots \wedge \psi_n) \to \varphi$.*

[5] Note, however, that the change is talked about as a change in the meta-language, and not in $\mathcal{L}_{\mathsf{MDL^+II}}$. We will return to this point in the next section.

The above rules obviously preserve validity, and all the axioms are easily seen to be valid. Thus this proof system is sound.[6]

The completeness of this proof system can be proved in a completely standard way by building a canonical model. Thus we have:

Theorem 1 (Completeness of $\mathsf{MDL^+II}$). *Let $\Sigma \cup \{\varphi\}$ be a set of $\mathcal{L}_{\mathsf{MDL^+II}}$-formulas. Then, if $\Sigma \models_{\mathsf{MDL^+II}} \varphi$ then $\Sigma \vdash_{\mathsf{MDL^+II}} \varphi$.*

3 A Dynamic Extension ECL II

In the previous section, we have seen that $\mathcal{L}_{\mathsf{MDL^+II}}$-formulas can be used to describe the situations before and after the issuance of your boss's command. But the change brought about by your boss's command was not talked about as a change in $\mathcal{L}_{\mathsf{MDL^+II}}$ but in the meta-language, and it is simply impossible to use $\mathcal{L}_{\mathsf{MDL^+II}}$ to talk about the act of commanding which changed M into N. In Yamada [21], $\mathcal{L}_{\mathsf{MDL^+}}$ was extended to $\mathcal{L}_{\mathsf{CL}}$ by introducing operators indexed by the terms of the form $!_i\varphi$ in order to talk about effects of acts of commanding. Now we introduce expressions of the form $!_{(i,j)}\varphi$ for each pair $(i, j) \in I \times J$ in order to denote the type of an act of commanding in which an authority j commands an agent i to see to it that φ. The static base language $\mathcal{L}_{\mathsf{MDL^+II}}$ shall be expanded by introducing new modalities indexed by expressions of this form. Then, in the resulting language, the language $\mathcal{L}_{\mathsf{CL\,II}}$, of Command Logic, we have formulas of the form $[!_{(i,j)}\varphi]\psi$, which is to mean that after every successful act of commanding of type $!_{(i,j)}\varphi$, ψ holds. Thus we define:

Definition 5. *Take the same countably infinite set* Aprop *of proposition letters, the same finite set I of agents, and the same finite set J of command issuing authorities as before, with p ranging over* Aprop*, i over I, and j over J. The refined language of command logic $\mathcal{L}_{\mathsf{CL\,II}}$ is given by:*

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box\varphi \mid O_{(i,j)}\varphi \mid [\pi]\varphi$$
$$\pi ::= \,!_{(i,j)}\varphi$$

Terms of the form $!_{(i,j)}\varphi$ and operators of the form $[!_{(i,j)}\varphi]$ are called command type terms and command operators, respectively. The set of all well formed formulas of $\mathcal{L}_{\mathsf{CL\,II}}$ is referred to as $S_{\mathsf{CL\,II}}$, and the set of all the well formed command type terms as Com II.

$\bot$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\Diamond$, $P_{(i,j)}$, $F_{(i,j)}$, and $\langle !_{(i,j)}\varphi\rangle$ are assumed to be introduced by definition in the obvious way. Note that $S_{\mathsf{MDL^+II}} \subset S_{\mathsf{CL\,II}}$.

Then, the truth definition for the sentences of $\mathcal{L}_{\mathsf{CL\,II}}$ can be given with reference to $\mathcal{L}_{\mathsf{MDL^+II}}$-models as follows:

Definition 6. *Let $M = (W^M, R_A^M, R_D^M, V^M)$ be an $\mathcal{L}_{\mathsf{MDL^+II}}$-model, and $w \in W^M$. If $p \in$* Aprop*, $\varphi, \psi, \chi \in S_{\mathsf{CL\,II}}$, and $(i, j) \in I \times J$, then:*

[6] Strictly speaking, $O_{(i,j)}$-Nec is redundant since it is derivable. It is included here just to record the fact that $\mathsf{MDL^+II}$ is normal.

(a) $M, w \models_{\mathsf{ECL\,II}} p$ *iff* $w \in V^M(p)$

(b) $M, w \models_{\mathsf{ECL\,II}} \top$

(c) $M, w \models_{\mathsf{ECL\,II}} \neg\varphi$ *iff* $M, w \not\models_{\mathsf{ECL\,II}} \varphi$

(d) $M, w \models_{\mathsf{ECL\,II}} (\varphi \wedge \psi)$ *iff* $M, w \models_{\mathsf{ECL\,II}} \varphi$ *and* $M, w \models_{\mathsf{ECL\,II}} \psi$

(e) $M, w \models_{\mathsf{ECL\,II}} \Box\varphi$ *iff* $M, v \models_{\mathsf{ECL\,II}} \varphi$ *for every* v *such that* $(w, v) \in R_A^M$

(f) $M, w \models_{\mathsf{ECL\,II}} O_{(i,j)}\varphi$ *iff* $M, v \models_{\mathsf{ECL\,II}} \varphi$ *for every* v *such that* $(w, v) \in R_D^M(i, j)$

(g) $M, w \models_{\mathsf{ECL\,II}} [!_{(i,j)}\chi]\varphi$ *iff* $M_{!_{(i,j)}\chi}, w \models_{\mathsf{ECL\,II}} \varphi$,

where $M_{!_{(i,j)}\chi}$ *is an* $\mathcal{L}_{\mathsf{MDL^{+}II}}$*-model obtained from* M *by replacing* R_D^M *with the function* $R_D^{M_{!_{(i,j)}\chi}}$ *such that:*

(i) $R_D^{M_{!_{(i,j)}\chi}}(k, l) = R_D^M(k, l)$ *, for each* $(k, l) \in I \times J$ *such that* $(k, l) \neq (i, j)$

(ii) $R_D^{M_{!_{(i,j)}\chi}}(k, l) = \{(x, y) \in R_D^M(i, j) \mid M, y \models_{\mathsf{ECL\,II}} \chi\}$ *if* $(k, l) = (i, j)$.

We abbreviate $\{(x, y) \in R_D^M(i, j) \mid M, y \models_{\mathsf{ECL\,II}} \chi\}$ *as* $R_{(i,j)}^M \upharpoonright \chi^{\downarrow}$*. A formula* φ *is true in an* $\mathcal{L}_{\mathsf{MDL^{+}II}}$*-model* M *at a point* w *of* M *if* $M, w \models_{\mathsf{ECL\,II}} \varphi$*. We say that a set* Σ *of formulas of* $\mathcal{L}_{\mathsf{CL\,II}}$ *is true in* M *at* w*, and write* $M, w \models_{\mathsf{ECL\,II}} \Sigma$*, if* $M, w \models_{\mathsf{ECL\,II}} \psi$ *for every* $\psi \in \Sigma$*. If* $\Sigma \cup \{\varphi\}$ *is a set of formulas of* $\mathcal{L}_{\mathsf{CL\,II}}$*, we say that* φ *is a semantic consequence of* Σ*, and write* $\Sigma \models_{\mathsf{ECL\,II}} \varphi$*, if for every* $\mathcal{L}_{\mathsf{MDL^{+}II}}$*-model* M *and every point* w *of* M *such that* $M, w \models_{\mathsf{ECL\,II}} \Sigma$*,* $M, w \models_{\mathsf{ECL\,II}} \varphi$*. We say that a formula* φ *is valid, and write* $\models_{\mathsf{ECL\,II}} \varphi$*, if* $\emptyset \models_{\mathsf{ECL\,II}} \varphi$*.*

The crucial clause here is (g). The truth condition of $[!_{(i,j)}\chi]\varphi$ at w in M is defined in terms of the truth condition of φ at w in the updated model $M_{!_{(i,j)}\chi}$. Let a pair (w, v) of points be referred to as the R-arrow from w to v if it is in an accessibility relation R. Then the workings of an act of commanding of the form $!_{(a,b)}\varphi$ can be captured by saying that it eliminate every $R_{(a,b)}^M$-arrow (w, v) such that $M, v \not\models_{\mathsf{ECL\,II}} \varphi$ from $R_{(a,b)}^M$ if it is performed at some world in M. Note that the only possible difference between $M_{!_{(i,j)}\chi}$ and M consists in the possible difference between $R_{(i,j)}^M \upharpoonright \chi^{\downarrow}$ and $R_{(i,j)}^M (= R_D^M(i, j))$. All the other constituents are common to them. Since we always have $R_{(i,j)}^M \upharpoonright \chi^{\downarrow} \subseteq R_{(i,j)}^M$, we also have $R_{(i,j)}^M \upharpoonright \chi^{\downarrow} \subseteq R_A^M$ as required in the clause (iii) of Definition 2. Thus $M_{!_{(i,j)}\chi}$ is guaranteed to be an $\mathcal{L}_{\mathsf{MDL^{+}II}}$-model. [7]

Note also that each of the remaining clauses in the above definition reproduces the corresponding clause in the truth definition for $\mathcal{L}_{\mathsf{MDL^{+}II}}$. Obviously, we have:

[7] If we impose additional frame conditions on models by adding extra axioms to the proof system of MDL⁺II, however, the above model updating operation may yield models which violate these conditions. Thus we will have to impose matching constraints upon updating operation, but it might not always be possible. For example, see the discussion on Dead End principle in Section 5. Model updating operations are used and studied in dynamic epistemic logics and a more general discussion can be found in van Benthem & Liu [3].

Corollary 1. *Let M be an $\mathcal{L}_{\mathsf{MDL^{+}II}}$-model and w a point of M. Then for any $\varphi \in S_{\mathsf{MDL^{+}II}}$, $M, w \models_{\mathsf{ECL\,II}} \varphi$ iff $M, w \models_{\mathsf{MDL^{+}II}} \varphi$.*

The following corollary can be proved by induction on the length of ψ:

Corollary 2. *Let ψ be an (i, j)-free formula. Then, for any $\varphi \in S_{\mathsf{CL\,II}}$, $M, w \models_{\mathsf{ECL\,II}} \psi$ iff $M_{!_{(i,j)}\varphi}, w \models_{\mathsf{ECL\,II}} \psi$.*

One of the things this corollary means is that acts of commanding do not affect so-called brute facts and alethic possibilities in any direct way.

Consider the previous example again. Let p, a and b be understood as before. Let c represent your political guru, and let q express the proposition that you will attend the political demonstration c mentioned. In the situation before the issuance of your boss's command, it was not obligatory upon you to see to it that p, nor was it so to see to it that $\neg p$. Let (M, s) represent that situation as before. Then we have:

$$M, s \models_{\mathsf{ECL\,II}} \neg O_{(a,b)} p \wedge \neg O_{(a,b)} \neg p \ . \tag{3}$$

This means that we have:

$$M, s \models_{\mathsf{ECL\,II}} P_{(a,b)} \neg p \wedge P_{(a,b)} p \ . \tag{4}$$

As we have assumed that whatever is permitted is possible, we have:

$$M, s \models_{\mathsf{ECL\,II}} \Diamond \neg p \wedge \Diamond p \ . \tag{5}$$

In this situation, we also have:

$$M, s \models_{\mathsf{ECL\,II}} \neg O_{(a,c)} q \wedge \neg O_{(a,c)} \neg q \ . \tag{6}$$

Hence we have:

$$M, s \models_{\mathsf{ECL\,II}} \Diamond p \wedge \Diamond q \ . \tag{7}$$

Since $\Diamond p$ and $\Diamond q$ are (a, b)-free and (a, c)-free, Corollary 2 guarantees:

$$(M_{!(a,b)p})_{!(a,c)q}, s \models_{\mathsf{ECL\,II}} \Diamond p \wedge \Diamond q \ . \tag{8}$$

Thus, Corollary 2 enables us to capture, at least partially, unchanging aspects of the changing situations.

As regards the changing aspects, the semantics defined above validates:

$$M, s \models_{\mathsf{ECL\,II}} [!_{(a,b)} p] O_{(a,b)} p \ . \tag{9}$$

Your boss's command eliminates all the $R^M_{(a,b)}$-arrows (w, v) such that $M, v \not\models_{\mathsf{ECL\,II}} p$, and consequently we have:

$$M_{!_{(a,b)}p}, s \models_{\mathsf{ECL\,II}} O_{(a,b)} p \ . \tag{10}$$

In fact this is an instantiation of the following principle:

Proposition 1 (CUGO Principle). *If* $\varphi \in S_{(i,j)\text{-free}}$, *then* $\models_{\mathsf{ECL\,II}} [!_{(i,j)}\varphi]O_{(i,j)}\varphi$.

CUGO Principle here characterizes, at least partially, the workings of acts of commanding; though not without exceptions, commands usually generate obligations. The restriction on φ here is motivated by the fact that the truth of φ at a point v in M does not guarantee the truth of φ at v in $M_{!_{(i,j)}\varphi}$ if φ involves deontic modalities for the pair (i, j). For example, $[!_{(i,j)}P_{(i,j)}q]O_{(i,j)}P_{(i,j)}q$ is not valid.[8]

Let's go back to the example. As $O_{(a,b)}p$ is (a, c)-free, Corollary 2 guarantees:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} O_{(a,b)}p \ . \tag{11}$$

As another instantiation of CUGO Principle, we have:

$$M_{!_{(a,b)}p}, s \models_{\mathsf{ECL\,II}} [!_{(a,c)}q]O_{(a,c)}q \ . \tag{12}$$

By definition, this is equivalent to:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} O_{(a,c)}q \ . \tag{13}$$

Hence we have:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} O_{(a,b)}p \wedge O_{(a,c)}q \ . \tag{14}$$

Thus, it is obligatory upon you to see to it that p with respect to your boss while it is obligatory upon you to see to it that q with respect to your guru.

Unfortunately, however, as we have supposed earlier, no means of transportation that is fast enough to enable you to join the demonstration in Tokyo and attend the conference in São Paulo on the same day happened to be available. It is not possible for you to obey both commands. One possible way of expressing this supposition is to assume:

$$M, s \models_{\mathsf{ECL\,II}} \neg\Diamond(p \wedge q) \ . \tag{15}$$

Then, as $\neg\Diamond(p \wedge q)$ is (a, b)-free and (a, c)-free, Corollary 2 guarantees:

$$(M_{!(a,b)p})_{!(a,c)q}, s \models_{\mathsf{ECL\,II}} \neg\Diamond(p \wedge q) \ . \tag{16}$$

Thus, if we accept (15), we will have:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} O_{(a,b)}p \wedge O_{(a,c)}q \wedge \neg\Diamond(p \wedge q) \ . \tag{17}$$

If you obey your boss's command you will disobey your guru's command; if you obey your guru's command you will disobey your boss's command. You are in an obligational dilemma. As $p \wedge q$ is not a logical contradiction, there may be a possible situation in which you could obey both commands, but unfortunately it is not the situation you are in.

Whether this is really a good way of representing the situation you are in, however, doesn't seem to be obvious, since the impossibility involved in this situation is not an alethic (*i.e.* metaphysical) impossibility. If a sufficiently fast means of transportation were available, it would be possible for you to obey both commands. I will return to this point after looking at an obligational dilemma of a different kind.

[8] For more on this point, see Yamada [21].

4 Proof System for ECL II

Now we define proof system for ECL II.

Definition 7. *The proof system for* ECL II *contains all the axioms and all the rules of the proof system for* $\mathsf{MDL}^{+}\mathsf{II}$, *and in addition the following reduction axioms and rules:*

(RAt)	$[!_{(i,j)}\varphi]p \leftrightarrow p$ *where* $p \in Aprop$	(Reduction to Atoms)
(RVer)	$[!_{(i,j)}\varphi]\top \leftrightarrow \top$	(Reduction to Verum)
(FUNC)	$[!_{(i,j)}\varphi]\neg\psi \leftrightarrow \neg[!_{(i,j)}\varphi]\psi$	(Functionality)
($[!_{(i,j)}\varphi]$-Dist)	$[!_{(i,j)}\varphi](\psi \wedge \chi) \leftrightarrow ([!_{(i,j)}\varphi]\psi \wedge [!_{(i,j)}\varphi]\chi)$	($[!_{(i,j)}\varphi]$-Distribution)
(RAleth)	$[!_{(i,j)}\varphi]\Box\psi \leftrightarrow \Box[!_{(i,j)}\varphi]\psi$	(Reduction for Alethic Modality)
(RObl)	$[!_{(i,j)}\varphi]O_{(i,j)}\psi \leftrightarrow O_{(i,j)}(\varphi \rightarrow [!_{(i,j)}\varphi]\psi)$	(Reduction for Obligation)
(RInd)	$[!_{(i,j)}\varphi]O_{(k,l)}\psi \leftrightarrow O_{(k,l)}[!_{(i,j)}\varphi]\psi$ *where* $(i,j) \neq (k,l)$	(Independence)
($[!_{(i,j)}\varphi]$-Nec)	$\dfrac{\psi}{[!_{(i,j)}\varphi]\psi}$ *for each* $(i,j) \in I \times J$.	($[!_{(i,j)}\varphi]$-necessitation)

An ECL II*-proof of a formula* φ *is a finite sequence of* $\mathcal{L}_{\mathsf{CL\,II}}$*-formulas having* φ *as the last formula such that each formula is either an instance of an axiom, or it can be obtained from formulas that appear earlier in the sequence by applying a rule. If there is a proof of* φ*, we write* $\vdash_{\mathsf{ECL\,II}} \varphi$*. If* $\Sigma \cup \{\varphi\}$ *is a set of* $\mathcal{L}_{\mathsf{CL\,II}}$*-formulas, we say that* φ *is deducible in* ECL II *from* Σ *and write* $\Sigma \vdash_{\mathsf{ECL\,II}} \varphi$ *if* $\vdash_{\mathsf{ECL\,II}} \varphi$ *or there are formulas* $\psi_1, \ldots, \psi_n \in \Sigma$ *such that* $\vdash_{\mathsf{ECL\,II}} (\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \varphi$.

It is easy to verify that all these axioms are valid and the rules preserve validity. Hence the proof system for ECL II is sound. Obviously the following condition holds:

Corollary 3. *Let* $\Sigma \cup \{\varphi\} \subseteq S_{\mathsf{MDL^{+}II}}$*. Then, if* $\Sigma \vdash_{\mathsf{MDL^{+}II}} \varphi$*, then* $\Sigma \vdash_{\mathsf{ECL\,II}} \varphi$*.*

RAt and Rver axioms allow us to eliminate command operators prefixed to a proposition letter and $\top$ respectively, and other axioms enable us to reduce the length of sub-formulas to which command operators are prefixed. Consequently, any sentence of $\mathcal{L}_{\mathsf{ECL\,II}}$ can be translated into a sentence of $\mathcal{L}_{\mathsf{MDL^{+}II}}$ that is provably equivalent to it. Thus:

Definition 8 (Translation). *The translation function t that takes a formula from* $\mathcal{L}_{\mathsf{CL\,II}}$ *and yields a formula in* $\mathcal{L}_{\mathsf{MDL^{+}II}}$ *is defined as follows:*

$$
\begin{array}{lll lll}
t(p) & = & p & t([!_{(i,j)}\varphi]p) & = & p\\
t(\top) & = & \top & t([!_{(i,j)}\varphi]\top) & = & \top\\
t(\neg\varphi) & = & \neg t(\varphi) & t([!_{(i,j)}\varphi]\neg\psi) & = & \neg t([!_{(i,j)}\varphi]\psi)\\
t(\varphi \wedge \psi) & = & t(\varphi) \wedge t(\psi) & t([!_{(i,j)}\varphi](\psi \wedge \chi)) & = & t([!_{(i,j)}\varphi]\psi) \wedge t([!_{(i,j)}\varphi]\chi)\\
t(\Box\varphi) & = & \Box t(\varphi) & t([!_{(i,j)}\varphi]\Box\psi) & = & \Box t([!_{(i,j)}\varphi]\psi)
\end{array}
$$

$$
\begin{array}{llllll}
t(O_{(i,j)}\varphi) & = & O_{(i,j)}t(\varphi) & t([!_{(i,j)}\varphi]O_{(i,j)}\psi) & = & O_{(i,j)}(t(\varphi) \to t([!_{(i,j)}\varphi]\psi)) \\
 & & & t([!_{(i,j)}\varphi]O_{(k,l)}\psi) & = & O_{(k,l)}t([!_{(i,j)}\varphi]\psi) \text{ where } (i,j) \neq (k,l) \\
 & & & t([!_{(i,j)}\varphi][!_{(k,l)}\psi]\chi) & = & t([!_{(i,j)}\varphi]t([!_{(k,l)}\psi]\chi)) \\
 & & & & & \text{for any } (k,l) \in I \times J\ .
\end{array}
$$

It is easy, though sometimes tedious, to prove that this translation has the properties stated by the following corollaries and lemmas:

Corollary 4 (Translation Effectiveness). *For every formula $\eta \in S_{\mathsf{CL\,II}}$, $t(\eta) \in S_{\mathsf{MDL^+II}}$.*

Lemma 1 (Translation Correctness). *Let M be an $\mathcal{L}_{\mathsf{MDL^+II}}$-model, and w a point of M. Then for any formula η of $\mathcal{L}_{\mathsf{CL\,II}}$, $M, w \models_{\mathsf{ECL\,II}} \eta$ iff $M, w \models_{\mathsf{ECL\,II}} t(\eta)$.*

Corollary 5. *Let M be an $\mathcal{L}_{\mathsf{MDL^+II}}$-model, and w a point of M. Then for any formula η of $\mathcal{L}_{\mathsf{CL\,II}}$, $M, w \models_{\mathsf{ECL\,II}} \eta$ iff $M, w \models_{\mathsf{MDL^+II}} t(\eta)$.*

Lemma 2. *For any formula $\eta \in S_{\mathsf{ECL\,II}}$, $\vdash_{\mathsf{ECL\,II}} \eta \leftrightarrow t(\eta)$.*

These properties enable us to derive the completeness of ECL II from the known completeness of MDL$^+$II. The use of translation based on reduction axioms has been a standard method in the development of the logic of public announcements.[9] The proof of the completeness of ECL II is exactly similar to that of the completeness of ECL given in Yamada [20]. Here we only state the result.

Theorem 2 (Completeness of ECL II). *For any set $\Sigma \cup \{\varphi\}$ of formulas of $\mathcal{L}_{\mathsf{CL\,II}}$, if $\Sigma \models_{\mathsf{ECL\,II}} \varphi$, then $\Sigma \vdash_{\mathsf{ECL\,II}} \varphi$.*

5 Built-In Assumptions and Interesting Validities and Non-validities

As $I \times J$ is a finite set, from a purely formal point of view, all instances of $\mathcal{L}_{\mathsf{MDL^+II}}$ and $\mathcal{L}_{\mathsf{CL\,II}}$ are instances of $\mathcal{L}_{\mathsf{MDL^+}}$ and $\mathcal{L}_{\mathsf{CL}}$ respectively, and all $\mathcal{L}_{\mathsf{MDL^+II}}$-models are $\mathcal{L}_{\mathsf{MDL^+}}$-models. As the truth definition for $\mathcal{L}_{\mathsf{CL\,II}}$ exactly parallels that for $\mathcal{L}_{\mathsf{CL}}$, ECL II inherit all three built-in assumptions from ECL; (1) acts of commanding are assumed to be always eliminative so that we always have $R^M_{(i,j)} \restriction \chi^{\downarrow} \subseteq R^M_{(i,j)}$; (2) acts of commanding of the form $!(i,j)\varphi$ performed at some world in an model M are assumed to have no effects on the deontic accessibility relation other than $R^M_{(i,j)}$; and (3) commands are assumed to have no preconditions for their issuance.[10] Moreover, all the validities are inherited mutatis mutandis.

But in concrete applications, the distinction between command issuing authorities provide us with a finer grained treatment of examples. Suppose, for example, your boss were so stupid that he gave you a command of the form $!_{(a,b)}\neg p$ on the same day he had commanded you to see to it that p. Now, ECL II inherits the following principles from ECL:

[9] Van Benthem & Liu [3] proved that every relation changing operation that is definable in PDL without iteration has a complete set of reduction axioms in dynamic epistemic logic.

[10] For a detailed discussion of these assumptions, see [21].

(DE) $[!_{(i,j)}(\varphi \wedge \neg\varphi)]O_{(i,j)}\psi$ (Dead End)

(RSC) $[!_{(i,j)}\varphi][!_{(i,j)}\psi]\chi \leftrightarrow [!_{(i,j)}(\varphi \wedge \psi)]\chi$ where $\varphi, \psi \in S_{(i,j)}$-free (Restricted Sequential Conjunction)

(ROI) $[!_{(i,j)}\varphi][!_{(i,j)}\psi]\chi \leftrightarrow [!_{(i,j)}\psi][!_{(i,j)}\varphi]\chi$ where $\varphi, \psi \in S_{(i,j)}$-free . (Restricted Order Invariance)

As $\neg p$ is (a, b)-free, by Restricted Sequential Conjunction Principle, we have:

$$[!_{(a,b)}p][!_{(a,b)}\neg p]\chi \leftrightarrow [!_{(a,b)}(p \wedge \neg p)]\chi \ . \tag{18}$$

By Dead End Principle, we have:

$$M, s \models_{\mathsf{ECL\,II}} [!_{(a,b)}(p \wedge \neg p)]O_{(a,b)}\psi \ . \tag{19}$$

Hence:

$$M, s \models_{\mathsf{ECL\,II}} [!_{(a,b)}p][!_{(a,b)}\neg p]O_{(a,b)}\psi \ . \tag{20}$$

This is equivalent to:

$$(M_{!_{(a,b)}p})_{!_{(a,b)}\neg p}, s \models_{\mathsf{ECL\,II}} O_{(a,b)}\psi \ . \tag{21}$$

As $(R^M_{(a,b)} \upharpoonright p^{\downarrow}) \upharpoonright \neg p^{\downarrow}$ is empty, no world that is compatible with the obligations with respect to your boss is accessible from s for you; you are in an absurd state.[11]

But if it is not your boss but your guru that commanded you to see to it that $\neg p$, you will be in a slightly different situation. We then have:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}\neg p}, s \models_{\mathsf{ECL\,II}} (O_{(a,b)}p \wedge O_{(a,c)}\neg p) \wedge \neg\Diamond(p \wedge \neg p) \ . \tag{22}$$

As $p \wedge \neg p$ is a contradiction, it is logically impossible for you to obey both your boss's command and your guru's command. But there might still be worlds $R_{(a,b)}$-accessible from s and worlds $R_{(a,c)}$-accessible from s in $(M_{!_{(a,b)}p})_{!_{(a,c)}\neg p}$. And so, you are not in an obligational dead end but in an obligational dilemma.

Now let's go back to the first example, in which your guru commanded you to see to it that q after your boss commanded you to see to it that p. We have considered one possible way of representing the situation you are supposed to be in after the issuance of your guru's command in this example, namely (17). I reproduce it here as (23).

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} O_{(a,b)}p \wedge O_{(a,c)}q \wedge \neg\Diamond(p \wedge q) \ . \tag{23}$$

The most important difference between (22) and (23) consists in the fact that $p \wedge q$ is not a contradiction while $p \wedge \neg p$ is. So there might be a world t, even in M, for which the following condition holds:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, t \models_{\mathsf{ECL\,II}} p \wedge q \ . \tag{24}$$

[11] Note that Dead End Principle precludes the possibility of adding multi-agent variant of the so-called D axiom, $O_{(a,b)}\phi \to P_{(a,b)}\phi$, to ECL II. For more on this point, see Yamada [21].

So, the fact that the impossibility involved in this situation is not a logical impossibility can be said to be reflected in a sense even if we accept (23).

Accepting (23) as a way of representing the situation here, however, still seems to be a bit problematic. As I remarked earlier, we may say that if a sufficiently fast means of transportation were available, it would be possible for you to obey both commands. Thus, if $((M_{!_{(a,b)}p})_{!_{(a,c)}q}, s)$ is to represent the situation you are supposed to be in, it seems that we ought to have:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} \Diamond(p \wedge q) \ . \tag{25}$$

But then, (23) cannot be correct. Thus, the only remaining way of representing the sort of impossibility involved here in $\mathcal{L}_{\mathsf{CL\,II}}$ seems to be to say:

$$(M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} O_{(a,b)}p \wedge O_{(a,c)}q \wedge \neg(p \wedge q) \ . \tag{26}$$

Thus, if you obey your boss's command, then you will disobey your guru's command $((M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} p \rightarrow \neg q)$, and if you obey your guru's command, then you will disobey your boss's command $((M_{!_{(a,b)}p})_{!_{(a,c)}q}, s \models_{\mathsf{ECL\,II}} q \rightarrow \neg p)$, in the real world you are in. Even if a sufficiently fast means of transportaion were available to you in any other possible worlds, it would not be of much help to you. In this example, you are in an obligational dilemma in the real world just because of a contingent fact about the present state of the system of transportation in it. The situation looks very closely similar to those situations in which you are in moral dilemmas.[12]

6 Conclusion

In this paper, an eliminative command logic ECL is slightly refined into ECL II by allowing command terms and deontic operators to be indexed by a Cartesian Product of a given finite set of agents and a given finite set of command issuing authorities. Complete axiomatization and interesting validities are presented, and a concrete example of a situation in which conflicting commands are given to one and the same agent by different authorities is discussed extensively.

In ECL and ECLII, model updating operations are used to model effects of acts of commanding. This idea is imported from dynamic epistemic logics developed in Plaza [15], Groeneveld [9], Gerbrandy and Groeneveld [6], Gerbrandy [5], Baltag, Moss, & Solecki [2], and Kooi & van Benthem [11] among others. In these logics, model updating operations are used to model effects of various forms of information transmissions. In the field of deontic reasoning, van der Torre & Tan [18] and Žarnić [22] extended update semantics of Veltman [19] and uses model updating operations to interpret normative sentences and natural language imperatives respectively. As is noted in Yamada [21], the relation between their semantics on the one hand and ECL and ECLII on the other is analogous to that between Veltman's update semantics and various epistemic logics. In this respects, DLP_{dyn} of Pucella & Weissman [16] and DLP^{+}_{dyn} of Demri [4] are closer to ECL and ECLII in spirit. They use model updating operations to model

[12] An illuminating discussions on moral dilemmas can be found in Marcus [13].

changes in legal policies and thereby dynamified DLP, a logic of permission, of van der Meyden [14]. And more recently, van Benthem & Liu [3] proposed "preference upgrade" as a counter part to information update. According to them, my "command operator for proposition A can be modeled exactly as an upgrade sending R to $R; ?A$" in their system, and their paper "provides a much more general treatment of possible upgrade instructions" ([3]). Their preference upgrade really has a much wider application than the deontic update of the present paper. But, as is noted in [21], the notion of preference upgrade seems to be connected with perlocutionary consequences of various utterances, while the deontic update is used to capture effects of acts of commanding as a specific kind of illocutionary acts. They can be seen as mutually complementary.

With regards to the possibilities of further research, there is an apparent need of dynamifying richer deontic languages. The dynamified languages $\mathcal{L}_{\mathsf{CL}}$ and $\mathcal{L}_{\mathsf{CLII}}$ inherit various inadequacies from the static base languages $\mathcal{L}_{\mathsf{MDL}^+}$ and $\mathcal{L}_{\mathsf{MDL}^+\mathsf{II}}$.[13] Moreover, the possibilities of update logics of various other kinds of illocutionary acts suggest themselves. For example, an act of promising can be considered as another deontic updator, and an act of asserting as an updator of propositional commitments. Here I only mention one interesting immediate application. A command type term of the form $!_{(i,j)}\varphi$ can be reinterpreted as a term for a type of an act of promising with a promisor i and an promisee j to the effect that i will see to it that φ. Then the analogue of CUGO principle will state that acts of promising usually generates obligations. Comparing this with Searle's discussion on the relation between acts of promising and obligations in Searle [17] will be a task for another paper.

References

1. Austin, J. L.: How to Do Things with Words. Urmson, J. O., Sbisà, M. (eds.), 2nd. ed. Harvard University Press, Cambridge (1975)
2. Baltag, A., Moss, L. S., Solecki, S.: The Logic of Public Announcements, Common Knowledge, and Private Suspicions. Technical Report, TR534. Department of Computer Science (CSCI), Indiana University (1999)
3. van Benthem, J., Liu, F.: Dynamic Logic of Preference Upgrade. Journal of Applied Non-Classical Logics (to appear)
4. Demri, S.: A Reduction from DLP to PDL. Journal of Logic and Computation, 15-5 (2005), 767–785
5. Gerbrandy, J.: Bisimulations on Planet Kripke. Ph. D. Thesis. ILLC Dissertation Series DS-1999-01, University of Amsterdam (1999)
6. Gerbrandy, J., Groeneveld, W.: Reasoning about Information Change. Journal of Logic, Language, and Information, Vol.6 (1997) 147–169

[13] Kooi & Tamminga [12] introduced a formula of the form $\odot^{\mathcal{F}}_{\mathcal{G}}\varphi$ in order to deal with conflicting obligations. Intuitively, a formula of this form is supposed to mean that group $\mathcal{G}$ of agents ought to see to it that φ in the interest of group $\mathcal{F}$. When $\mathcal{G}$ and $\mathcal{F}$ are unit sets, say, $\{i\}$ and $\{j\}$ respectively, we get a formula of the form $\odot^{j}_{i}\varphi$. A formula of this form can be used to express what we express by using a formula of the form $O_{(i,j)}\varphi$ in $\mathcal{L}_{\mathsf{MDL}^+\mathsf{II}}$. As their logic extends a simplified version of Horty's multi-agent deontic logic based on stit theory developed in [10], it suggests interesting possibilities of extending $\mathsf{MDL}^+\mathsf{II}$, though such extended logics will obviously be more difficult to dynamify than $\mathsf{MDL}^+\mathsf{II}$.

7. Goble, L., Meyer, J-J. Ch.: Deontic Logic and Artificial Normative Systems: 8th International Workshop on Deontic Logic in Computer Science, DEON 2006, Utrecht, The Netherlands, July 2006, Proceedings. Lecture Notes in Artificial Intelligence, vol. 4048. Springer-Verlag, Berlin / Heidelburg / New York (2006)
8. Grice, H. P.: Meaning (1948). Reprinted in: Grice, H. P., Studies in the Way of Words. Harvard University Press. Cambridge, Massachusetts / London, England (1989) 213–223
9. Groeneveld, W.: Logical Investigations into Dynamic Semantics. Ph. D. Thesis. ILLC Dissertation Series DS-1995-18, University of Amsterdam (1995)
10. Horty, J. F.: Agency and Deontic Logic, Oxford University Press, Oxford (2001)
11. Kooi, B., van Benthem, J.: Reduction Axioms for Epistemic Actions. In: Schmidt, R., Pratt-Hartmann, I., Reynolds, M. and Wansing, H. (eds): Preliminary Proceedings of AiML-2004: Advances in Modal Logic. Technical Report Series, UMCS-04-9-1, Department of Computer Science, University of Manchester (2004) 197–211
12. Kooi, B., Tamminga, A.: Conflicting Obligations in Multi-agent Deontic Logic. In [7]. (2006) 175–186
13. Marcus, R. B.: Moral Dilemmas and Consistency. Journal of Philosophy, LXXVII, 3 (1980) 121-136. Reprinted in: Marcus, R. B.: Modalities: Philosophical Essays. Oxford University Press, New York/Oxford (1993) 125-141
14. van der Meyden, R.: The Dynamic Logic of Permission. Journal of Logic and Computation, Vol. 6, No. 3 (1996) 465–479
15. Plaza, J. A.: Logics of public communications. In: Emrich, M. L., Pfeifer, M. S., Hadzikadic, M., Ras, Z. W. (eds.): Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems (1989) 201–216
16. Pucella, R., Weissman, V.: Reasoning about Dynamic Policies. In: Walukiewicz (ed.): FOSSACS 2004. Lecture Notes in Computer Science, vol. 2987. Springer-Verlag, Berlin / Heidelburg / New York (2004) 453-467
17. Searle, J.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge (1969)
18. van der Torre, L. W. N., Tan, Yao-Hua: An Update Semantics for Deontic Reasoning. In: McNamara, P., Prakken, H. (eds.): Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science, Volume 49, Frontiers in Artificial Intelligence and Applications, IOS Press (1999) 73–90
19. Veltman, F.: Defaults in Update Semantics. Journal of Philosophical Logic, 25 (1996) 221–261
20. Yamada, T.: Completeness of the Logic of Eliminative Commands. In Tanaka, H. (ed.): Language Understanding and Action Control. Annual Project Report, Grant-in-Aid for Creative Basic Research (13NP0301). Tokyo Institute of Technology (2006) 9-20
21. Yamada, T.: Acts of Commanding and Changing Obligations. In: Inoue, K., Sato, K., Toni, F. (eds.): Post-Proceedings of the Seventh International Workshop on Computational Logic in Multi-Agent Systems (CLIMA VII), Future University Hakodate, Hakodate, Japan, 8-9 May, 2006, in association with AAMAS2006. Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin / Heidelburg / New York (to appear)
22. Žarnić, B.: Imperative Change and Obligation to Do. In Segerberg, K., Sliwinski (eds.): Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Åqvist. Uppsala Philosophical Studies 51. Department of Philosophy, Uppsala University, Uppsala (2003) 79–95

On Factive Islands: Pragmatic Anomaly vs. Pragmatic Infelicity

David Y. Oshima

Department of Languages and Literatures, Arizona State University,
Tempe, AZ 85287, U.S.A.
davidyo@gmail.com
http://www.public.asu.edu/~dyoshima

Abstract. Certain types of wh-phrases (e.g. *how*, *why*) cannot be extracted from the complement clause of a factive predicate, nor can they occur *in situ* within it (the factive island effect). This paper argues that the factive island effect is a pragmatic phenomenon, which follows from two independent factors: (i) the speaker's expectation about possible answers of wh-interrogatives, and (ii) presuppositions induced by factive predicates. The proposed account illustrates a special kind of pragmatic infelicity (which I term "pragmatic anomaly"), which can be opposed to "contingent" pragmatic infelicity such as presupposition failure, violation of Gricean maxims, etc.

1 Introduction

This paper provides a pragmatic account of the factive island constraint (FIC), the phenomenon whereby the complement clause of a factive predicate (e.g., *know*, *regret*) serves as an island environment (domain which does not allow extraction; [1]) for certain kinds of wh-phrases.[1] Section 2 illustrates basic data. Section 3 reviews previous syntactic/semantic analyses of the FIC, and points out their limitations and drawbacks. Section 4 demonstrates that the FIC can be derived from two independently motivated factors: (i) the speaker's expectation about possible answers of wh-interrogatives, and (ii) presuppositions induced by factive predicates. Section 5 argues that the proposed account illustrates a special kind of pragmatic infelicity, which can be opposed to "contingent" pragmatic infelicity such as presupposition failure.

2 Basic Facts

2.1 Types of Wh-Phrases That Are Subject to the FIC

The complement clause of a factive predicate has been considered a "weak island", which allows extraction of certain kinds of wh-phrases, but not of others

[1] Throughout the paper, unless noted otherwise, "extraction" refers to the phenomenon whereby a wh-phrase in a complement clause takes wide scope over the clause-taking predicate at the semantic level, whether or not it is fronted over the clause-taking predicate at the (surface) syntactic level.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 147–161, 2007.

([2]–[7]). Except for some clear-cut cases, acceptability judgments on sentences that involve extraction from a factive complement tend to be subtle, be heavily affected by the surrounding context, and vary among speakers. Speakers seem to agree, however, (i) that argument wh-phrases like *who* and *which man* allow extraction more readily than adjunct wh-phrases like *when* and *where*,[2] (ii) that the extraction of *when* and *where* is more acceptable than or at least as acceptable as the extraction of *how*, and (iii) that the extraction of *how* is more acceptable than or at least as acceptable as the extraction of *why*.[3]

---

[2] Adams ([8]) and Rooryck ([3]) maintain that a factive complement clause blocks the extraction of the subject, while it allows the extraction of an object:

(i) (Judgements are Rooryck's)
   a. *Who do you regret/understand/forget likes this book? (= (1a) in [3]; cited from [8])
   b. ?Which article did you regret/understand/forget I had selected? (= (1c) in [3])
   c. Who do you believe likes this book? (= (2a) in [3]; cited from [8])

Hegarty ([6]), on the other hand, claims that subject-extraction is as a rule possible for both factive and non-factive complements. The contrasts illustrated in (i), I believe, do not directly have to do with the factivity. The contrast between (ia) and (ib), on the one hand, can be attributed to the fact that syntactic extraction of a subject tends to be more difficult than that of an object ([4], [9]). The contrast between (ia) and (ic), on the other hand, can be attributed to the fact that clause-taking verbs like *believe*, *think*, and *feel*, which can be used "annotatively" (as in *They, I believe, did't make it after all*; cf. ??*They, I regret, didn't make it after all*), allow extraction more readily than those which cannot (see [10]:178–181, [11]).

The contrast between subject-extraction/object-extraction out of a factive complement is not strongly felt in Japanese, as long as wh-interrogatives that do not involve a long-distance dependency at the surface syntactic level are concerned:

(ii) a. Taro-wa [dare-ga Ken-o tukitobasi-ta koto]-ni kizui-ta-no?
   T.-Top who-Nom K.-Acc push.off-Past fact-Dat realize-Past-Q?
   'Who did Taro realize pushed Ken off?'
   b. Taro-wa [Ken-ga dare-o tukitobasi-ta koto]-ni kizui-ta-no?
   T.-Top K.-Nom who-Acc push.off-Past fact-Dat realize-Past-Q?
   'Who did Taro realize Ken pushed off?

[3] Rooryck ([3]:357) points out that sentences like the following, where it is clear that *how* cannot be construed as modifying the matrix predicate, are marginally acceptable:

(i) ??*How* did he deeply regret [that his son had fixed the car _]?

(ii) ??*How* did John very well know [that his son would have fixed the car_]?

(1) **Scale of Extractability:**
argument wh-phrases (WHO, etc.) > {WHEN, WHERE} > HOW > WHY[4]

(2) a. Who does Max know [that Alice criticized _]?
b. (?)When does Max know [that Alice took a day off _]?
c. (?)Where does Max know [that Alice watched a baseball game _]?
d. ?*How does Max know [that Alice went to San Francisco _]?
e. *Why does Max know [that Alice insulted Pat _]?

The factive island effect obtains also with *in situ* wh-interrogatives, which comprise the paradigmatic type of wh-interrogatives in languages like Japanese. In Japanese too, extractability of wh-phrases varies in accordance with the scale given in (1). The extraction of WHEN, WHERE, and HOW appears to be somewhat easier in Japanese than in English; this is probably because there is no extra processing load occasioned by the syntactic long-distance dependency in Japanese. The extraction of WHY is, however, downright unacceptable in Japanese too.

(3) a. Taro-wa [Hanako-ga dare-o hihan-si-ta koto]-o
T.-Top H.-Nom who-Acc critisize-Past fact-Acc
sitte-i-ru-no?
know-Asp-Pres-Q?
'Who does Taro know [that Hanako criticized _]?'
b. (?)Taro-wa [Hanako-ga itu sigoto-o yasun-da koto]-o
T.-Top H.-Nom when work-Acc rest-Past fact-Acc
sitte-i-ru-no?
know-Asp-Pres-Q?
'When does Taro know [that Hanako took a rest from work _]?'
c. (?)Taro-wa [Hanako-ga doko-de kaimono-o si-ta-koto]-o
T.-Top H.-Nom where shopping-Acc do-past-fact-Acc
sitte-i-ru-no?
know-Asp-Pres-Q?
'Where does Taro know [that Hanako went shopping _]?'
d. ??Taro-wa [Hanako-ga doo Tokyo-made it-ta koto]-o
T.-Top H.-Nom how Tokyo-to go-Past fact-Acc
sitte-i-ru-no?
know-Asp-Pres-Q?
(What is the means/manner x such that Taro knows that Hanako went to Tokyo by x?)
e. *Taro-wa [Hanako-ga doosite Ken-o buzyoku-si-ta koto]-o
T.-Top H.-Nom why K.-Acc insult-Past fact-Dat
sitte-i-ru-no?
know-Asp-Pres-Q?

[4] By capitalized WHO, WHEN, etc., I refer to English *who*, *when*, etc. and their counterparts in other languages.

(What is the reason x such that Taro knows that Hanako insulted Ken for x?)

In the rest of the present paper, somewhat abstracting away from inter-speaker/ cross-linguistic differences,[5] I will assume (i) that WHO, WHEN, and WHERE can be extracted from a factive complement, (ii) that the extraction of HOW is marginal (nearly impossible), and (iii) that the extration of WHY is utterly unacceptable.

2.2 Unique wh-Phrase

Szabolcsi and Zwarts ([2]:269–72) point out that a wh-phrase cannot be extracted from a factive complement clause when the resulting complement clause with a gap would denote a "one-time-only" property, i.e., a property that can be true of at most one object (*succeed George V*, *assassinate Abraham Lincoln*, etc. are one-time-only properties, while *greet George V*, *vote for Abraham Lincoln*, etc. are not):

(4) a. To whom do you regret having shown this letter?
b. *From whom do you regret having gotten this letter?
([2]:271)

(5) a. Who does Max know [that Alice sent a Christmas card to _]?
b. *Who does Max know [that Alice got married to _ on June 1st]?

Sz&Z's finding can be reformulated as follows: the extraction of a wh-phrase α from a factive complement C is blocked when [α C] (or $[_C \ldots \alpha \ldots]$ in languages with *in situ* wh-interrogatives) would form an interrogative such that among its resolutions, at most one can be true (its resolutions are mutually exhaustive).[6] For (4a), for example, [α C] is "to whom [(your) having shown this letter _]", which has resolutions like 'you showed this letter to Max', 'you showed this letter to Pat', etc. Among such resolutions, more than one can be true at the same time, as one can show the same letter to more than one individual. For (4b), likewise, [α C] is "from whom [(your) having gotten this letter _]", which has resolutions like 'you got this letter from Max', 'you got this letter from Pat', etc. This time, at most one resolution can be true, because there can be only one sender (that may be either an individual or group of individuals) for a single letter.

[5] Some English speakers find it unacceptable/marginal to extract wh-adjuncts in general from factive complement clauses (or complement clauses in general; see [10]:ch.4). The present study does not address the questions of (i) why it is unacc-ceptale/marginal to some speakers, and (ii) how this inter-speaker variance arises.

[6] A "resolution" of a wh-interrogative is, roughly speaking, the proposition denoted by a sentence that is formed by de-fronting the wh-phrase and instantiating it with a specific constant (e.g., 'John laughed' is a resolution of "Who laughed?"). See below for a fuller definition.

3 Previous Analyses and Their Problems

3.1 Syntax-Based Approach

In previous syntactic analyses ([3], [4], among others), it has been argued that the factive island effect arises as a result of an Empty Category Principle violation, which is caused by an empty operator occupying [Spec, CP] of the factive complement. This line of analysis seems rather *ad hoc*, in that it postulates an invisible operator whose existence is not independently motivated. Furthermore, it fails to provide an account of the asymmetry between different wh-adverbials (WHY, HOW vs. WHERE, WHEN), or the impossibility of extraction of an argument/adjunct of a one-time-only predicate.

3.2 Semantics-Based Approach

Szabolcsi and Zwarts ([2]) propose a semantic account of the FIC (and other types of island constraints, such as the negative island constraint), which is based on algebraic properties of wh-phrases. Their account of the un-extractability of a manner denoting wh-phrase (e.g. *how*), for example, runs as follows:

(6) (i) The meaning of a factive predicate involves conjunction (e.g. "Mary regrets that John left" has a logical form along the lines of: **regret**(**Mary**)(**that John left**) $\wedge$ **fact**(**that John left**)).
(ii) To answer a question where a wh-phrase α takes wide scope over a factive predicate, thus, one needs to intersect two members that belong to the domain of α.
(iii) The denotation domain for manners is a semi-lattice. Since a semi-lattice is not closed under intersection, one cannot carry out the procedure to calculate the answer of a question like "How does Mary regret [John left _]?".

Sz&Z's claim that we cannot generally take the intersection of manners, however, seems rather unintuitive and unmotivated (e.g. 'quickly and violently' can be formed by intersecting 'quickly' and 'violently'; see [12]).[7] Also, it is not clear how Sz&Z's analysis can deal with the asymmetry among wh-adverbials. They do not discuss the extraction of WHEN or WHERE; they also exclude WHY from consideration, because of its "atypical" behavior ([2]:fn.14).

4 An Alternative Analysis

4.1 Some Ontological Assumptions and Terminological Clarifications

Prior to presenting an alternative, discourse-based analysis of the FIC, some background assumptions and terminological conventions need to be laid out.

[7] In this connection, Sz&Z claim that the ordering (inclusion, etc.) of manners is based on union rather than intersection, and that, for example, *nicely and stupidly* denotes a manner that subsumes both *nicely* and *stupidly* ([2]:264). This assumption seems implausible, as it fails to capture our intuition about the direction of entailment (e.g. 'John behaves nicely and stupidly' entails 'John behaves nicely', and not vice versa).

First, following the standard view, I assume that the meaning of an interrogative (question) can be characterized in terms of what (statement) counts as an answer to it, whether it is a "correct" or "wrong" answer ([13]–[15]).

I distinguish two types of answers: immediate answers, which are designated as "resolutions", and non-immediate answers. The set of resolutions of a yes-no interrogative I is: $\{[\![S]\!]^{\mathrm{M,g}}, [\![\neg S]\!]^{\mathrm{M,g}}\}$, where S is a declarative sentence corresponding to I. The set of resolutions of a wh-interrogative I: [wh$_i$ ... $\_{}_i$ ...] (or [... wh$_i$ ...]) is: $\{[\![S']\!]^{\mathrm{M,g}}, [\![S'']\!]^{\mathrm{M,g}}, \ldots\}$, where S', S'' ... are declaratives that are formed by instantiating wh$_i$ with a non-wh-expression of the appropriate category (e.g., individual-denoting terms like *John* for WHO, manner-denoting terms like *slowly* for HOW). The resolution set for an interrogative may contain both true and false statements.

The definition of answer-hood is derived from the notion of resolution: a statement is an answer to an interrogative I iff, if it is true, it would entail some or all of the *true resolutions* of I (note that a resolution is an answer by this definition). To give a concrete example, when "Who likes sushi?" is uttered in a situation where only Pat and Alice like sushi, its resolutions, true resolutions, and non-immediate answers will be as follows:[8]

(7) "Who likes sushi?"

a. **resolutions** = {Max likes sushi; Pat likes sushi; Alice likes sushi; ...}
b. **true resolutions** = {Pat likes sushi; Alice likes sushi}
c. **non-immediate answers** = {Max and Pat like sushi; Everyone likes sushi; ...}
d. **answers** = (7a) ∪ (7c)

It is generally believed that a wh-interrogative presupposes that it has at least one true resolution; I too take this view, although it can be a matter of debate. Some wh-interrogatives can have more than one true resolution; for example, the question: "Who likes sushi?" can have more than one true resolution, although it is also possible for it to have only one. Wh-interrogatives with a one-time-only predicate, on the other hand, can have only one true resolution; the question: "Who left first?", for example, may have countless resolutions like 'John left first', 'Max left first', 'Pat left first', and so on, but only one of them can be true. Wh-interrogatives that can have only one true resolution will be referred to as *unique wh-interrogatives* (wh-interrogatives that *happen to* have a unique true resolution are not included).

4.2 Unique wh-Phrase

As mentioned in section 2.2, a wh-phrase α occurring in a factive complement C cannnot be extracted (from C) when [α C] would form a unique wh-interrogative.

[8] Note that, according to this definition of answer-hood, 'Max or Pat likes sushi', 'Max may like sushi', 'Nobody likes sushi', etc. do not count as (immediate or non-immediate) answers to "Who likes sushi?".

This phenomenon can be given a simple, discourse-based explanation, without making any syntactic or ontological stipulation.

By uttering a wh-interrogative, the speaker seeks to determine which of its resolutions is true. When the wh-phrase is an argument or adjunct of the predicate of a factive complement, some of the resolutions are already “screened out” (are expected not to be true) at the time the interrogative is uttered.

Let us take the pair of sentences in (5) as examples. Resolutions of (5a) are of the form: ‘Max knows that Alice sent a Christmas card to x’. Because of the factive presupposition triggered by *know*, however, some of these resolutions are incompatible with the common ground, i.e., the collection of knowledge shared by the interlocutors (unless the speaker knows that Alice sent a Christmas card to *everyone*). The resolution: ‘Max knows that Alice sent a Christmas card to Pat’, for example, can be a true statement only when it is in the common ground that Alice sent a Christmas card to Pat (otherwise, the factive presupposition induced by *know* would not be satisfied). Conversely, when it is not in the common ground that Alice sent a Christmas card to Pat, the speaker can infer that ‘Max knows that Alice sent a Christmas card to Pat’ is not a true resolution before asking (5a). Therefore, what (5a) asks is practically equivalent to: “Of the x’s such that it is in the common ground that Alice sent a Christmas card to x, who is such that Max knows Alice sent a Christmas card to him/her?”.

Now, let us consider the case of (5b). In the same vein, what it asks can be paraphrased as: “Of the x’s such that it is in the common ground that Alice got married to x on June 1st, who is such that Max knows that Alice got married to him on June 1st?”. This results in pragmatic oddity for an obvious reason. In a quotidian context, it is natural to assume that a single person would not get married twice (to different persons) on the same day. If there is only one x such that it is known to the speaker that Alice got married to x on June 1st, the speaker should know *the* answer to his question before asking it.

Note that questions like (4b) and (5b) are not odd merely because they fail to meet the speaker’s expectation, like in the case of a simple presupposition failure as in:

(8) A: Who does Max know that Alice sent a Christmas card to?
B: (Max knows that Alice sent a Christmas card to) Pat.
A: Oh, I didn’t know that Alice sent a Christmas card to Pat.

In this dialogue, speaker B’s answer is not ideally felicitous, presupposing a proposition that does not belong to the common ground; speaker B might have misunderstood what belongs to A’s knowledge, or he might be just expecting A to accommodate the common ground based on the answer. In either case, B’s answer does not meet A’s expectation, to the effect that A is forced to reorganize his belief. When uttering (4b) and (5b), in contrast, the speaker has a chance to obtain an informative answer only if the answer is pragmatically deviant. This means that the speaker should expect the answer to be already known to him, and at the same time expect this expectation not to be fulfilled, which is possible only with an irrational mind (see discussion in section 5).

4.3 WHY

The proposed, pragmatic account of the un-extractability of unique wh-phrases can be extended to the case of WHY. In the following, I demonstrate that WHY cannot be extracted from a factive complement simply because an interrogative with WHY is always a unique wh-interrogative.

The semantics of reasons (reason-denoting adverbial phrases) has been relatively understudied; scholars generally agree, however, that a reason is a proposition, which stands in a certain relation to another proposition (consequence), and that contextual relevance is a crucial factor to determine "what counts as a reason" ([16], [17]). For propositions p and q to stand in the reason-consequence relation with respect to world of evaluation w_0 ('p, because q'), it must be the case (i) that $p(w_0) = q(w_0) = 1$ and (ii) that in every world w_1 that is accessible from w_0, $p(w_1) = q(w_1)$. This is, however, too weak to be a sufficient condition. Consider the following examples:

(9) Why does John live in a bachelor apartment?

(10) John lives in a bachelor apartment ...
- a. because he is single.
- b. because he doesn't like having roommates.
- c. because he cannot afford to buy a house.
- d. because he doesn't want to live with his mother.
- ...

Let us assume that in the current state of affairs, all of (10a-d) satisfy necessary conditions (i) and (ii) for the reason-consequence relation described above. Which, if any, of them counts as an answer to the question in (9) is, however, contingent on the interlocutors' interests/attention. When (10a) is provided as an answer, the utterer of (9) may accept it as the reason that John lives in a bachelor apartment, but he could as well refute it by saying: "I know, but what I'm asking is why {he doesn't live with a roommate/he doesn't buy a house/he doesn't live with his mother}".

The following, thus, appears to be a reasonable approximation of the semantics of reason:

(11) Proposition q is a reason for proposition p with respect to world w_0 iff:
- (i) $p(w_0) = q(w_0) = 1$;
- (ii) in every world w_1 that is accessible from w_0, $p(w_1) = q(w_1)$; and
- (iii) among q, q', q'', ... that satisfy conditions (i) and (ii), q is the most contextually relevant.

Clause (iii) of (11), in effect, implies that a proposition can have only one reason in a given context. This claim is given support by data like the following:

(12)
- a. Alice insulted Max because she was on edge, among other reasons.
- b. Alice insulted Max because she was on edge.

From the statement in (12a) that involves an 'among other x' phrase, we cannot infer (12b) (the latter statement can be refuted by saying: "No, that's only part of the reason"). In other words, (12a) does *not* mean that: 'Alice insulted Max because she was on edge, and Alice insulted Max because of some other reasons', but rather: 'Alice insulted Max for the reason R such that R is the conjunction of the proposition that Alice was on edge and some other propositions'. The (invalidness of the) inference pattern illustrated in (12) is shared by unique arguments/adjuncts in general. From (13a), where the VP denotes a one-time-only property, we cannot infer (13b) (assume that *among others* in (13a) modifies the subject):

(13) a. ?Alice, among others, killed Max.
b. Alice killed Max.

(13a) sounds quite awkward, but if it can mean anything, it means: 'Alice is a member of the group of individuals who collectively killed Max'. On the other hand, from (14a), where the VP is not a one-time-only predicate, we can infer (14b) (unless we take the dispreferred reading of (14a) which is equivalent to: 'Alice is a member of the group who collectively insulted Max'):

(14) a. Alice, among others, insulted Max.
b. Alice insulted Max.

From the fact that a proposition has only one reason in a given context, it follows that a why-question of the form: "why p?", whose resolution is a statement that a certain proposition serves as a reason for p, is always a unique wh-interrogative. The pragmatic account proposed in the previous subsection thus carries over.

4.4 HOW

As we observed in section 2.1, HOW resists extraction from a factive complement, but allows it somewhat more easily than WHY. In the following, I argue that this is because an interrogative with HOW *tends to* be interpreted as a unique wh-interrogative.

A priori, we can think of two types of situations where an interrogative with HOW may have more than one true resolution. The first is cases where multiple manners, which are provided as (parts of) resolutions, are understood as modifying different events.[9] The second is situations where multiple manners are understood as modifying a single event. Either case is rarely found, for the reasons explained below.

(i) Multiple manners - multiple events

Following the standard view, let us assume that the meaning of a sentence generally involves existential quantification of the event argument introduced by the (main) predicate, and further that manner, time, and place adverbs (adverbial

[9] In the present work, I assume that "manners" subsume means and methods.

phrases) pose semantic restrictions on the event argument ([12], among others). The sentence: "John spoke to Mary rudely in the cafeteria on Monday" is, for example, given the logical form along the lines of (15) (the tense is ignored):

(15) $\exists e_1[\textbf{speak-to}(e_1, \textbf{john}, \textbf{mary}) \wedge \textbf{rude}(e_1) \wedge \textbf{in}(e_1, \textbf{the-cafeteria}) \wedge \textbf{on}(e_1, \textbf{monday})]$

Time and place adverbial phrases can be interpreted distributively when they are conjoined with AND:

(16) a. John spoke to Mary in the cafeteria and in the library.
b. $\exists e_1[\textbf{speak-to}(e_1, \textbf{john}, \textbf{mary}) \wedge \textbf{in}(e_1, \textbf{the-cafeteria})] \wedge \exists e_2[\textbf{speak-to}(e_2, \textbf{john}, \textbf{mary}) \wedge \textbf{in}(e_2, \textbf{the-library})]$

(17) a. John spoke to Mary on Monday and on Tuesday.
b. $\exists e_1[\textbf{speak-to}(e_1, \textbf{john}, \textbf{mary}) \wedge \textbf{on}(e_1, \textbf{monday})] \wedge \exists e_2[\textbf{speak-to}(e_2, \textbf{john}, \textbf{mary}) \wedge \textbf{on}(e_2, \textbf{tuesday})]$

Manner adverbs, however, tend not to allow the distributive interpretation. (18a), for example, is intuitively felt to be contradictory (unless it is understood as: 'John spoke to Mary in an intricate manner that has some flavor of rudeness and some flavor of niceness').

(18) a. #John spoke to Mary rudely and nicely.
b. $\exists e_1[\textbf{speak-to}(e_1, \textbf{john}, \textbf{mary}) \wedge \textbf{rude}(e_1)] \wedge \exists e_2[\textbf{speak-to}(e_2, \textbf{john}, \textbf{mary}) \wedge \textbf{nice}(e_2)]$

The more "reasonable" interpretation provided in (18b) is not easily available, if not impossible. This suggests that a manner adverbial phrase tends to be interpreted as modifying the *maximal event*, i.e., the collection of all events that satisfy the descriptions posed by the predicate and arguments. The sentence: "John spoke to Mary rudely", for example, is interpreted as (19b), rather than as (19a):

(19) John spoke to Mary rudely.
a. $\exists e_1[\textbf{speak-to}(e_1, \textbf{john}, \textbf{mary}) \wedge \textbf{rude}(e_1)]$
b. $\exists E_1[\textbf{speak-to}(E_1, \textbf{john}, \textbf{mary}) \wedge \textbf{rude}(E_1) \wedge \forall e_1[\textbf{speak-to}(e_1, \textbf{john}, \textbf{mary}) \rightarrow e_1 \lhd E_1]]$

where,

(20) E_1, E_2, E_3, ... are variables over collections of events (which consist of one or more atomic events)

and

(21) For any world w and any time i, $[\![e \lhd E]\!]^{\mathrm{M,w,i,g}} = 1$ iff $[\![e]\!]^{\mathrm{M,w,i,g}}$ is a member of $[\![E]\!]^{\mathrm{M,w,i,g}}$

The maximal interpretation requirement posed by a manner adverbial phrase is not always categorical, but can be removed with appropriate settings. (22a), for example, allows quite easily the reading where the two adverbial phrases are associated with different events.

(22) a. Mary has travelled to Chicago by train and by bus.
b. $\exists e_1$[**travel-to**(e_1, **mary**, **chicago**) $\wedge$ **by-train**(e_1)] $\wedge$
$\exists e_2$[**travel-to**(e_2, **mary**, **chicago**) $\wedge$ **by-bus**(e_2)]

It seems natural to assume that HOW, as a kind of manner adverb, tends to force the maximal interpretation. This means that, for example, what the wh-interrogative: "How did John spoke to Mary?" asks is in effect "What manner is such that it characterizes *all* instances of events whereby John spoke to Mary?". From this it follows that typically a HOW-question cannot have multiple true resolutions in such a way that different true resolutions are "about" different events.

When HOW modifies a kind of predicate that easily allows the non-maximal interpretation of manner adverbial phrases, it is predicted that its extraction over a factive predicate is relatively easy. The judgements are admittedly subtle, but the contrast illustrated below appears to bear it out.

(23) a. ?*How does Ken know [that John spoke to Mary _]?
b. ??How does Ken know [that Mary has travelled to Chicago _]?

(ii) Multiple manners - single event
A single event can be characterized by more than one manner, in contrast to the cases of time and place.

(24) a. Alice killed Max impulsively and quickly.
b. #Alice killed Max in New York and in Chicago.
c. #Alice killed Max on Monday and on Wednesday.

Indeed, as "manners" subsume a wide range of semantic aspects (domains), there can be a large, if not infinite, number of manners that hold of a single event. What the utterer of a HOW-question seeks to obtain is obviously not the exhaustive list of such manners. Somewhat parallel to the case of WHY-questions, what counts as a legitimate answer to a HOW-question largely depends on what semantic aspect of the relevant event the utterer is interested in. Let us assume that the following is true in the current state of affairs.

(25) Alice taught Max French effectively, with emphasis on pronunciation, by private lessons, based on a contract through a major language school.

When the HOW-question in (26) is uttered, answers like (27a-d) may be considered appropriate, but they could as well be refuted as "irrelevant".

(26) How did Alice teach Max French?

(27) Alice taught Max French . . .

a. effectively
b. with emphasis on pronunciation
c. by private lessons
d. based on a contract through a major language school

When (27a) is provided as an answer, for example, the utterer of (26) may be satisfied with it, but may as well want to continue his inquiry by saying: "Well, that's not what I'm asking. Did she teach him by private lessons, or by group lessons?". It thus appears that typically a HOW-question solicits an answer concerning *a particular semantic aspect* of the relevant event. Crucially, a single event usually cannot be modified by more than one manner that belongs to the same semantic domain. For example, if an event (of teaching) is effective, it cannot be ineffective; similarly, if an event is "with emphasis on pronunciation", it cannot be "with emphasis on listening comprehension". In other words, manners belonging to the same semantic domain are generally mutually exclusive:

(28) a. effectively, ineffectively
b. with emphasis on pronunciation, listening comprehension, writing skills, . . .
c. by private lessons, by group lessons, in a large classroom setting, . . .
d. based on a contract through a major language school, through a private contract, on a volunteer basis, . . .

This means that when a particular semantic aspect is "under discussion", an interrogative with HOW has only one true resolution, i.e., is a unique wh-interrogative.

It seems not impossible, however, for a HOW-question to have true resolutions concerning multiple semantic domains. (29), for example, can be a legitimate and true answer to (26).

(29) Alice taught Max French effectively, with emphasis on pronunciation.

Furthermore, in a situation where (29) counts as a legitimate and true answer to (26), so do (30a) and (30b). This means that it is possible for both (30a) and (30b) to be true resolutions of (26), in the same context of utterance.

(30) a. Alice taught Max French effectively.
b. Alice taught Max French with emphasis on pronunciation.

In sum, the marginal status of HOW with respect to the FIC can be attributed to the fact that a HOW-question *tends to* be interpreted as a unique wh-interrogative, but can be interpreted as a non-unique wh-interrogative as well with appropriate settings/interpretative efforts.

5 Pragmatic Anomaly vs. Pragmatic Infelicity

Finally, it is worth pointing out that the account of the FIC proposed in the present work illustrates a special kind of pragmatic infelicity, which can be opposed to "contingent" pragmatic infelicity such as presupposition failure, violation of Gricean maxims, etc.

An utterance that involves a presupposition failure (e.g., "John stopped smoking" uttered in a situation where John has never smoked), for example, cannot accomplish a successful speech-act of assertion because the relevant presupposition does not conform to the way things are; under different circumstances, the same sentence can be felicitously uttered. This contrasts with the case of an FIC violation. There can be no possible situations, for instance, where an utterance like: "Why does Max regret [that Alice killed Pat _]?" makes a pragmatically felicitous question. I name this type of pragmatic infelicity *pragmatic anomaly.* The distinction between the two types of pragmatic infelicity can be analogized with the one between contingent falsehood and contradiction, and the one between accidental misdenotation and oxymoron.

Table 1. Two types of pragmatic infelicity

	contingent	**non-contingent**
falsehood	contingent falsehood	contradiction (e.g. 'it is raining and it is not raining.')
misdenotation	contingent misdenotation (e.g. 'the present king of France')	oxymoron (e.g. 'a round square')
pragmatic infelicity	contingent infelicity (e.g. presupposition failure)	**pragmatic anomaly** (e.g. FIC violation)

The question of what other cases of pragmatic infelicity belong to the class of pragmatic anomaly is left open for future research.[10]

An especially noteworthy point about pragmatic anomaly is that it is "inherited" through a (non-factive) attitude predicate. That is, when an interrogative *I* is unacceptable due to pragmatic anomaly, so is a complex sentence that contains *I* as the complement of a non-factive attitude predicate like *wonder* and *ask.*

(31) a. *Why does Ken know [that Alice insulted Pat _]?
b. *John wonders [why Ken knows [that Alice insulted Pat _]].

(32) a. ?*How does Ken regret [that Alice left _].
b. ?*John asked [how Ken regretted [that Alice left _]].

[10] Violation of the so-called negative island constraint may also be counted as a kind of pragmatic anomaly, if we adopt a discourse-based account along the lines of Kuno and Takami ([18]).

This contrasts with ordinary types of pragmatic infelicity. For example, when a (declarative or interrogative) clause C would induce a presupposition failure with respect to the present context of utterance, a sentence that contains C as a complement clause does not (necessarily) sound infelicitous ([19], [20]).

(33) (**Situation:** Linda has never written novels.)
 a. #Has Linda stopped writing novels?
 b. Ken wonders if Linda stopped writing novels.
 c. Ken asked if Linda stopped writing novels.

(34) Ken believes that Linda wrote a number of novels in the past few years, and he wonders if she stopped writing novels.

(35) Ken is trying to make John believe that Linda is keen on literature. That's why Ken asked John if Linda stopped writing novels. (Ken tried to mislead John by asking this question.)

Likewise, "John is either in the kitchen or in the bathroom" pragmatically implicates that the utterer does not know if John is in the kitchen or in the bathroom; it violates the Maxim of Quantity if uttered in a situation where the speaker knows that John is in the kitchen. If the same sentence is embedded under *believe*, *say*, etc., this implicature disappears.

(36) a. John is in the kitchen or in the bathroom.
 $+>$ The speaker does not know whether John is in the kitchen or in the bathroom.
 b. Ken believes that John is in the kitchen or in the bathroom.
 $+\not>$ The speaker does not know whether John is in the kitchen or in the bathroom.

At the present time, I do not have an answer to the question of how this contrast between pragmatic anomaly and other types of pragmatic infelicity arises. It is interesting to note that, however, a somewhat similar contrast is found between contingently false statements and contradictions, in relation to a modal possibility operator. The falsehood of a contingently false statement is not preserved when it is prefixed by a possibility operator; a contradiction, in contrast, remains to be a contradiction when embedded under a possibility operator:

(37) (**Situation:** John likes blueberries.)
 a. John does not like blueberries. (FALSE)
 b. It is (logically) possible that John does not like blueberries. (TRUE)

(38) a. John likes blueberries and John does not like blueberries. (FALSE)
 b. It is (logically) possible that John likes blueberries and John does not like blueberries. (FALSE)

I leave it an open question what this apparent parallelism between pragmatic anomaly and logical falsehood points to.

References

1. Ross, J. R.: Constraints on Variables in Syntax. Ph.D. thesis, Massachusetts Institute of Technology (1967)
2. Szabolcsi, A.; Zwarts, F.: Weak Islands and an Algebraic Semantics for Scope Taking. Natural Language Semantics **1** (1993) 235–284
3. Rooryck, J.: Negative and Factive Islands Revisited. Journal of Linguistics **28** (1992) 343–373
4. Manzini, R.: Locality: A Theory and Some of its Consequences. MIT Press, Cambridge (1992)
5. Melvold, J.: Factivity and Definiteness. MIT Working Papers in Linguistics, Vol. 15: More Papers on *Wh*-movement (1991) 97–117
6. Hegarty, M.: On Adjunct Extraction from Complements. MIT Working Papers in Linguistics **13** (1990) 101–124
7. Cinque, G.: Types of A'-dependencies. MIT press, Cambridge (1990)
8. Adams, M.: Government of Empty Subjects in Factive Clausal Complements. Linguistic Inquiry **16** (1985) 305–313
9. Rizzi, L.: Relativized Minimality. MIT Press, Cambridge (1990)
10. Pollard, C., Sag, I. A.: Head-driven Phrase Structure Grammar. The University of Chicago Press, Chicago (1994)
11. Liberman, M.: Some Observations on Semantic Scope. M.A. thesis, Massachusetts Institute of Technology (1973)
12. Landman, F.: Events and Plurality. Kluwer, Dordrecht (2000)
13. Hamblin, C.: Questions. Austorasian Journal of Philosophy **36** (1958) 159–168
14. Karttunen, L.: Syntax and Semantics of Questions. Linguistics and Philosophy **1** (1977) 3–44
15. Groenendijk, J., Stokhof, M.: Studies on the Semantics of Questions and the Pragmatics of Answers. Ph.D. thesis, University of Amsterdam (1984)
16. Hintikka, J., Halonen, I.: Semantics and Pragmatics for Why-questions. Journal of Philosophy **92** (1995) 636–657
17. Sudo, Y.: The Semantics of Reasons. Ms., University of Tokyo.
18. Kuno, S.; Takami, K.: Remarks on Negative Islands. Linguistic Inquiry **28** (1997) 553–576
19. Heim, I.: Presupposition Projection and the Semantics of Attitude Verbs. Journal of Semantics **9** (1992) 183–221
20. Oshima, D.: Perspectives in Reported Discourse. Ph.D. thesis, Stanford University (2006)

Aspects of the Indefiniteness Effect*

Linton Wang[1] and Eric McCready[2]

[1] Department of Philosophy, National Chung Cheng University
lintonwang@ccu.edu.tw
[2] Department of English, Aoyama Gakuin University
mccready@cl.aoyama.ac.jp

Abstract. This paper proposes an account for the indefiniteness effect discussed in [1]. We shall argue that the effect arises from (a) the tendency of indefinites in certain interrogative constructions to take wide scope over WH-terms, resulting in 'quantifying into' questions, and (b) the pragmatic impropiety of interrogatives with wide scope taking indefinites to form speech acts for requesting answers. Our proposal is formally implemented using the interpretation of quantifying into questions as choice readings provided by [2] and [3].

1 Introduction

[1] first noticed the *the indefiniteness effect*, seen in (1a,b) and (2a,b), where occurrences of indefinites in the interrogative construction *What is wrong with ___?* and *Where is ___?* (hereafter *anti-indefinite constructions*) are infelicitous.

(1) a. What is wrong with John/the student/every student/him (the three students/most students/you/his students)?
b. What is wrong with a student (/*three students/*no more than three students/*at least three students/*no students)? ([1]: 11)

(2) a. Where is (are) the man/John/every man/(you /most men)?
b. Where is (are) *a man/*no man/(*two men/*at least three men)?

The indefiniteness effect, as noted in [1], can also be observed in a variety of other interrogative constructions such as *What do you think about ___?*, *Are you ok with ___?*, *What can you say about ___?*, etc.[1] [1] did not offer an account for the effect, however. The objective of this paper is to remedy this lack by proposing an account.

* The authors would like to thank the comments from the participants in the conference LENLS 2006. The first author gratefully acknowledges the support from Taiwan National Science Council under the project number 94-2411-H-194-034.

[1] If the reader finds it difficult to detect the infelicity in (1b) and (2b), imagine how one might respond when asked questions.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 162–176, 2007.

The interest of the indefiniteness effect initially comes from its being a counterpart of the *definiteness effect* in the existential *there-be* construction (cf. [4], [5], and [6]) and the possessive *have* construction (cf. [7] and [8]), shown in examples (3a,b) and (4a,b), respectively.

(3) a. There is (are) a man/at least one man/(two men/many men/at least two men)?
    b. There is (are) *John/*the man/*every man/(*most men/*the men)?
(4) a. John has a daughter/ at least one daughter/(two daughters/many daughters)?
    b. John has *the daughter/*every daughter/(*most daughters)?

We shal call the constructions which exhibit the definiteness effect *anti-definite constructions*. One might expect that the indefiniteness effect to be accounted for as an "inverse" of a theory of the definiteness effect. Nonetheless, as already indicated in [1], there is no existing theory of the definiteness effect that can be transformed into an account for the indefiniteness effect. Existing theories of the definiteness effect (cf. [9], [5], [10], etc.) may at best provide a classification of the quantifiers in the effect, but fail to provide an explanation for the effect.[2]

The indefiniteness effect is even more perplexing than it looks. Even though there is a robust intuition that (1b) and (2b) are infelicitous, they also seem to have some felicitous interpretations, which make the indefiniteness effect further diverge from the definiteness effect. To begin with, some speakers have the intuition that, for example, *Where is a man?* in (2b) can receive a felicitous interpretation as *Where is there a man?*. If this is right, there will be felicitous interpretations for the other examples in (2b). We are not sure how clear this intuition is; we don't share it completely. In any case, if *Where is a man?*, as well as the other examples in (2b), can really receive a felicitous interpretation like *Where is there a man?*, it is not the interpretation that we aim to account for (we'll say more about the interpretation we think is most salient shortly). Moreover, this paraphrase strategy cannot be extended to the cases in (1b); it's totally unclear how to apply the same paraphrase strategy for the infelicitous *What is wrong with a man?*. Doing so would yield something like *What is there wrong with a man?* or *What is wrong with that there is a man*, which is senseless. What we aim to pursue is an interpretation of indefinites that robustly induces the infelicitous occurrences in the interrogative constructions like *What is wrong with ___?*, *Where is ___?*, etc.

[1] shows a second type of felicitous interpretations for indefinites in anti-indefinite constructions, that arises from the generic interpretation of indefinites.

[2] We would like to further indicate that a classification of quantifiers based on the definiteness effect cannot be applied to the indefiniteness effect. For example, by a classification according to the definiteness effect, the quantifier *neither man* should be classified as compatible with the anti-definite constructions. Based on the classification, taking the indefiniteness effect as an inverse of the definiteness effect predicts that *neither man* should have felicitous occurrences in anti-indefinite constructions. However, this prediction is wrong, since *What is wrong with neither man?* sounds infelicitous.

The effect of the generic interpretation of indefinites is even more salient in (5a,b) where the indefinites are easily understood as generic.

(5) a. What is wrong with a man with two mistresses?
b. Are you ok with a man with two mistresses?

(5a) has a salient generic interpretation with a quasi-universal flavor, which is similar to *What is wrong with every man with two mistresses?* or *In general, what is wrong with a man with two mistresses?*; (5b) has a salient generic interpretation which is similar to *Are you ok with every man with two mistresses?* or *In general, are you ok with with a man with two mistresses?* It has long been noticed that the generic interpretation of indefinites carries certain features of 'definiteness.' The role of this definiteness feature in the indefiniteness effect will be elaborated on later in the paper.

Some people may have the intuition that (1b) and (2b) can receive felicitous interpretations from either wide scope or specific interpretations of indefinites.[3] We do not share this exact intuition. Nonetheless, it is clear to us that indefiniteness effect constructions can be made felicitous by embedding them into declarative sentences, e.g. (6a,b). Here the indefinite indeed receives a wide scope or specific interpretation.

(6) a. What is wrong with many professors is that they don't research anymore.
b. I know where a (certain) man is.

Once we embed the infelicitous *What is wrong with many professors?* in (6a), and transform the infelicitous interrogative *Where is a man?* into an indirect embedded interrogative in (6b), the sentences become fine. We also need an explanation for this effect.

In this paper, we explain why theories of interrogatives in the literature fail to account for the indefiniteness effect, and we provide a proposal to account for it. Our proposal has it that there is nothing syntactically and semantically wrong with the occurrences of indefinites in the anti-indefinite constructions. Syntactically, indefinites in anti-indefinite constructions are grammatical. Semantically, we take it that they can receive felicitous narrow scope interpretations according to [11] and [2], and felicitous wide scope interpretations (quantifying into questions) from [2] or [3] as choice interpretations. However, pragmatically speaking, indefinites in anti-indefinite constructions tend to take wide scope interpretations, and wide scope interpretations of indefinites in interrogatives do not constitute proper speech acts that request answers. We suggest that it is this pragmatic fact that induces the indefiniteness effect.

The paper is organized as follows. In section 2 and 3, we consider an account for the indefiniteness effect based on certain syntactic and semantic aspects of interrogatives discussed in the literature and show that we still fail to find a

[3] Some speakers have told us that they find *Where is no man?* acceptable under some interpretation, but we do not have any idea about what it might be.

proper account from such considerations. In section 4, we propose our pragmatic account for the indefiniteness effect based on the idea that indefinites in anti-indefinite constructions tend to take wide scope, which results in a failure to express coherent answer-requesting speech acts. Due to space limitations, we shall skip all formalism that we find inessential (though nonetheless possibly helpful) in the present paper.

To see why the the effect resists the regimentation of existing theories of interrogatives (according to standard accounts in the literature), we consider the syntactic, semantic, and pragmatic aspects of interrogatives corresponding to interrogatives as sentences, questions, and speech acts, respectively. In the next section, we begin with their syntax.

2 The Syntactic Aspect

To begin with a possible syntactic approach to the indefiniteness effect, we see no reason to think the infelicity of indefinites in these constructions is the result of any syntactic feature of anti-indefinite constructions. Although we can imagine a proposal that accounts for the indefiniteness effect by assuming that anti-indefinite constructions are incompatible with the [+indefinite] feature, this proposal is relatively *ad hoc*, and fails to account for the related felicitous interpretations of indefinites in (2b).

We would like to further suggest that no syntactic theory should even try to account for the indefiniteness effect, especially in a transformational grammar. It is generally assumed in transformational theories that interrogatives are constructed through movement from declaratives.[4] If this is so, then once an declarative sentence is grammatical, it should be that the interrogative generated from it is also grammatical. For example, since an indicative sentence like *John is ok with a man* is grammatical, the grammaticality of the interrogative *Is John ok with a man?* should be guaranteed. If the indefiniteness effect can be accounted for by syntax, a syntactic theory will need to show that the infelicity of *Is John ok with a man?* is its ungrammaticality, which seems to be quite implausible.

Even if a theory is willing to accept that *Is John ok with a man?* is ungrammatical, it also needs at the same time to account for the felicity of (7).

(7) I do not know whether John is ok with a student.

A syntactic account for the indefiniteness effect should tell us why the indefiniteness effect disappears when *Is John ok with a man?* turns into an indirect interrogative and is embedded. We do not see any simple, non-stipulative way of doing this.

Another reason to resist a syntactic account for the indefiniteness effect comes from the heterogeneity of anti-indefinite constructions. Some anti-indefinite constructions can be base generated, e.g. *Is John ok with a man?*, but some of them

[4] Though this may not hold for recent Minimalist theories.

do not seem to be base generated, e.g. *What is wrong with a man?* and *What do you think about a man?*.[5]

3 The Semantic Aspect

Now let us consider the semantic aspect of the indefiniteness effect. Following the standard semantic analysis of interrogatives in the literature, we take the meaning of interrogatives, questions, to be identified as their answers. We shall start our semantic consideration of the indefiniteness effect from narrow scope, wide scope, and pair-list interpretations. As we shall show, no previously mentioned interpretation of interrogatives is suitable to account for the indefiniteness effect. Of course, there are other semantic theories concerning the interpretation of quantifiers in interrogatives (cf. [12] and [13]), but we shall skip them in this paper since we fail to see their significance as a solution for the indefiniteness effect.

Narrow Scope Interpretations
To begin with, consider the interpretation of quantifiers taking narrow scope in interrogatives that exhibit the indefiniteness effect. This type of interpretations is characterized by answers such as 'every man is in New York' for (8a) and 'a man is in New York' for (9a). Formally, (8a) and (9a) receive the narrow scope readings (8b,c) and (9b,c) from the Karttunen-style analysis in [11] and Groenendijk & Stokhof-style analysis in [2], respectively.

(8) a. Where is every man?
b. $\lambda p.\exists y({}^{\vee}p \wedge p ={}^{\wedge} \forall x(man(x) \rightarrow at(x, y)))$
c. $?y.\forall x(man(x) \rightarrow at(x, y))$

(9) a. *Where is a man?
b. $\lambda p.\exists y({}^{\vee}p \wedge p ={}^{\wedge} \exists x(man(x) \wedge at(x, y)))$
c. $?y.\exists x(man(x) \wedge at(x, y))$

With respect to the Karttunen-style analysis and the Groenendijk & Stokhof-style analysis, both narrow-scope readings for (8a) and (9a) are felicitous. Moreover, if we take a closer look into the interpretations of (9b,c), we shall find that they represent the interpretation of the interrogative *Where is there a man?*. We do not deny that the narrow scope reading for (9a) is available to some degree, but we take it that it is not the intended interpretation for (9a) because, although the logical forms present no difficulties in construction, still the persistent intuition is that the construction is odd.

[5] Some constructions do not like to have indeifnites in subject positions. For example, *A man is John* is grammatical, but still sounds weird. One may suggest that the anti-indeinite constructions are such constructions that disprefer indefinites. We do not take this suggestion as impossible, but we find it difficult to see a common property among various anti-indefinite constructions.

Wide Scope Interpretations
The second possibility is to consider the wide scope interpretation for the quantifiers in (8a) and (9a), paraphrased as (10a,b) respectively, which takes the form of so-called *quantifying into questions*.[6]

(10) a. For every man x, where is x?
b. For a man x, where is x?

As indicated in [14], possible Karttunen-style analyses of (10a) are (10a-1) and (10a-2) and (10b) as (10b-1) and (10b-2), but they all fail to be correct.

- (10a-1) $\forall x(man(x) \rightarrow \lambda p.\exists y(^{\vee}p \wedge p =^{\wedge} at(x,y)))$
- (10a-2) $\lambda p.\forall x(man(x) \rightarrow \exists y(^{\vee}p \wedge p =^{\wedge} at(x,y)))$
- (10b-1) $\exists x(man(x) \wedge \lambda p.\exists y(^{\vee}p \wedge p =^{\wedge} at(x,y)))$
- (10b-2) $\lambda p.\exists x(man(x) \wedge \exists y(^{\vee}p \wedge p =^{\wedge} at(x,y)))$

In (10a-1) and (10b-1), the consequent of $\rightarrow$ and $\wedge$ does not have the right type t. On the other hand, (10a-2) does gives us a suitable interpretation: however, it denotes the empty set, if there is more than one man.[7] (10b-2) provides an interpretation which is quite similar to the narrow scope interpretation except that both man and location must be specified. So it is clear that, even if it is the case that the indefiniteness effect arises from quantifying into questions, Karttunen-style analyses do not account for it since, on this analysis, both definites and indefinites are infelicitous in anti-indefinite constructions, *qua* (10a-1) and (10b-1), or they are both felicitous in anti-indefinite constructions, *qua* (10a-2) and (10b-2).

On the other hand, according to Groenendijk & Stokhof-style analysis, (10a) receives (10a-3) and (10b) receives (10b-3) as their interpretations.

- (10a-3) $\lambda j.\lambda i \forall y(man(j)(y) \rightarrow \lambda x.(at(i)(x)(y)) = \lambda x.(at(j)(x)(y)))$
- (10b-3) $\lambda j.\lambda i \exists y(man(j)(y) \wedge \lambda x.(at(i)(x)(y)) = \lambda x.(at(j)(x)(y)))$

While (10a-3) gives a sensible interpretation, (10b-3) fails to gives us the correct interpretation for (10b): we do not get a partition of possible worlds unless there is exactly one man in the model.

The Groenendijk & Stokhof-style analysis for quantifying into questions may cause people to think that we have an account for the indefiniteness effect: definites are allowed for quantifying into questions, but indefinites are not. This suggestion is correct in a sense, but there are two worries concerning it. First, according to the proposal, the wide scope interpretation for *Where are most men?* should be infelicitous, since Groenendijk & Stokhof-style analysis predicts that it does not form a partition of worlds. This prediction fails to be correct.

---

6 Our informal analysis of quantifying into questions is adopted from [14].

7 Consider two men m_1 and m_2. (10a-2)=$\{p|\exists y(^{\vee}p \wedge p =^{\wedge} at(m_1,y)) \wedge \exists y(^{\vee}p \wedge p =^{\wedge} at(m_2,y))\}$, which is the empty set. [14] wrongly states that (10a-2) is the empty set unless every man is in the same place. This mistake was first pointed out to us by Bernhard Schwarz.

Second, in this account, there is nothing specifically mentioned concerning the anti-indefinite constructions, which are after all the reason that the indefiniteness effect arises.

Moreover, [2] takes both (10a,b) to be felicitous, so they further introduce so-called *choice readings* (10a-4,b-4) for both (10a,b).

- (10a-4) $\lambda Q.\exists W($ W is a witness set of [every man]$\wedge$ $Q = \lambda j.\lambda i.([\lambda y.\lambda x.W(y) \wedge at(i)(x,y)] = [\lambda y.\lambda x.W(y) \wedge at(j)(x,y)]))$
- (10b-4) $\lambda Q.\exists W($ W is a witness set of [a man]$\wedge$ $Q = \lambda j.\lambda i.([\lambda y.\lambda x.W(y) \wedge at(i)(x,y)] = [\lambda y.\lambda x.W(y) \wedge at(j)(x,y)]))$

According to [9], a set A is a witness set of a quantifier $D(B)$ iff $A \subseteq B$ and $D(B)(A)$, so every choice reading is a set of equivalence relations. Both (10a,b) receive felicitous choice readings, unless the set of men is empty. If the set of men is empty, then (10a,b) both denote the empty set, which does not gives rise to a suitable restricted question.[8] Informally, the choice reading for (10b), (10b-4) (for instance) can be understood as saying: *Pick a man x, where is he?*[9] Again, since both definite and indefinites are available for choice readings, we still lack an account for the indefiniteness effect.

Pair-List Readings
Our last consideration is the availability of the pair-list readings for (8a) and (9a). While it is clear that indefinites do not have pair-list readings, definites like *most men* also fail to have pair-list readings, e.g. (11a,b) (assuming that there are only three guests).

(11) a. Which dish did most guests make?
b. *Al the pasta, and Bill the salad. ([14])

So the indefiniteness effect cannot be simply accounted for by the fact that pair-list readings are unavailable for for indefinites in general. Moreover, the unavailability of pair-list readings for indefinites is a general feature of many interrogatives. If the unavailability of pair-list readings accounted for the indefiniteness effect, we would expect the indefiniteness effect to be even more pervasive than we have seen it to be.

4 The Pragmatic Aspect

Neither syntactic nor semantic considerations led to a reasonable explanation of the indefiniteness effect. Therefore we would like to turn our focus now to the pragmatic aspect of interrogatives.

[8] Why does an empty set of equivalence relation not give rise to a suitable question interpretation? We shall come back to this in section 4.

[9] To capture this intuition, one more constraint should be added into the definition of witness sets for $a(B)$: A is a witness set of $a(B)$ only if $|A| = 1$.

The Pragmatic Reconsideration of the Effect

We shall begin with a reconsideration of anti-indefinite constructions in the context of conversations. First, consider the following dialogue between two agents.

(12) a. A: Where is a man?
b. B: Wait a minute, who are you asking about?
c. A: I am asking about the man who stole my computer.
d. B: He is at the police station now.

(12a) sounds infelicitous out of context, but, as shown by (12), it improves in conversation.

Consider what the conversation (12) tells us. The felicity of (12b) cannot come from the narrow scope interpretation of (12a) asking *Where is there a man?*. Otherwise, (12d) would already be a direct partial answer for (12a). Instead, (12b) reflects the agent B's inability to answer A's question (12a). We can conclude several things from the conversation (12). First, the felicity of (12a) in a conversational context indicates that the infelicity of (12a) by itself, as we have suggested, has nothing to do with the semantic interpretation of (12a); so it is semantically interpretable. Second, we propose that the out of conversation infelicity of (12a) arises from B's inability to answer A's question without further information concerning A's question. This indicates that the out of conversation infelicity of (12a) is a pragmatic consequence rather than a semantic consequence, i.e. (12a) cannot be answered without further information. Since we suppose that the out of conversation infelicity of (12a) comes from pragmatics, this result does not conflict with the robust intuition in the literature that the meaning of a interrogative is characterized by its answers.

Further evidence for the idea that the infelicity of (12a) arises from needing further elaboration on the question comes from the following two cases. First, consider the conversation (13).

(13) a. A: Where is a certain man?
b. B: Wait a minute, who are you asking about?
c. A: I am asking about the man who stole my computer.
d. B: He is at the police station now.

(13a) alone sounds infelicitous. Moreover, it does not help to take (13a) on the existential interpretation *Where is there a certain man?*, which still sounds infelicitous. A proper understanding of (13) is to take (13a) as a semantically felicitous question, but a pragmatically unanswerable one: the specific adjective *certain* indicate that the question asker has someone specific in mind, but the listener cannot answer the question unless what's in the speaker's mind is further specified explicitly. Once what the speaker has in mind is further specified, e.g. as in (13c), the listener can answer the question.

On the other hand, we can consider the conversation (14) as a comparison. Assume that the domain of quantification is not under discussion here.

(14) a. A: Where is every man?
b. B: *Wait a minute, who are you asking about?

The infelicity of (14b) can be understood as showing that (14a) is both semantically felicitous and pragmatically clear, so it is improper to ask for information for further specification.

To summarize what we have said, we suggest that the indefiniteness effect arises from a need for further information in order to answer the question under certain semantic interpretations of interrogatives. Or, in other words, the indefiniteness effect arises from the fact that some questions fail to be a proper request for answers on certain interpretations, unless they are further specified.

Our proposal can only make sense if it can be extended to determiners other than *a* and *every*. Consider, first, other indefinite determiners.

(15) a. A: Where are two men/at least two men/at most two/no men/neither men?
b. B: Wait a minute, who are you asking about?

(15a,b) indicates that indefinite determiners in general pass the test for information specification of questions.

On the other hand, consider the following definite determiners.

(16) a. A: Where is (are) the man/(you/the men/the three men)?
b. B: *Wait a minute, who are you asking about?

As shown by (16a,b), determiners for definites in general cannot serve as a basis for further informational specification.

A Formal Account

Our formal account for the indefiniteness effect based on the previous observation has two components: first, quantifiers in anti-indefinite type interrogatives tend to receive the wide scope interpretations, which are assumed in the case of indefinites to be captured by choice interpretations, which are sets of equivalence relations, since indefinites may have multiple witness sets. Second, since an equivalence relation represents a speech act for requesting answers, a non-singleton set of equivalence relations is *pragmatically inappropriate* since it either represents *multiple (concurrent)* question asking speech acts(for sets that have cardinality more than one, e.g. for the determiner *a*), or does not even ask a question at all (for sets with cardinality zero, e.g. for the determiner *no*).

The first part of the proposal concerns the tendency of indefinites to take wide scope in anti-indefinite constructions. Take *Where is a man?* as an example. This is infelicitous, but this shouldn't be the case if the indefinite gets narrow scope. One reason for the non-salience of the *Where is there a man?* interpretation is, we think, that the interpretation has a better way to be represented, i.e. the unambiguous utterance *Where is there a man?*. Clearly this is a better way to ask a question about where there is a man than *Where is a man?*. By the Gricean principle of Manner, if the speaker is really interested in asking where there is a man, he would use *Where is there a man?* rather than *Where is a man?*. The same idea applies to *Is John ok with a man?*; there, the narrow scope interpretation can be better asked by *Is there a man such that John is ok with?*[10]

Further, it is difficult to provide a narrow scope interpretation for indefinites in some anti-indefinite constructions. For example, what does the narrow scope interpretation for *What is wrong with a man?* mean? It cannot be the infelicitous *What is wrong with that there is a man?* Similarly, the infelicitous *What do you think about that there is a man?* does not represent the narrow scope reading of *What do you think about a man?* We think that for the above two cases, no narrow scope readings are available.[11] This explains why no feliciotus interpretation is available for *What is wrong with a man?*, as indicated in section 1. We may compare the above two cases with the interrogative *which dish did a guest order?*, which has a felicitous narrow scope interpretation as *Which dish is such that there is a guest who ordered it?*

For the second part of our proposal, we first consider three types of quantifiers with respect to choice interpretations. The first type of quantifiers are those which can have multiple witness sets, e.g. *a man*, *some men*, etc. Their corresponding choice readings are sets of equivalence relations which can have cardinality more than one. The second type of quantifiers are those which cannot have any witness sets, e.g. *no men*, *neither men*, etc. Their corresponding choice readings are the sets of equivalence relations which can only have cardinality zero. The third type of quantifiers are those which can have at most one witness set, e.g. *every men*, *the men*, etc. Their corresponding choice readings are the sets of equivalence relations which can only have cardinality one or zero. With repsect to the indefiniteness effect, can see that the first and the second type have infelicitous occurrences in anti-indefinite constructions, but the third type are felicitous there. The only exceptional case involves quantifiers with the determiner *most*, e.g. *most men*, which can have multiple witness sets and so belongs to the first type of quantifiers, but has felicitous occurrences in anti-indefinite constructions. We shall deal with this case later.

The issue is now how the pragmatic felicity of choice readings is related to the cardinality of their corresponding sets of equivalence relations. Our proposal is to assume that every equivalence relation represents a speech act of requesting answers. Three cases must be considered.

---

[10] Rick Nouwen (p.c.) asks us about quantifiers like *at least n students*, which may have a tendency to take narrow scope. If this tendency is real, there is a difficulty for our proposal in that it will not extend to such quantifiers. We are not sure about this; we seem to be able to get wide scope readings with respect to universal quantifiers and modals without much difficulty. For example, it is quite natural to give a wide scope interpretation for *at least three papers* in the sentence *every student should read at least three papers assigned by the professor*. If there is such a tendency for *at least n students* to take narrow scope, however, it is clearly pragmatic in nature, and thus overridable; we guess that anti-indefinite constructions are contexts in which it does in fact get overriden, again for pragmatic reasons.

[11] One common feature for these anti-indefinite constructions is that the quantifiers are the objects of prepositions. We suspect that this feature may have something to do with the tendency of such quantifiers to take wide scope.

First Case. Consider the first case, in which the set of equivalence relations arises from more than one witness set and has cardinality more than one. It represents what we shall call a group of *concurrent* questions, where the speaker is actually asking many questions at the same time. For example, assume that there are exactly three men, John, Jack, and Bill. According to (9d-1), (9a) is actually asking as many questions at the same time (concurrently) as there are different witness sets for *a man*.[12] Thus in our situation the question is asking *Where is John?*, *Where is Jack?*, and *Where is Bill?* at the same time. Even though there is no semantic difficulty with asking many questions concurrently, the listener cannot identify which question the speaker is actually asking and so cannot determine how to answer the question in order to satisfy the request for an answer. This leads to pragmatic inappropriateness. Note that it would be a mistake to answer all of the concurrently asked questions, since this would be answering the question *Where is every man?*.

Second Case. The second case concerns choice readings whose corresponding sets of equivalence relations arise from no witness sets at all and so have cardinality zero. Again, assume that there are exactly three men, John, Jack, and Bill. According to this scenario, as well as to other scenario, the choice interpretation for *Where is no man?* is an empty set of equivalence relations. Given that every equivalence relation represents a speech act for requesting answers, a question that requests nothing is taken to be pragmatically infelicitous.

Third Case. The third case concerns choice readings whose corresponding set of equivalence relations arises from exactly one witness set and so has cardinality one. Again, assume that there are exactly three men, John, Jack, and Bill. According to this scenario, the choice interpretation for *Where is every man?* is a set of equivalence relations which has cardinality one, so only one request for answers is represented. This is a pragmatically felicitous question.

The above consideration of three cases corresponds nicely with the discussion about question acts in [14], based on the definition 1.

Definition 1. ***Question Acts Transformation***

1. $(q_1 \& q_2) \mapsto \{answer(q_1) \cup answer(q_2)\}$
2. $(q_1 \vee q_2) \mapsto \{answer(q) : q \in \{q_1, q_2\}\}$

According to definition 1, every conjunctive question act corresponds to a set of equivalence relations with cardinality one, and every disjunctive question act corresponds to a set of equivalence relation with cardinality more than one. Of course, we can then take a null question act to correspond to a a set of equivalence relations with cardinality zero. As argued in [14] and above, conjunctive question acts are felicitous, but disjunctive and null questions acts are infelicitous. This is exactly parallel to our previous analysis for the three cases. A crucial difference between our proposal and Krifka's proposal in [14] is that we would like to take the (in)felicity of question acts to be pragmatic rather than semantic. This is a debatable issue, but we shall not go into it further here.

[12] Remember we define a witness set A of a quantifier $a(B)$ as a set such that (i) $A \subseteq B$, (ii) $a(B)(A)$, (iii) $|A| = 1$.

Now we come back to the first type of quantifier. Depending on the model of interpretations, such quantifiers may have zero, one, or more than one witness sets. However, we take situations in which there is more than one witness set to be salient in a conversation. Otherwise, the conversation would be pretty bizarre. For example, it is pretty odd to say *A man is happy* if everyone involved in the conversation knows that there is only one man in the world. So we take the choice readings corresponding to the first type of quantifier to be pragmatically infelicitous in conversation. The choice readings corresponding to the second type of quantifiers are also pragmatically infelicitous, since the corresponding set of equivalence relation is always empty. The choice readings corresponding to the third type of quantifiers are pragmatically felicitous, since the salient corresponding set of equivalence relations has cardinality one, even though there is a non-salient possibility that the set has cardinality zero. More generally, in order for a question to be answerable, the set of equivalence relations must be a singleton. Otherwise, the listener has no way to identify an equivalence relation and then provide a proper answer.

Now we can provide an account for the conversation (12) based on our proposal. With wide scope taking indefinites, (12a) represents a set of equivalence relations. Nonetheless, B cannot answer the question (12a), for he cannot identify an equivalence relation as a basis for providing an answer. Instead, B's (12b) asks A to identify the proper equivalence relation for asking an answer. Then A's (12c) signifies the proper equivalence relation, and at the same time eliminates other equivalence relations. As a result, B has a way to answer A's question: (12d). But keep in mind that B's answer (12d) is not an answer for (12a). Rather, it is an answer to for the elaborated question *Where is a man? I am asking about the man who stole my computer.* The elaboration here is in fact an instance of backgrounding in the sense of [15] and [16].

There are some cases in need of further attention: the determiner *most*, generic interpretations of indefinites, and embedded indirect interrogatives.

The Case of Most. When definite quantifiers containing the determiner *most* (e.g. *most men*) take wide scope over anti-indefinite constructions, they can generate sets of equivalence relations with cardinality greater than one, since they can have multiple witnesses. Therefore, our proposal seems to predict that quantifiers like *most men* should be infelicitous in anti-indefinite constructions. Nonetheless, *most men* has feliciotus occurrences in anti-indeinite constructions as in (1a) and (2b). This seems to be a problem for us.

In fact, though, the situation is more complex. Even though *most men* has felicitous occurrences in anti-indefinite constructions as in (1a) and (2b), unlike other definite quantifiers like *every man* it passes the *wait a minute* test, as shown by the conversation (17). Here the model contains only three men, John, Bob, and Jack).

(17) a. A: Where are most men?
b. B: Wait a minute, who are you asking about?
c. A: I am asking about John and Bob.
d. B: John, New Your. Bob, Amsterdam.

The felicitous conversation (17) seems to indicate that, according to our proposal, *most men* should have infelicitous occurrences in anti-indefinite constructions.

We propose that there is an interpretation for *Where are most men?* concerning a certain set of men, but it is not a salient interpretation. We have a intuitively good answer by answering that most men are in New York for the narrow scope interpretation of *Where are most men?*; thus the narrow scope reading is salient. If right, this idea predicts that both (18b,b-1) are intuitively inappropriate answers for (18a).

(18) a. Where are most men?
b. *They are at New York. (concerning a certain set of men)
b-1. *Bob, New York. John, Amsterdam. (for pair-list reading)

As shown by (18), the prediction is correct. On the other hand, in the conversation (17), the question (17b) makes the non-salient, or even generally infelicitous, interpretation become salient. This accounts for the felicity of the conversation.

We still need to answer why the narrow scope readings of *most men* in anti-indefinite constructions are salient, but the wide scope readings of *most men* are not. Again, we appeal to the Gricean principle of Manner. If the speaker is interested in asking where a certain set of men is, the question *where are most men?* does not convey his intention fully. On the other hand, if a set of men can be singled out, then the reading of *most men* concerning a certain set of men will be available. This, we believe, is what is going on in (17).

The Case of Generic Interpretations. Now we consider how the generic interpretation of indefinites licenses indefinites in anti-indefinite constructions, as shown by (5a,b). Our account is based on the idea that the cases for multiple witness sets do not arise for generic interpretation of indeifnites. For example, generic interpretations of indefinites in anti-indefiniteness constructions like *What is wrong with a man with two wives?* receive an interpretation similar to the narrow scope interpretation of *What is wrong with every man with two wives?*. The same is true of *Are you ok with a man with two mistresses?* with the narrow scope interpretation *Are you ok with every man with two mistresses?*, as well as for *What does a woman want?* with the narrow scope interpretation *What does every woman want?*

The Case of Embedded Indirect Interrogatives. Our last consideration involves anti-indefinite constructions embedded as indirect interrogatives.

(19) a. I know where a man is.
b. I know where every man is.

Both the determiner *a* and the determiner *every* are felicitous in (19a,b). But keep in mind that the felicity of *a* in (19a) does not necessarily come from the interpretation *I know where there is a man.* Another available interpretation for (19a) concerns a certain man: for a certain man, I know where he is, as shown by (19a-1).

– (19a-1) For a man x, I know where x is.

In case we have a wide scope indefinite interpretation for (19a), as shown by (19a-1), we do not form a set of equivalence relations for the embedded question, but an equivalence relation with respect to a certain man. Under this semantic interpretation, (19a-1) does not actually request answers. As we have argued, the problem with anti-indefinite constructions is pragmatic and involves the speech act of questioning; thus, there will be nothing wrong with a wide scope indefinite interpretation of (19a), which accounts for the felicity of (19a-1). Simply put, embedded indirect interrogatives are not speech acts for requesting answers.

5 Concluding Remarks

Some interesting points concerning the investigation of interrogatives follow from out proposal for the indefiniteness effect. First, an investigation of interrogatives is complicated by the interaction among syntactic, semantic, and pragmatic aspects of interrogatives: a request for answers is conveyed by certain linguistic meanings under certain syntactic constructions. When we consider the infelicity of an interrogative, it is essential but difficult to know which aspect the infelicity follows from. What we propose in this paper is that the indefiniteness effect results from the pragmatic aspect rather than the syntactic and semantic aspects of interrogatives: (a) indefinites in anti-indefinite constructions tend to take wide scope, and (b) wide scope taking quantifiers over Wh-terms can be understood as quantifying into questions, and receive choice interpretations which are pragmatically inappropriate for indefinites. Essentially, the indefiniteness effect can be understood as arising from the ability of indefinites to have multiple witness sets, or no witness sets (for determiner *no*).

Careful readers may have noticed that our proposal has a close relationship with the issue concerning pair-list interpretations, if we take pair-list readings to arise from quantifying into questions. The idea is roughly this: by definition 1, we can convert Krifka's account for pair-list interpretations into an account for pair-list interpretations based on choice interpretations, by making use of question acts. As indicated earlier, a crucial difference is that this new sort of account is pragmatically oriented, rather than semantically oriented as in [14]. On the other hand, we also believe that our understanding on scope taking of quantifiers in interrogatives can also be applied to account for other quantifier related phenomena concerning interrogatives, for example, the *simplest hard problem* from Higginbotham concerning the infelicity of *Do you see the man in a red hat?* (cf. [17]). We shall also leave all these issues for future research.

References

1. Wang, L., McCready, E.: The Indefiniteness Effect. *Snippets* 11 (2005) 11-12
2. Groenendijk, J., Stokhof, M.: Studies on the Semantics of Questions and the Pragmatics of Answers. PhD Thesis, University of Amsterdam (1984)

3. Higginbotham, J.: Interrogatives. In Hale, K., Keyser, S., eds.): The View from building 20: Essays in honor of Sylvain Bromberger. MIT Press, Cambridge, MA (1993) 195-227
4. Milsark, G.: Toward an Explanation of Certain Peculiarities of the Existential Construction in English. Linguistic Analysis 3 (1977) 1-29
5. Keenan, E.: A Semantic Definition of 'Indefinite NP'. In Reuland, E., ter Meulen, A., eds.: The Representation of (In)definiteness. MIT Press, Cambridge, Mass. (1987) 286-317
6. McNally, L.: Existential Sentences without Existential Quantification. Linguistics and Philosophy 21 (1998) 353-392
7. Bach, E.: Have and Be in English Syntax. Language 43 (1967) 462-485
8. Partee, B.: Weak NP's in HAVE Sentences. In Gerbrandy, J., Marx, M., de Rijke, M., Venema, Y., eds., JFAK [a Liber Amicorum for Johan van Benthem on the occasion of his 50th Birthday; CD-ROM]. University of Amsterdam, Amsterdam (1999) (Accessible from `http://www.illc.uva.nl/j50/contributions.html`)
9. Barwise, J., Cooper, R.: Generalized Quantifiers and Natural Language. Linguistics and Philosophy 4 (1981) 159-219
10. Diesing, M.: Indefinites. MIT Press, Cambridge Mass. (1992)
11. Karttunen, L.: Syntax and Semantics of Questions. Linguistics and Philosophy 1 (1977) 3-44
12. Higginbotham, J., and May, R.: Questions, Quantifiers and Crossing. Linguistic Review 1 (1981) 41-80.
13. Chierchia, G.: Questions with Quantifiers. Natural Language Semantics 1 (1993) 181-234
14. Krifka, M.: Quantifying into Question Acts. Natural Language Semantics 9 (2001) 1-40
15. Asher, N., and Lascarides, A.: Questions in Dialogue. Linguistics and Philosophy 21 (1998) 237-309
16. Asher, N., and Lascarides, A.: Logics of Conversation. Cambridge University Press, Cambridge (2003)
17. Higginbotham, J.: The Simplest Hard Problem I Know. In Gärtner, H-M, Beck, S., Eckhardt, R., Musan, R., Stiebels, B., eds.: 40-60 Puzzles for Krifka. ZAS, Berlin (2006) (Accessible from `http://www.zas.gwz-berlin.de/40-60-puzzles-for-krifka/`)

Interpreting Metaphors in a New Semantic Theory of Concept*

Yi Mao[1,2] and Beihai Zhou[1,3]

[1]Center for Logic, Language and Cognition, Peking University
[2]Gemalto Inc., U.S.A.
yi.mao@gemalto.com
[3]Department of Philosophy, Peking University, China
zhoubh@phil.pku.edu.cn

Abstract. We present a formalism of metaphor that is based on a newly developed semantic interpretation of concept in the line of Montague's type theory, and exposit the role that metaphoric expressions play in linking two concepts together. We argue that metaphors have to be understood intensionally. We show that a metaphoric expression states the inclusion relation between the intensions of two concepts in a certain context. A formal intensional semantics that defines the truth condition of metaphor is established. This semantics accounts for a number of important features of metaphors such as intensionality, nonreversibility, partial systematicity and context-sensitivity.

1 Motivation

Human communication relies heavily on metaphors like "Time is money" and "Argument is war." This applies to everyday interactions. We use metaphors to express abstract concepts that we cannot easily express using literal language as in "Love is a journey." Metaphors allow the communication of complex configurations of information in a succinct manner to capture the similar structure of a well-understood entity. For instance, "A cell is a factory." The use of literal language to communicate the same meaning would not only be cumbersome and inefficient but also rather impossible. The ideas communicated via metaphors, say, "Sermons are sleeping pills," are so vivid and rich that they cannot be achieved using literal language. Metaphorical expressions are ubiquitous in our languages. Indeed we live by metaphors, as the title of the well-known book by Lakoff and Johnson ([7]) suggests.

There are predicative metaphors (e.g., (1)-(4)) and noun-phrase metaphors (e.g., (5)-(6)). The predication can express a metaphorical property as in (1) and (2), or a complex metaphorical relation between multiple entities as in (3) and (4). The metaphorical noun-phrases could be a definite description as in (5) or a proper name as in (6). Any theory that provides semantic analysis of

* The authors wish to thank Dr. Bertrand du Castel for his inspiring discussions on metaphors and pointers to important references.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 177–190, 2007.

metaphors should at least cover that much ground, though it is desirable to apply to "whole-sentence" metaphors like (7) as discussed in [1].

(1) Juliet is the sun.
(2) Lawyers are foxes.
(3) My car drinks gasoline.
(4) The doctor hit James with bad news.
(5) The cunning fox is up to his familiar tricks.
(6) Einstein over here just stuck his finger into the electrical socket.
(7) Hark! The sun's first light breaks o'er the mountaintop. (Suppose that it is uttered by Romeo in the context that Juliet appears at her bedroom window.)

Metaphors have several important characteristics: (1) Metaphorical meaning is essentially intensional. Most metaphors under the extensional interpretation are literally false. (2) The topic (the first term) and vehicle (the second term) of metaphors are nonreversible. The roles that they play to constitute a metaphorical meaning are asymmetric. (3) Metaphors demonstrate partially systematic structural mappings between the topic and the vehicle. (4) Metaphorical meaning is context sensitive and demonstrates an open-ended quality. (5) The literal meaning is still active in the nonliteral interpretation. (6) The meaning that was once nonliteral can become literal as time goes by. New metaphors can be created, while old metaphors are taken for granted to the extent that they are not metaphors any more.

Any semantic theory of metaphor should be in accord with these unique features of metaphors. The existing three main models (i.e., the *salience imbalance model*, the *structural mapping model* and the *class inclusion model* discussed in [11]) fall short of covering all of them in one way or the other. In what follows we give detailed explanation of these characterizing features of metaphors.

1.1 The Intensionality of Metaphors

Despite the vital importance of metaphorical use of language, it is yet often thought to be at odds with compositional, truth-conditional approaches to semantics as, after all, most metaphors are literally false. The theories of truth within the classic Tarski tradition of truth definition are mostly extensional. The interpretation of a constant or predicate symbol is referred to the denotation or extension of the constant or the predicate symbol. On this account the interpretation of Example (1) is true if and only if the denotation of the logical constant j (i.e., the individual that is referred by Juliet) is an element of the denotation of the one-place predicate *is-the-sun*. Such interpretations may provide literal meanings but cannot account for the truth condition of metaphors.

Some more sophisticated semantic theories of metaphor like [12] and [13] are still extensional in their essence. Van Genabith rooted his semantics for metaphors in type theory that was first developed by Church and got much evolved and standardized in Montague's work. According to [12], "John is a fox" is interpreted as "John has some of the typical properties of foxes." The truth condition of John's metaphorically being a fox requires the existence of a property P denoting a set of entities that includes both the extension of j and all members of *fox*. The referring to "property" sounds intensional. This approach boils down to use extensions of some shared property in interpretations. The attempt of extensional reduction cannot succeed in the examples like "Airplanes are metal birds." Apparently, the shared property under consideration is the flying capability. However, this intended property does not include the extension of birds because some birds like penguins do not fly. We argue that metaphors are essentially intensional and that any extensional approaches (simple-minded or sophisticated) are not sufficient to account for metaphorical meaning.

Vogel ([13]) let a predicate like *fox* have literal extension that includes commonly understood foxes in some possible worlds, but he extended extensions that may also include John in some others. It is with respect to the latter type of possible worlds, the predicate *fox* has a metaphorical meaning. Though the overall function that determines different extensions at different possible worlds is known as intension, the focus of Vogel's account for metaphorical meaning actually lies in referring to some particular possible world where the extension of a predicate in question is extended to cover some entities in a nonliteral sense. The overall function of intension loses its intended meaning. In the intensional logic, an intension of a term is not merely a mathematical function, but is supposed to be linked to the understanding and comprehension of a concept that is expressed by the term. As the name stands, it is responsible to determine the extension of the term in each possible world. If this is the case, then it would be quite against our intuition to think that the intension of *fox* could categorize John as one of the fox kind under some circumstance. The intension of *fox* defined on possible worlds is intended to capture the idea that the variety of foxes are covered under one and the same concept of fox. Foxes with three legs or with green furs, as abnormal as they may be, are nevertheless still foxes. Saying that John is a fox is very different from saying that a three-legged fox is a fox. They differ in essence, not just by degree. It seems that John could under no circumstance be identified as a fox if John were still a human being in the meantime, unless our concept of fox or that of human being were to be dramatically altered.

We have shown that the inadequacy of theories like [12] and [13] to account for metaphorical meanings stems from the extensional aspect. Though their work was presented in the framework of type theory or possible world semantics, unfortunately, the intensionality was indeed reduced to extensions in one way or another. We advocate that metaphors are intensional in their nature and the involved intensionality is by no means reducible to extensions. The intensional account of metaphor that we will present in this paper reflects this position.

1.2 The Nonreversibility of Metaphors

Most metaphors are not reversible. When a metaphor like "John is a fox" is stated in the reversed order "The fox is John", the original meaning is completely lost. Yet the resulting sentence in most of cases can hardly make any sense. "The fox is John" may make some sense only in a rare situation where John happens to have a prototypical property that could also be attributed to the fox[1]. We agree with Glucksberg et al. in a general sense on their claim ([6]) that metaphors are nonreversible.

The change in meaning between the forward and reverse metaphors are supported by the empirical data. Thomas and Mareschal ([11]) did experiment to have the participants rate the sense of literal and metaphorical comparisons in original and reversed formats. The experimenters judged the forward and reverse metaphors and their paraphrases for how much sense they made. One example used is "The apple is a ball." Stating an apple as a ball is to emphasize its properties of being hard, round and perhaps being thrown. Its reversed form (e.g., "The ball is an apple") may still make sense under certain paraphrase; the new meaning is nevertheless very different from the original one. The effect of comparing the ball to an apple is to exaggerate the softness, irregularity and edibility of the ball, reducing its size, its roundness and its likelihood of being thrown.

The reason that metaphors are nonreversible is that the topic and vehicle make very different contributions to the metaphor's meaning. A metaphor highlights some properties of which the vehicle is a prototypical member. The metaphorical comparisons are used to attribute the salient properties of the vehicle to the topic which may share these properties. The existing features of the topic are exaggerated via metaphor. If there is no preexisting similarity between the topic and vehicle, then the acceptance of a metaphor can result in the creation of similarities by attributing some notable properties of the vehicle to the topic.

The property-matching models such as the salience imbalance model and the structural mapping model cannot account for nonreversibility.

1.3 The Partial Systematicity of Metaphors

Lakoff and Johnson enumerated many examples in [7] to illustrate that the metaphorical expressions in our language are tied to metaphorical concepts in a systematic way. In the metaphor of "Argument is war", the expressions from the vocabulary of war form a systematic way of talking about the battling aspects of arguing. It is not just one property that is shared by the vehicle and the topic but rather a portion of the conceptual network of the vehicle that can be systematically applied to the topic. Because of the systematic structural mapping

[1] Dr. Bertrand du Castel gave interesting examples to show that "John is a fox" could be reversed and the meaning of the reversed version is close to the original one. Here are his examples:

(A) There are a lot of foxes in this company. Certainly, John is a fox, and so are you.

(B) There is got to be a fox in this company, otherwise they would not have gotten the authorization. I think that the fox is John, or perhaps you.

We think that these examples should be read as someone x in this company is a fox, which is a metaphor and is not reversible. Referring to this metaphor, the seemingly reversibility is from the identity statement "John = x", which means the same as "x = John".

between the two terms, the communication via a metaphor has the advantage of being compact. The vehicle embodies a covariance of features that, so long as the topic can key into them, may be transferred to the topic as a whole.

On the other hand, while some aspects are highlighted in a metaphor, some others are suppressed or hidden. Comparing argument to battle overlooks the cooperative aspects of arguing. Thus the systematicity revealed by metaphors is always partial.

The class inclusion model, understanding metaphors as categorical assertions, does not have a good story to explain the systematicity of metaphors.

1.4 The Context-Sensitivity of Metaphors

Metaphor is a crucially context-dependent linguistic phenomenon. In the context of different sentences, the same word could be used metaphorically to express many, very different meanings. When Romeo uttered "Juliet is the sun," his words treated Juliet's virtues as like the natural and original sun's light in brightening and warming his world. During the cultural revolution, Chinese people sang the praises of Chairman Mao in a song as being the sun of the country. The metaphoric content of "Chairman Mao is the sun" in this context is fixed as being worthy of worship, nurturing of one's soul, fairer than all others, and so on. Apparently, "the sun" in "Chairman Mao is the sun" conveys a very different metaphorical meaning as it is in "Juliet is the sun."

Moreover, even the same metaphor can mean something dramatically different when it is uttered by different speakers on different occasions. Camp ([1]) had a detailed discussion on this phenomenon. The same metaphorical comparison of Juliet to the sun, when uttered by Romeo, or Romeo's rival Paris, or Romeo's friend Benvolio, could have distinct or even incompatible meanings.

Metaphorical meaning is context sensitive and demonstrates an open-ended quality. The more candidate properties available from the vehicle, the more potential metaphorical interpretations would be. Although pragmatic factors are involved in determining the intended metaphorical meaning, the semantics should provide all possible meanings from which the intended one is to be chosen. It is in this direction that we pursue to develop a semantics of metaphor that will assign interpretations to metaphors in a uniform manner and in the meantime interpretations are parameterized by context, which represents the pragmatic factors.

1.5 The Coexistence of Literal and Nonliteral Meanings

A metaphorical expression has both literal and nonliteral meanings in one expression. This is a special property that a normal expression does not have. Ohkura ([9]) showed two reasons that the metaphorical meaning is not reducible to the literal meaning: (1) Paraphrasing of a metaphorical expression cannot present all the meanings that the original metaphorical expression has; and (2) The metaphorical meaning cannot be inferred from logical contradiction when we understand the expression with its literal meaning. The reduction in the other direction is also not possible. When the metaphorical meaning is comprehended, the literal meaning is still active. We think it is right to say that metaphorical expressions have both literal meaning and nonliteral meaning. They are interactive, but none of them can substitute for the other one. The metaphorical meaning demands a semantics to account for.

1.6 The Death of Metaphors

Once being metaphorical, an expression may eventually "freeze" the original metaphorical meaning as part of its literal meaning via its frequent use. This phenomena is known as the death of metaphor. The boundary between the literal meaning and nonliteral meaning blurs in dead metaphors. There are many such metaphors in our language that we are so used to them and often do not count them metaphors.

For example, how many people actually realize that the expression "is close to" in "Chanukah is close to Christmas" is used metaphorical, as Lakoff and Johnson ([7], 266) pointed out? The expression "is close to" is initially a spatial metaphor for time, as the relationship between the times of the two holidays is captured by the correlated experience in location. However, as time "goes by", the term "close to" applies equally well to the domain of time as it does to the domain of space. It does not sound metaphorical any more after it has been used so often and for so long in the temporal domain. The fact that it is applicable to describe a temporal relation has become part of its normal meaning.

While old metaphors die, new metaphors are created everyday in poetic, political, religious and, even scientific writings. Theories based on the similarity between pre–existing properties have difficulty to account for the emerging of new metaphors. A semantic theory of metaphor that successfully handles both the death of metaphors and creation of metaphors is called for.

We shall present a model-theoretical account for metaphorical meaning based on a newly developed semantic theory of concept. The interpretation of metaphor that we propose heavily relies on our understanding of how a concept is structured. Before we can outline the semantics of metaphor, we need to introduce the basics of our semantic structure of concept to set the ground for our current task. It is to the new semantic theory of concept we now turn.

2 Semantic Interpretation of Concept

There has been a long tradition to divide the use of a term into its extensional meaning and intensional meaning. This bipartition is also discussed in other vocabulary like sense and denotation, and concept and object. It has been commonly agreed that the extension of a term like a common noun is meant to be a number of individuals of which the term may predicate, while the intension of a term signifies a group of attributes. The objects that fall in the extension of a term possess attributes contained in the intension of the term. The words "intension," "sense" and "concept" are used interchangeably, as opposed to objects that they denote.

We argued in Zhou and Mao ([14]) that intension, sense and concept are all different notions and they are related to each other as follows: (a) Sense determines extensions; (b) An intension is a set of senses; (c) Concept determines intensions. Extension, sense, intension and concept together consist of four semantic layers of a linguistic term. Sense and concept are modelled as functions defined on possible worlds. The formalization of four-layer semantic entities was expressed in the framework of Montague's type theory ([8]), and linked back to Frege's theories on sense, denotation ([5]) and concept ([4]) and to Church's type theory of concepts ([2], [3]).

We followed the possible world approach used in the intensional semantics and viewed that meanings are construed as certain functions from possible worlds to various entities. Possible worlds in some sense may be understood as contexts of utterances. We will use possible world and context interchangeably. Montague's type theory provides us the needed semantic entities of four different types which correspond to four different layers of interpretations that a term could have. In Montague's type theory, there are three distinct symbols – e, t and s, and two primitive types – e (for entities) and t (for truth values). A functional type $\langle \sigma, \tau \rangle$ can be composed from σ and τ if σ and τ are types, and an intensional type $\langle s, \tau \rangle$ can be constructed from type τ. The relation between extension and intension is captured by the function application: Intensions are functions from possible worlds to entities of the appropriate sorts, and the extension of an intension in a world is the result of applying the function to the world in question.

Referring to Montague's type theory, we claimed that four semantic entities on different levels that a term could mean are of four different types. Namely, extension of a term is of type $\langle e, t \rangle$, sense is of type $\langle s, \langle e, t \rangle \rangle$, intension is of type $\langle \langle s, \langle e, t \rangle \rangle, t \rangle$, and concept is of type $\langle s, \langle \langle s, \langle e, t \rangle \rangle, t \rangle \rangle$. Thus, four semantic layers of a term are distinguished in connection with the formal types. It is quite familiar to researchers to talk about the denotation and sense of a term as of types $\langle e, t \rangle$ and $\langle s, \langle e, t \rangle \rangle$. As intension and concept have long be confused with each other and with sense as well, it is the novel part of our semantic theory of concept in claiming that (i) an intension is a set of senses; and (ii) concepts are functions from possible worlds to intensions in a similar way as senses being functions from possible worlds to extensions.

An intension being a set of senses covers associated properties of a term in a certain context. The commonly understood property is more or less on the level of sense in our four-layer picture. For example, {Having wings, Capable of flying, Laying eggs, ... } is an intension of bird. The sense of a term is solely responsible for determining the denotations of the term in all contexts, while an intension of a term is not directly tied up to the denotations of the term. For this reason, we insisted that an intension is not just a compound sense.

Like senses, concepts are functions parameterized by context. Unlike senses, concepts take intensions as their values. When the concept of a term is considered under a fixed context (which is often the case), concept and its intension in this particular context are meant to be the same. This explains why they are often confused with each other. Concepts being functions become apparent with respect to multiple contexts. This can be demonstrated in the metaphorical use of expressions and supported by the research in systematic polysemy. With a better understanding of the relationship between metaphors and construction of concepts, our semantic theory of concept also finds itself an application area of interpreting metaphorical meaning.

3 Concept and Metaphor

The study of polysemy shows that words not only have literal meanings in a concrete domain but also have systematically related meanings in abstract domains. Lakoff and Johnson ([7]) have given us a dozen of such examples: up, down, rise, fall, high, low, hit bottom, and so on. The polysemous word "rise" is used, for example, to mean both increase in elevation and increase in quantity. The phrase "close to" applicable to both time and space domains is another example. We can add a lot more like "home run" and "chicken" to the list. "Home run" is originally a baseball slang referring to a hit that allows the batter to make a complete circuit of the diamond and score a run. It is also applicable to describe a job well done, a highly successful achievement, or a doubling of one's profits. Chicken is a popular synonym for "coward."

Viewing the multi-meanings of polysemous words as intensions under different contexts, the phenomena of polysemy fits well in the picture of concepts being functions over contexts. The systematicity is in accord with our claim that these meanings are not grouped under the same word by coincidence, but rather they are governed by the same concept denoted by the word. The seemingly different but related meanings (i.e., intensions, in our terminology) of a word are the function values that are obtained by applying the same concept in various domains (i.e., contexts). Or, looking it from a reversed direction, multiple meanings of a word jointly construct the concept that they share.

The shared concept is what connects intensions in various contexts under the same expression. When an old concept is applied to a new domain like "close to" and "chicken", the meaning is always metaphorical at the beginning. When the domain is no longer considered new, the concept as a function is extended to cover the context that was once new and assigns a stable intension to it. To that point, the new aspect of the meaning has become a normal part of the concept and then the metaphorical flavor varnishes. What has been said explains the death of metaphor, but we do not predict that every metaphor will die at a later time. The "time is money" metaphor has come a long way but is still alive.

The intension of a concept in a context is open-ended in a long run and can be enriched via metaphors. A metaphor bridges two concepts – the concept expressed by the topic and the concept expressed by the vehicle. Some typical properties (i.e., senses in the four-layer semantic structure) of vehicle are highlighted and accommodated as part of the intension of the topic in the context of the utterance of the metaphor. This could results in expanding the intension of the topic concept, or amplifying and exaggerating the similar properties of topic that were previously included in its intension. Based on this observation, we propose that a metaphor is true if and only if, with respect to the context of utterance, the intension of the vehicle concept is included in the intension of the topic concept. This is a description of truth condition for metaphor and does not capture the expansion of intension or amplifying some elements in an intension that may be involved in comprehending metaphors. We formalize this truth condition in the next section.

4 A Formalism of Metaphors

The formalism of metaphors we will present here is developed from our very basic claim: metaphors have to be treated intensionally rather than extensionally. We second a position held by Lakoff and Johnson that metaphors are conceptual in nature rather than a matter of "mere language" ([7], 159). The formalism reflects our view that a metaphoric expression states the inclusion relation between the intentions of two concepts in a certain context. One of these two concepts serves as topic and the other as vehicle.

In some metaphorical statements where the topic term is about an individual like "John" or the vehicle term is a relational predicate like "drink," the topic concept is not explicitly expressed by some word in the sentence, and it is not obvious what the topic concept is. For example, the topic concept in metaphor "John is a fox," under our analysis, is "the property of being John." The "consume" relation that matters to my car and gasoline is the topic concept in "My car drinks gasoline," and it is this relational concept that is metaphorically compared to the concept of "drink." For these metaphors, a transform step is needed to get their logical form from the original sentences. This step abstracts from an object to the property of being this object or from objects to a relation holding among these objects. To capture this abstraction, we introduce the λ operator in the formal language to construct λ expressions that are needed for the logical representation of some metaphors.

Many metaphors like "Time is money" and "Argument is war" are in the form of "A is B". If we consider A and B as the topic and vehicle concepts that may be constructed via λ operator, then this is indeed a general form that covers all atomic (as opposed to mixing) metaphors. "Is" in "A is B" is a connective that links two terms to form a metaphor. We introduce Mbe in the language to denote it. The logical form of a metaphor can thus be written as $Mbe(A, B)$, where A and B are predicates or λ expressions with the same arity. With these explanations in mind, we are now ready to lay out a formal language and its corresponding semantics.

4.1 Syntax

The language $\mathfrak{L}_M$ that we use to characterize metaphors contains the following symbols: (1) a denumerable set of individual variables $Var = \{v_0, v_1, ...\}$; (2) a denumerable set of individual constants $C = \{c_0, c_1, ...\}$; (3) a denumerable set of n-ary ($n \geq 0$) predicate variables $Pred = \{P_0^n, P_1^n, ...\}$; (4) an abstraction operator λ; (5) a term connective Mbe; (6) two sentential connectives: $\neg$ and $\rightarrow$; (7) a universal quantifier $\forall$; and (8) a separator ','.

Hereafter, lower-case letters a, b, x, y etc. are used for variables or constants, and capital letters P, Q, R etc. for predicates. A λ expression in the form of $\lambda P(Pa_0a_1...a_n)$ are treated, for the simplicity, as a compound n-ary predicate, instead of a set of n-ary predicates as it normally denotes in literature. Other connectives $\vee$, $\wedge$, and $\leftrightarrow$, and quantifier $\exists$ can be customarily introduced. In addition to the formulas that are defined as usual, $Mbe(A, B)$ is also a formula, where A and B are predicates or λ expressions. The intended reading for $Mbe(A, B)$ is "A metaphorically is B."

In $\mathfrak{L}_M$, we can write the logical representations for metaphors listed in Section 1 as follows:

(1') $Mbe(\lambda P(Pj), \lambda Q(Qs))$
(2') $Mbe(L, F)$
(3') $Mbe(\lambda C(Ccg), D)$
(4') $Mbe(\lambda T(Tdjn), H)$
(5') $\exists x(Mbe(\lambda P(Px), F) \wedge Uxt)$
(6') $\exists x(Mbe(\lambda P(Px), \lambda Q(Qe)) \wedge Sxfs)$
(7') $\exists x\exists y(Mbe(\lambda P(Px), \lambda Q(Qs)) \wedge Mbe(\lambda R(Ry), M) \wedge Axy)$

(1')-(4') are what we called atomic metaphors. They metaphorically link together two concepts, which may or may not directly expressed by some words in the sentences. The use of λ expressions shows the abstraction getting concepts involved in metaphors. (5')-(7') have embedded metaphors. (5') talks about someone who is a cunning fox, which is a metaphor, and he is up to his familiar tricks. A similar analysis applies to (6'). Namely, someone here is an Einstein in the metaphorical sense, and he just stuck his finger into the electrical socket. (7') is more complex as it has two embedded metaphors: someone is the sun and something is the mountaintop. The set up of language $\mathfrak{L}_M$ is sufficient to manage examples with this complexity. This is not achieved in other theories of metaphor like [10].

The "metaphorically be" (i.e., Mbe) connective we introduced at the level of logical form is different from the "metaphoricity" operator $Mthat$ that Stern postulated in [10]. Stern made $Mthat$ a unary operator and applied it only to the vehicle of the metaphor, but we believe that metaphor is an interactive phenomena between the topic and the vehicle and hence use a binary connective to mark a metaphor. In consequence, our approach does not encounter the difficult of syntactic ambiguity as where to append Mbe in the sentence — a problem that Stern's $Mthat$ faces.

4.2 Semantics

Corresponding to our thesis that a term can be interpreted on four levels (i.e., denotation/extension, sense, intension, and concept), we will first of all define *sense space* and *concept space*, from a given set of possible worlds and a give domain of all possible objects. Then a term will get assigned a sense and a concept, and its extensions and intensions in various possible worlds (i.e., contexts) can be derived as values obtained via the applications of its assigned sense function and concept function. We are to provide a mathematical model in which the truth condition of metaphors can be described. This model is not intended for the comprehension of concept, nor should it be viewed that we claim concept is an independent objective thing.

Definition 1. *(Sense space) Given a set W of possible worlds, a set D of objects, and any $n > 0$, let* $\mathsf{SEN}(W, D^n) = \{sen \mid sen \text{ is a function from } W \text{ to } \wp(D^n)\}$.

$\mathsf{SEN}(W, D^n)$ contains the senses that can be assigned to any n-ary predicates or λ expressions. The sense *sen* of an n-ary predicate determines its extension on every possible world in such a way that it selects a subset of n-tuples of objects from the domain D^n (i.e., for each $w \in W$, $sen(w) \subseteq D^n$) that satisfy the property or relation depicted by the predicate.

Definition 2. *(Concept space) Given a set W of possible worlds, a set D of objects, and any $n > 0$, let* $\mathsf{CON}(W, D^n) = \{con \mid con \text{ is a function from } W \text{ to } \wp(\mathsf{SEN}(W, D^n)\}$.

$\mathsf{CON}(W, D^n)$ contains all concepts that may be expressed by some n-ary predicates or λ expressions. The concept *con* of an n-ary predicate determines, in every possible world, its intension whose elements are from $\mathsf{SEN}(W, D^n)$.

Definition 3. *An $\mathfrak{L}_M$-model $\mathfrak{M}$ is a five-tuple $\langle W, D, \sigma, \eta_s, \eta_c\rangle$, where*

1. *W is a non-empty set of possible worlds;*
2. *D is a non-empty set of objects;*
3. *σ is a function that assigns objects to varibles (i.e., $\sigma(v) \in D$, for any $v \in Var$);*
4. *η_s is a sense assignment function such that*
   (i) *$\eta_s(c) \in D$, for any $c \in C$;*
   (ii) *$\eta_s(P) \in \mathsf{SEN}(W, D^n)$, for any n-ary $P \in Pred$;*
   (iii) *$\eta_s(\lambda P(Pa_1...a_n)) \in \mathsf{SEN}(W, D^n)$ such that $\langle a_1^{\mathfrak{M}}, ..., a_n^{\mathfrak{M}}\rangle \in \eta_s(P)(w)$ in all $w \in W$ (if $a_1 \in Var$, $a_1^{\mathfrak{M}} = \sigma(a_1)$; else if $a_1 \in C$, $a_1^{\mathfrak{M}} = \eta_s(a_1)$);*
5. *η_c is a concept assignment function such that*
   (i) *$\eta_c(c) = \eta_s(c) \in D$, for any $c \in C$;*
   (ii) *$\eta_c(P) \in \mathsf{CON}(W, D^n)$ satisfying $\eta_s(P) \in \eta_c(P)(w)$ at some $w \in W$, for any n-ary predicate or λ expression P.*

In the above definition, functions η_s and η_c assign trivialized senses and concepts (i.e., extensions) to constant symbols. The constraint placed on η_s for the case of λ expression is to guarantee that the property or relation assigned to the λ expression applies to the object(s) from which it is abstracted. The constraint for η_c is to ensure that the concept and sense assigned to the same predicate or λ expression will not be completely irrelated to each other. Some other constraints are possible to be added to the model, but we are not concerned about them for the moment.

Definition 4. *Given an $\mathfrak{L}_M$-model $\mathfrak{M}$, a possible world w and a formula α, we recursively define that α is true at w in $\mathfrak{M}$ (denoted as $\mathfrak{M} \vDash_w \alpha$) in the clauses below.*

1. *If $\alpha = Pa_1...a_n$, $\mathfrak{M} \vDash_w \alpha$ iff $\langle a_1^{\mathfrak{M}}, ..., a_n^{\mathfrak{M}} \rangle \in \eta_s(P)(w)$.*
2. *If $\alpha = Mbe(A, B)$, $\mathfrak{M} \vDash_w \alpha$ iff $\eta_c(B)(w) \subseteq \eta_c(A)(w)$.*
3. *If $\alpha = \neg\beta$, $\mathfrak{M} \vDash_w \alpha$ iff $\mathfrak{M} \nvDash_w \beta$.*
4. *If $\alpha = \beta \to \gamma$, $\mathfrak{M} \vDash_w \alpha$ iff $\mathfrak{M} \nvDash_w \beta$ or $\mathfrak{M} \vDash_w \gamma$.*
5. *If $\alpha = \forall x\beta$, $\mathfrak{M} \vDash_w \alpha$ iff $\mathfrak{M}(d/x) \vDash_w \beta$ for all $d \in D$.*

All clauses except the second one in the above definition are classical truth conditions. The second clause specifies the condition under which a metaphor is true. The condition states that a metaphor like "John is a fox" is true in a context w where all properties (e.g., cunning, sly, clever, etc.) of foxes particularly considered as relevant in w are also properties of being John. In other words, some typical features of foxes, as far as the metaphor is considered, is part of John's character.

The use of the context w here is to get an appropriate intension that contains only relevant properties for the metaphor and excludes irrelevant properties like having a tail or stealing chicken. In contrast to Stern's *Mthat* operator that contributes the set of properties which happen to be m-associated with the vehicle term in the context of utterance, the selection of relevant properties for a context in our semantics is governed by the assigned vehicle concept. A wide variety of factors (social, cultural or even personal experience) can shape a concept to have a different intension (i.e., a set of senses, and loosely, a set of properties) in a certain context. Our semantic model is to determine the contents of metaphorical utterances, given information about how involved concepts are constructed.

The truth condition of metaphor we provide matches our intuitive judgement for the truth of metaphors. If one knows John very well that John is an honest person, he would then deny that the metaphor "John is a fox." On the other hand, suppose that one does not know Mary and is told that Mary is a fox. If he takes this metaphor to be true, then he would attribute the fox-like properties to her and be careful in doing business with her.

In our model, while the metaphorical meaning is interpreted via η_c, the literal meaning depends on η_s. Literally, it is false that John is a fox because $\eta_s(j) \notin \eta_s(F)(w)$ where w is the context of utterance. Thus, the truth of a metaphor does not conflict with the falsity of its literal meaning. Both of them are accounted for in the same model.

5 Conclusions

Looking back to the list of unique features of metaphor that motivated us to develop a comprehensive theory of metaphor, we think that our formalism of metaphor and its semantic interpretation have met this challenge.

The core idea of our approach is to capture the intensionality of metaphor, and everything else is built around this idea. We insist that metaphors express relations between two concepts and our semantics reflects this fundamental view. The intensions in our model are not reducible to extensions. The truth of metaphors is determined by the inclusion relation between two intensions. The theory of metaphor that we have presented is nothing but intensional.

The inclusion relation used in the truth condition for metaphor is an anti-symmetric relation. When two intensions of two concepts in a context are not identical, which is the case for most metaphors used in the natural languages, the inclusion relation is not reversible. So are metaphors. On the other hand, in two different contexts w_1 and w_2, it could happen that the intension of one concept is included in that of another at w_1, but it is the other way around at w_2. This explains the apple-ball example. Both "The apple is a ball" and "The ball is an apple" can be true, but they are true from the different perspectives (i.e. contexts) where the intensions of apple and ball have changed to contain a different set of properties.

An intension for us is not just one sense (which can roughly be understood as property) but a set of senses. The truth of metaphor implies that a set of properties of vehicle should be the properties of topic as well. This shows the systematicity. However, this systematic feature will always be partial because an intension of a concept in a certain context does not contain all possible properties that the same concept may have in some other context.

The interpretation of metaphor we gave is context-sensitive. Our notion of concept is parameterized by context. The intension of a concept varies from one context to another. We believe a metaphor makes a claim about intensions of two concepts in a certain context. The truth of a metaphor depends on the context in which it is evaluated. In the context of Romeo's utterance, it is true that Juliet is the sun because she has the sun's properties of being bright and warm. In the context of Paris's utterance, it is also true that Juliet is the sun because she has the sun's other properties like the center of the surroundings.

Our semantics provides truth conditions to both literal meaning and nonliteral meaning of a metaphorical expression. The literal meaning is interpreted existentially, whereas the nonliteral meaning is taken from an intensional angle. Although the literal meaning and nonliteral meaning of metaphorical sentences are distinct and modelled on a different layer, they nevertheless use the same formal apparatus.

Via the four-layer semantic model for concept and its relation to metaphors, we have shown that the death of metaphor occurs when a concept has evolved and gained conventional meaning in a domain.

Based on our study of metaphor, a point we would like to deliver is that metaphoricity is not outside the remit of natural language semantics. Rather, explanations of certain aspects of metaphor are integral to the theory of meaning of any sentence in a natural language.

References

1. Camp, Elisabeth: *Josef Stern, Metaphor in Context (Cambridge, MA: MIT Press, 2000)*, **Noûs**, Vol. 39, No. 4. (December 2005), pp. 715-731.
2. Church, Alonzo: *Carnap's introduction to semantics*, **The Philosophical Review 1943 Vol. LII**, pp298-304.
3. Church, Alonzo: *A formulation of the logic of sense and denotation*, in Paul Henle, et al. (eds.), **Structure, Method and Meaning**, Liberal Arts Press, New York, 1951.
4. Frege, Gottlob: *Function and Concept*, in **Frege, Collected Papers on Mathematics, Logic and Philosophy** (ed. McGuinness), pp. 137-56.
5. Frege, Gottlob: *On Sense and Reference*, in JSTOR: Phil. Review (May 1948).
6. Glucksberg, S., McGlone, M.S., & manfredi, D.: Property attribution in metaphor comprehension. Journal of Memory and Language, 36 (1997), pp. 50-67.
7. Lakoff, George and Johnson, Mark: **Metaphors We Live By**, The University of Chicago Press, Chicago and London, 1980.
8. Montague, Richard: *Universal Grammar*, in **Formal Philosophy**, (ed. Richmond H. Thomason), New Haven and London, Yale University Press, 1974.
9. Ohkura, Chiaki: *The Semantics of Metaphor in the Game Theoretic Semantics with at Least Two Coordination Equilibria*, ACL 2003 Workshop on The Lexicon and Figurative Language, URL = http://acl.ldc.upenn.edu/W/W03/W03-1408.pdf.
10. Stern, Josef: **Metaphor in Context**, Cambridge, Mass., 2000.
11. Thomas, Michael S.C. and Mareschal, Denis: *Metaphors as Categorization: A Connectionist Implementation*, **Metaphor and Symbol**, Vol. 16, No. 1&2 (2001), pp. 5-27.
12. Van Genabith, Josef: *Metaphors, Logic, and Type Theory*, **Metaphor and Symbol**, Vol. 16, No. 1&2 (2001), pp. 43-57.
13. Vogel, Carl: *Dynamic Semantics for Metaphor*, **Metaphor and Symbol**, Vol. 16, No. 1&2 (2001), pp. 59-74.
14. Zhou, Beihai and Mao, Yi: *Four Semantic Layers of a Common Noun*, in review.

Covert Emotive Modality Is a Monster

Sumiyo Nishiguchi

Stony Brook University, Department of Linguistics
Stony Brook, NY 11794-4376
Osaka University, Graduate School of Letters, Department of English Linguistics
1-5 Machikaneyama, Toyonaka 560-0043, Japan
snishigu@ic.sunysb.edu
http://homepage3.nifty.com/sumiyo_nishiguchi/

Abstract. It has been argued that in some languages, attitude verbs shift the reference of indexicals in the embedded clauses [1, 2, 3]. I argue that not only overt attitude predicates but also implicit emotive modality is a context-shifting operator that changes the context parameters. I base my argument on the following three mono-clausal constructions: (i) fake past (ii) fake present, and (iii) discourse initial sentence focus *also/even/too*. Surprise licenses non-past interpretation of the past tense [4] with negative presupposition. Alternation between past and present directs readers to re-experience narratives [5]. A speaker's emotion or sentiment licenses wide-scope *also/too* without explicit antecedents. Finally, covert modal is grammatically represented as a determiner-like element that takes negative presupposition in the restrictor and overt predicate in the nuclear scope [6, 7, 8, 9].

Keywords: monster, attitude, fake past, fake present, wide scope *also*.

1 Shifted Context in Unembedded Sentences

I discuss the following three types of mono-clausal constructions:

Fake Past

(1) a. Oh, it was here (all along).
b. A, koko-ni {at-ta/#a-ru}. (Japanese)
Oh here-LOC be-PAST/be-PRES
'Oh, it was here'
c. Mintian you-le wanyan. (Mandarin)
tomorrow have-PERF party
'I had a party tomorrow'

Fake Present in Narratives

(2) Picchaa nage-ta. Ut-ta. Ichiro hashi-ru. Oshii. Auto. (Japanese)
Pitcher throw-PAST hit-PAST Ichiro run-**pres** sorry out
'The pitcher threw a ball. (Ichiro) hit it. Ichiro *runs*. Oh, no. He is out.'

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 191–204, 2007.

Discourse Initial *Also/Too*

(3) a. Yo-**MO** huke-ta. (Japanese)
night-also pass-PAST
'It grew late'
b. It's nice here, **too**.

The existence of monsters that shift the context parameters has been argued in the literature [10, 1, 2, 3]. Typically, indexical shift has been observed in embedded contexts. The data discussed in this paper indicate that even implicit speaker attitude shifts the context parameters. In each of the three cases, surprise, empathy, and sentimentality affect temporal interpretations and satisfy presuppositions.

(4) $\text{MODAL}_{emotive}(\llbracket\phi\rrbracket^{<<t_c,w_c,a_c>,<t_i,w_i>>}) = \llbracket\phi\rrbracket^{<<t_i,w_i,a_c>,<t_i,w_i>>}$
(t = time, w = world, a = speaker, c = context, i = index)

Speaker's emotion switches the context into worlds in which the presupposition of discourse initial *mo/too* is satisfied. Covert attitude predicates such as *I regret*, *I'm sorry*, *I'm glad that*, *happily*, or *be touched with* are monstrous functions that satisfy the presuppositions.

Modality is grammatically represented by an abstract determiner-like element that takes negative presupposition as the restrictor and the overt unaccusative VP in the nuclear scope [cf. 11, 7, 8]. Further, modals resemble psych-verbs such as *surprise* or *affect* in the argument structure. Moreover, emotives take the speaker to be the experiencer and the event to be the theme.

Sections 2 and 3 examine mono-clausal fake past[1]and fake present sentences and indicate that modality distorts temporal interpretation. Section 4 shows that emotive modality accommodates the presupposition of wide-scope *also/too* used unexpectedly in Japanese, Korean, Chinese, and English.

2 Fake Past and Emotive Modality

Past tense in a simple sentence can be used without reference to past time in the context of expressing surprise at finding something that was lost, recalling something or at seeing the fulfillment of expectation [13, 14, 4, 15, 16, 17, 18, 19, 20]. [21] suggests that the Japanese past tense morpheme is ambiguous between non-past and past in view of the non-past interpretation in relative clauses.

I claim that mono-clausal "fake" past sentences are embedded under covert emotives that shift the context parameters. [1, 2] propose that attitude verbs shift references of indexicals in subordinate clauses. The mono-clausal fake past sentences indicate that, not only overt attitude verbs but also covert superordinate attitude predicates change temporal interpretations.

The necessary conditions for fake past interpretation are as follows: (i) speaker's surprise due to negative presupposition and (ii) the implicit factive emotive, *be glad*. Therefore, I argue that a speaker's speculative and bouletic modality [6] is a context-shifting operator [10].

[1] The terminology "fake" is taken from [12].

2.1 Data

The past tense in the following examples does not refer to the past.

Stative predicates. *Copular verb when finding an object*:

(5) a. Oh, the book was here (all along).
b. Chek-i yogi iss-ot-ne. (Korean)
book-NOM here be-PAST-EXC
'Oh, it was here'

When recalling a future schedule:

(6) a. There was a party tomorrow.[2]
b. Asu-wa Maria-no tanjobi-dat-ta. (Japanese)
tomorrow-TOP Maria-GEN birthday-be-PAST
'Tomorrow is Maria's birthday'
c. I had a meeting next Monday.

When recalling forgotten information:

(7) a. Where did you live, again?
b. Osumai-wa dochira-deshi-ta-ka. (Japanese)
residence-TOP where-HON-PAST-Q
'Where did you live?'
c. What was your name?

[4, 108]

d. Onamae-wa nan-deshi-tak-ke. (Japanese)
name-TOP what-HON-PAST-Q
What was your name?

Adjectives:

(8) a. Yo-kat-ta. (Japanese)
good-be-PAST
'Thank goodness' (when a lost wallet was returned with money)
b. (While I expected it to be blue) Kiiro-kat-ta. (Japanese)
yellow-be-PAST
'It is yellow'

(9) Shimat-ta. (Japanese)
close-PAST
'Oh, no'

[2] In the antecedent of mismatched counterfactual conditionals, fake past tense morphemes cooccur with future adverbial e.g., *tomorrow*, *next week* [22, 9]:
(i) If it rained tomorrow, I would go shopping.
(ii) If Lucy played with Charlie tomorrow, she would make him happy.

[9, 145]

Eventive predicates

(10) a. Basu-ga ki-ta. (Japanese)
bus-NOM come-PAST
'The bus is coming'

b. Ya ushla. (Russian) [3]
I go-PAST
'I am leaving'

c. A, warat-ta. (Japanese)
oh smile-PAST
'Oh, (the baby) is smiling'

Eventive verbs express less surprise or unexpectancy than stative predicates. Stative copular verbs are typical of non-past interpretations.

Unaccusativity. Stage-level [23] copular verbs are typical of fake past sentences. Further, the above data indicates that the verb classes used for the fake past construction tend to be unaccusative verbs [24, 25]. This past tense morpheme with non-past interpretation in relative clauses is known to take unaccusative predicates [26, 21, 27]. Moreover, this holds for the mono-clausal fake past construction, as well.

2.2 Fake Past

The past tense morpheme in the above examples does not refer to the actual past. The existence of a book in (7) and the coming of a bus in (10a) refer to the utterance time. The forgotten party is planned in the future in (8a,b).

Such vacuous tense phenomena are reminiscent of the Sequence of Tense phenomena:

(11) John said that Bill was sick.

In (11), the past tense in indirect discourse corresponds to the present tense in direct quotation. As the embedded tense is obligatorily controlled by the matrix tense, the subordinate tense is rendered null.

However, there is no superordinate clause that might manipulate the temporal reference of mono-clausal fake past sentences. Moreover, the non-past interpretation of the past tense is not limited to the sequence of tense languages.

Relevant to this aspect, there has been an argument on the direct referentiality of indexicals such as *I, you, it, now* and *yesterday* [28]. According to [10], these indexicals are rigidly specified once the character of a sentence is applied to the utterance context, before the context is derived. On the other hand, discussions on various languages claim that monstrous functions shift the context (Amharic [2]; Zazaki and Slave [3]; also see Aghem [29]; Navajo [30]).

[3] This example is courtesy Roksolana Mykhaylyk.

Tense is deictic [31], for instance, as the past tense in (12) refers to a definite interval whose identity is broadly distinguishable from the extra-linguistic content.

(12) I didn't turn off the stove.

The fake past sentences indicate that the context fails to provide direct reference to the temporal interpretations. A speaker's implicit attitude is the monstrous function that changes the context parameter.

2.3 "Fake" Operator Shifts Context Parameters

In the framework of the double index system [28], the ordinary past tense morphology shifts the temporal index into a prior time:

(13) Where $c =< w_c, t_c, a_c >, i =< w_i, t_i >$, t_i is prior to t_c,
$[\![Past\phi]\!]^{<t_c,w_c,a_c>,<t_i,w_c>} = 1$ iff there is t_i such that $t_i \prec t_c$ and $[\![\phi]\!] = 1$ at t_i in w_c

However, the fake past tense changes the temporal parameter of the context:

(14) *Fake tense operator*:
Fake: $((c \times i) \to t) \to ((c \times i) \to t)$
Fake $([\![\phi]\!]^{<c,i>}) = 1$ iff $[\![\phi]\!]^{<c^{[t_i/t_c]},i>} = 1$

2.4 Negative Presupposition

The fake past sentences presuppose negative antecedent as in (15) [19, 32]:

(15) (Nai-to omot-te-i-ta-ra,) at-ta.
NEG-COMP think-CON-be-PAST-then be-PAST
'(To my surprise,) it is here.'

Without the surprise that is caused by negative presupposition, the fake past interpretation cannot be obtained. The above examples would only refer to the past state or events.

(16) a. # At-ta. (without surprise, with fake past reading)
exist-PAST
'It was here'
b. At-ta. (without surprise, with real past reading)
exist-PAST
'It was here'

The negative implicature, i.e., the conventional implicatures [33], can take the covert forms of antecedent of conditionals (17a) or the main clause *be surprised* embedding the fake past sentences (17b):

(17) a. While expecting $\sim\phi$, it is ϕ
b. I am surprised (=surprisingly) that it is ϕ

The fake past tense morphology is the characteristic of counterfactual conditional [34]. The morphological behavior of fake past corresponds to the subjunctive behavior in "contrary-to-fact" conditionals [35].

2.5 Bouletic and Speculative Modality

[14] suggests that subjectivity alters the tense, namely, reporting as one's own experience or expressing as direct cognition affects the tense. To be more precise, the combination of speculative and bouletic modality affects the temporal interpretation [cf. 6].

Negative presupposition is the the result of an evaluation by speculative modality. According to the past speculation, circumstances appeared to be $\sim\phi$. The speaker's speculation changes at the time of utterance.

(18) *Exp*: expectation function based on the available data, a: speaker,
$Exp_a(w_i)(t_i)\,[\![\sim\phi]\!]^{w_i,t_i} \wedge Exp_a(w_c)(t_c)\,[\![\phi]\!]^{w_c,t_c} \wedge Know_a(w_c)(tc)\,[\![\phi]\!]^{w_c,t_c}$

The speaker might be positively presupposed as well, e.g., while waiting for a bus at a bus stop:

(19) $Exp_a(w_i)(t_i)\,[\![\phi]\!]^{w_i,t_i} \wedge Exp_a(w_c)(t_c)\,[\![\phi]\!]^{w_c,t_c} \wedge Know_a(w_c)(t_c)\,[\![\phi]\!]^{w_c,t_c}$

The speaker is not only negatively or positively presupposed but also happy to find an item that she has been looking for:

(20) (Ureshii-koto-ni,	nai-to	omotte-i-ta-ra,)	at-ta.
happy-fact-GOAL	NEG-COMP	think-be-PAST-PRES-then	be-PAST

'(I'm glad that) it is here (surprisingly).'

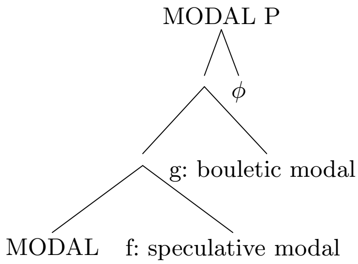
Therefore, in addition to the negative (or positive) presupposition, bouletic modality evaluates fake past sentences.

(21) Bouletic modality (in view of what I want): ϕ worlds are ranked higher than $\neg\phi$ worlds

There are two implicit modals interacting with fake past sentences. Assuming a syntacticized version of Kratzer's theory [6, 36], modal takes two conversational backgrounds, g and h, for the ordering sources:

(22) *Conversational backgrounds*:
f (in view of my speculation): $(c \times s) \rightarrow ((c \times s \rightarrow t) \rightarrow \mathrm{t})$
g (in view of what I want): $(c \times s) \rightarrow ((c \times s \rightarrow t) \rightarrow \mathrm{t})$

$$(D_s = D_w \times D_t)$$



(23) a. For a given strict partial order $<_P$ on worlds, define the selection function max_P that selects the set of $<_P$-best worlds from any set X of worlds:
$\forall$ X $\subseteq$ W: MAX$_P$ (X) {w$\in$ X: $\neg\exists$ w' $\in$ X: w'$<_P$w}
For any world w, time t, conversation grounds f, g, and proposition ϕ:
$[\![fake\_past]\!]^{<c,i>}(\phi)(\mathrm{f})(\mathrm{g}) = 1$ iff $\mathrm{w}_c \notin \bigcap \mathrm{f}(\mathrm{w}_i) \wedge w_c \in \mathrm{MAX}_g\ (\mathrm{w}_c)$, where $t_i \prec t_c$

b. Fake:$(c \times s \rightarrow t) \rightarrow ((c \times s) \rightarrow ((c \times s \rightarrow t) \rightarrow t)) \rightarrow ((c \times s) \rightarrow ((c \times s \rightarrow t) \rightarrow t)) \rightarrow ((c \times s) \rightarrow t)$

2.6 The Phrase Structure of Fake Past

Modal scopes over the proposition as a result of its quantificational force [37, 34, 38]. Modal takes the presupposition as its restrictor, and the assertion in its nuclear scope [7, 8, 11, 39].

(24)	MODAL	[$\lambda i. [\![\sim \phi]\!]$]	[$\lambda i. [\![\phi]\!]$]
	determiner	restrictor	nuclear scope

The quantificational operator MODAL adopts the negative presupposition in the restrictor, and the overt proposition in the nuclear scope [40, 9]. The emotive modal s-selects [41] a negative presupposition and a positive predicate. In a sense, negative presupposition resembles a negative propositional pronoun [cf. 42, 43].

The surprise modal behaves like the psych-verb *surprise*. Surprise c-selects the speaker as the experiencer and the event as the theme [cf. 44]:

(25) SURPRISE(speaker, event)

3 Narrative Fake Present

3.1 Data

In contrast to the *fake past*, there is the *fake present* phenomena. In narratives, the references of the past tense freely alternate between the past and present tenses [45, 14, 5, 4, 46]. While the past tense is the norm, the tense switches between present and past in novels.

(26) Yesterday, I **was** quietly sitting in my favorite hang-out and sipping a beer, as I always do at this time of the day. All of a sudden, a man to my left **turns** to me, **grins** and **says**: ‘Let’s get our of here!’

[45, 134]

Tense alternation is common in Japanese literature:

(27) Kocho-wa usuhige-no aru iro-no kuroi me-no okina
principal-TOP mustache-GEN be color-GEN black eye-GEN big
tanuki-no-yona otoko-de-**aru**. Yani mottaibut-te-i-**ta**
badger-GEN-like man-be-**non-past** terribly pompous-be-PAST
‘The principal **was** a dark complexioned man, with a whiskery mustache and large eyes like a badger. He **was** pompous.’

(Soseki Natsume, *Bocchan*)[4]

(28) Aru haru-no higure-**desu**. To-no miyako Rakuyo-no
some spring-GEN evening-be.**non-past** Tang-GEN capital Luo Yang-GEN
nishi-no mon-no shita-ni, bonyari sora-o aoide-iru, hitori-no
west-GEN gate-GEN under-LOC idly sky-ACC look up-PROG 1-CL-GEN
wakamono-ga ari-mashi-**ta**.
youth-NOM be-HON-**past**
‘It **was** in one spring evening, Under the west gate of Luo Yang, the capital of Tang, there **was** a young man who **was** looking up the sky blankly’
(Ryunosuke Akutagawa, 1920, *Toshishun*[5])

3.2 Context Change by Narrator’s Report Function

Fake present tense invites readers into a depicted world so that they experience the story as though they were present in it [5].

I claim that the proposition is embedded under the NARRATOR SAYS function that shifts the context. It is the empathy of both writer and reader that shifts the context parameter as in (29):

(29) NARRATOR SAYS: $(s \to (s \to t)) \to (s \to t)$
NARRATOR SAYS($[\![\phi]\!]^s$) $= [\![\phi]\!]^{s[w_i/w_c, t_i/t_c, a_i/a_c]}$

According to [45, 134], the narrative present depicts some of the events, occurrences, experiences, and feeling of those who participate in them ‘as if they were present.’ We state that propositions are embedded under the emotive modalities of both a writer and a reader. The narrator attempts to arouse the reader’s empathy by describing past events in such a manner that they appear present [cf. 47]. The writer’s implicit attitudes shift the temporal and locative parameters.

[4] This is quoted from [46], glossed by the author.
[5] Glossed by the author.

4 Discourse Initial *Also/Too*

In the following examples, *too* does not presuppose any antecedents.

(30) a. He is nice. He is a linguist, too.
b. It's nice here, too.[6]

In this section, I discuss wide-scope taking focus markers, namely, Japanese *mo* 'also/even,' Korean *to* 'also/even,' and Chinese *ye* 'also' that appear discourse initially without any antecedent such as *too* in the above examples. To the best of my knowledge, [48, 49] are the only literatures that discuss the wide-scope taking *mo* 'also/even'; thus far, there has not been much theoretical analysis. I argue that a speaker's sentimental attitude licenses such *also/too* and shifts the context parameters in order to plug the presuppositions.

4.1 The Third *Mo* 'also/even'

Mo 'also/even' is a particle attached to noun phrases in Japanese.

(31) Ken-**mo** ki-ta
Ken-also come-PAST
'Ken came, too'

Mo obtains the meaning of 'even' when the NP is focused [50].

(32) $[\text{Ken}]_F$-**mo** ki-ta
Ken-also come-PAST
'Ken came, too'

Japanese grammarians have considered *mo* polysemous [51]: (i) *also* (ii) *even* and (iii) attenuation [48] or admiration [52].

The meaning of the third *mo* is distinct from the other two, *also* and *even.*

(33) Haru-**mo** takenawa-ni nari-mashi-ta
spring-also peak-GOAL become-HON-PAST
'The spring has reached its peak'

[49, 172]

This *mo* does not need any antecedent such as "x reached its peak, *and* spring *also* reached its peak." *Mo*^{3} has its distinct meaning since replacing *mo*^{3} with *ga* (nominative case) or *wa* (topic marker) changes the meaning. Moreover, *mo* does not focus specifically on 'spring'; instead it focuses on the entire proposition and juxtaposes it with other things reminiscent of the season and the transition of time [49, 173]. The speaker presupposes prior situations such as the cherry blossoms are blooming, the trees are green, and the weather is warm. At LF, *mo* takes higher scope above the proposition:

(34) LF: mo [haru [takenawa-ni nari-mashi-ta]]
also spring peak-GOAL become-HON-PAST
'The spring has reached its peak'

[6] These examples are courtesy Hotze Rullman.

4.2 Negative and Positive Emotions

I argue that this type of *also/even/too* is licensed by the speaker's emotions. In (33), the speaker is happy to find that spring has reached its peak. However, most of the predicates of this kind are rather negative:

(35) a. Yo-**mo** fukete-ki-ta. Mo neru-to shi-yo. [7]
night-also late-become-PAST already sleep-COMP do-will
'It grew late (at night). It's time to go to bed'

b. Soto-**mo** hiete-ki-ta.
outside-also cold-ASP-PAST
'It has become cold outside'

c. Tabi-**mo** owari-ni chikazuite-ki-ta.
trip-also end-to approach-ASP-PAST
'The trip is nearing the end'

d. Ko-no saifu-**mo** furuku-nat-ta.
this-GEN wallet-also old-become-PAST
'This wallet has become old'

e. Mari-**mo** kashiko-i.
Mari-also smart-be
'Mary is smart indeed'

f. Omae-**mo** aho-ya-na.
you-mo silly-be-EXC
'You are silly, I should say'

Many of the above sentences can be embedded under negative factive predicates such as *I regret*, and not by positive predicates such as *be glad*:

(36) a. Zannenna-koto-ni soto-mo hiete-ki-ta.
unfortunate-fact-be outside-also cold-ASP-PAST
'I am sorry that it has become cold outside'

b. #Ureshii-koto-ni soto-mo hiete-ki-ta.
happy-fact-be outside-also cold-ASP-PAST
'I am glad that it has become cold outside'

4.3 Korean *To* 'also/even' and Chinese *Ye* 'also'

Korean *to* 'also/even' has a similar usage.

(37) Pom-**to** wat-ta.
spring-also come-PAST
'Spring came' (That's why I'm so sad)

To 'also/even' demonstrates speaker's attitude, and so does the sentence focus *ye* 'also' in Mandarin Chinese.[8]

[7] This example is modified from [51, 234].
[8] I thank Heejeong Ko and Chih-Hsiang Shu for making me aware of these facts.

(38) a. Qiutian **ye** lai-le.
fall also come-PERF
(In view of the foregoing events) 'Fall came'

b. #Qiutian **ye** bu lai-le.
fall also NEG come-PERF
'Fall has not come'

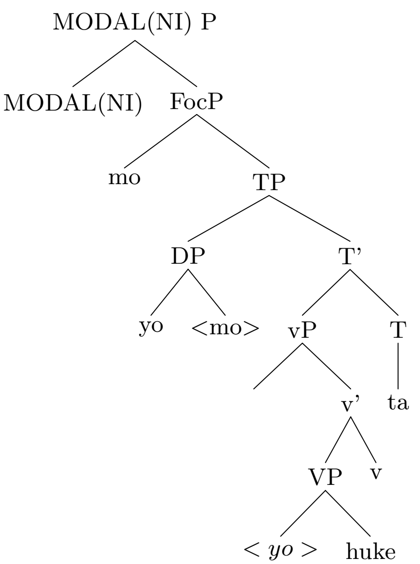
4.4 Sentence Focus and Presuppositions

The discourse initial *mo/to/ye* 'also' is the sentence focus [cf. 53] that quantifies over a proposition and triggers a presupposition [cf. 54]. For example, (35a) presupposes other events that lead to the utterance such as *it has become quiet, the neighbors turned off the lights* and *it has become quiet.*

(39) Sentential Focus:
$[[\text{night}]_F$ ALSO $[\text{grew late}]_F]$

At LF, *mo* moves and adjoins to TP in order to take a wide scope over the whole proposition.

(40) [NI[mo[yo-<mo>-huke-ta]]] (NI: negative implicature)

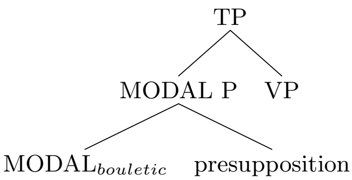


4.5 Bouletic Modality

A speaker's attitude licenses the kind of *also/even* that is mentioned above. The speaker is saddened or sentimental when seeing a worn-out bag, the end of a trip, a late night, and so on. The bouletic modal finds the situation to be undesirable.

(41) Conversational background g (bouletic modality):
ϕ is lowly ranked

Bouletic modality embeds utterances and licenses *also/even/too*. The covert modal functions as a determiner [6, 7, 8], taking implicit presupposition in the restrictor and overt unaccusative predicates in the nuclear scope.



The presupposition of *also/even/too* is accommodated since emotive modality is a monster that changes the context parameters of *also/too* sentences.

(42) MODAL:(s→t)→(s→t)
$$\text{MODAL}([\![\phi - also]\!]^{<c,i>}) = 1 \text{ iff } [\![\phi]\!]^{<c^{[w_i/w_c]},i>} = 1$$

4.6 Information Update

Utterances with *also/even* draw the hearer's attention to a fact that the hearer had not been so aware of, that is, *it had already grown late*, *the spring reached its peak*, and so on. On the other hand, a sentence without *also/even* is merely a statement of the fact and does not convey any new information.

In other words, the sentence with *also/even* offers additional information, compared to the one without them. The latter only contains the information that is common knowledge between the speaker and the hearer [cf. 55, 56].

On the contrary, the sentence with *also/even* adds 'I believe that we must know that ϕ':

(43) BELIEVE$_a$ MUST KNOW (speaker and hearer, ϕ)

5 Conclusion

In this paper, I investigated how the implicit speaker modality shifts context parameters. I discussed three phenomena that have rarely been discussed in the literature. Modality shifts the temporal parameters in fake past and fake present narrative sentences as well as the world parameters in the discourse initial *also/even/too* utterances.

References

[1] Schlenker, P.: Propositional Attitudes and Indexicality: A Cross-Categorial Approach. PhD thesis, MIT (1999)
[2] Schlenker, P.: A plea for monsters. Linguistics and Philosophy **26** (2003) 29–120
[3] Anand, P., Nevins, A.: Shifty operators in changing contexts. In: SALT 14. (2004) 20–37

[4] Teramura, H.: Nihongo no Shintakusu to Imi II. Kuroshio, Tokyo (1984)
[5] Soga, M.: Tense and Aspect in Modern Colloquial Japanese. University of British Columbia Press, Vancouver (1983)
[6] Kratzer, A. In: Modality. Walter de Gruyter, Berlin, New York (1991) 639–650
[7] Berman, S.: On the Semantics and Logical Form of Wh-Clauses. PhD thesis, University of Massachusetts at Amherst (1991)
[8] von Fintel, K.: Restrictions on Quantifier Domains. PhD thesis, University of Massachusetts at Amherst (1994)
[9] Ippolito, M.: Presuppositions and implicatures in counterfactuals. Natural Language Semantics **11** (2003) 145–186
[10] Kaplan, D. In: Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. Oxford University Press, New York, Oxford (1977, 1989)
[11] Heim, I.: The semantics of definite and indefinite noun phrases. PhD thesis, University of Massachusetts, Amherst (1982)
[12] Iatridou, S.: The grammatical ingredients of counterfactuality. Linguistic Inquiry **31** (2000) 231–270
[13] Kunihiro, T.: Kozoteki Imiron. Sanseido, Tokyo (1967)
[14] Mikami, A.: Gendai Goho Josetsu. Kuroshio, Tokyo (1953, 1972)
[15] Machida, K.: Nihongo no Jisei to Asupekuto. Arc, Tokyo (1989)
[16] Mo, S.: Jotai hyogen-no *Ru*-kei-to ta-kei: Ryokeishiki-no imi-to sono kinjisei. Nihongo to Nihonbungaku **17** (1992) 43–50
[17] Inoue, M., Ogoshi, N.: Kakokei no shiyo ni kakawaru goyoronteki yoin: Nihongo to chosengo no baai. In Kenkyujo, K.K., ed.: Nihongo Kagaku 1. Tosho Kanko Kai, Tokyo (1997)
[18] Inoue, M.: Gendai nihongo no ta. In: "Ta" no Gengogaku. Hitsuji Shobo, Tokyo (2001) 97–163
[19] Nishiguchi, S.: Temporal dynamic semantics of factual counterfactuals. In: Proceedings of JSAI 2004 International Workshop, Logic and Engineering of Natural Language Semantics (LENLS 2004), Kanazawa, Japan (2004) 73–81
[20] Nishiguchi, S.: Temporal dynamic semantics of factual counterfactuals. In Hasida, K., Nitta, K., eds.: Lecture Notes in Artificial Intelligence: New Frontiers in Artificial Intelligence: Joint Proceedings of the 17th and 18th Annual Conferences of Japanese Society for Artificial Intelligence. Springer-Verlag, Berlin (in press)
[21] Kusumoto, K.: The semantics of non-past *-ta* in japanese. In: MIT Working Papers in Linguistics. Volume 41. (2001) 163–180
[22] Ogihara, T.: Counterfactuals, temporal adverbs, and association with focus. In: SALT. Volume 10. (2000) 115–131
[23] Carlson, G.: Reference to Kinds in English. PhD thesis, University of Massachusetts, Amherst (1977)
[24] Perlmutter, D.M.: Impersonal passives and unaccusative hypothesis. In et al., P.M.P., ed.: Proceedings of the Fourth Annual Meeting of Berkeley Linguistics Society, Berkeley, Berkeley Linguistics Society (1978) 157–189
[25] Levin, B., Rappaport Hovav, M.: Unaccusativity: At the Syntax-Lexical Semantics Interface. The MIT Press, Cambridge (1995)
[26] Abe, Y.: Dethematized subjects and property ascription in japanese. In: Proceedings of PACLIC 10. (1993) 132–144
[27] Ogihara, T.: Adjectival relatives. Linguistics and Philosophy **27** (2004) 557–608
[28] Lewis, D. In: Index, Context, and Content. Reidel, Dordrecht (1980)

[29] Hyman, L.M.: Aghem grammatical structure. In: Southern California Occasional Papers in Linguistics. Volume 7, Part2. University of Southern California, Los Angeles (1979)
[30] Speas, M.: Person and point of view in navajo direct discourse complements. In Jelinek, E., ed.: WCCFL Papers in Honor of Ken Hale, Cambridge, MIT Press (2000)
[31] Partee, B.: Some structural analogies between tenses and pronouns in english. The Journal of Philosophy **70** (1973) 601–609
[32] Nishiguchi, S.: Bipolar items and emotive modality. In: Proceedings of Swarthmore Workshop on Negation and Polarity, Amsterdam, John Benjamins (to appear)
[33] Karttunen, L., Peters, S.: Conventional implicature. In Oh, C.K., Dinneen, D., eds.: Syntax and Semantics. Volume 11. Academic Press, (New York) 1–56
[34] Lewis, D.: Counterfactuals. Blackwell, Oxford (1973)
[35] Binnick, R.I.: Time and the Verb: A Guide to Tense and Aspect. Oxford University Press, New York Oxford (1991)
[36] von Fintel, K., Iatridou, S.: What to do if you want to go to harlem: Anankastic conditionals and related matters. (2005)
[37] Lewis, D.: Counterpart theory and quantified modal logic. Journal of. Philosophy **65** (1968) 113–26
[38] Kripke, S.A.: Naming and Necessity. Harvard University Press, Harvard (1972)
[39] Diesing, M.: Indefinites. MIT Press, Cambridge (1992)
[40] Kratzer, A.: What *Must* and *Can* must and can mean. Linguistics and Philosophy **1** (1977) 337–355
[41] Chomsky, N.: Minimalist Program. MIT Press, Cambridge (1995)
[42] van der Sandt, R.A. Journal of Semantics (1992) 333–377
[43] Roberts, C.: Anaphora in intensional contexts. In Lappin, S., ed.: The Handbook of Contemporary Semantic Theory. Blackwell, Oxford (1996) 215–246
[44] Belletti, A., Rizzi, L.: Psych-verbs and θ theory. Natural Language and Linguistic Theory **6** (1988) 291–352
[45] Klein, W.: Time in Language. Routledge, London, New York (1994)
[46] Nara, H.: Tense alternation in written narrative texts in japanese: A case study of natsume sooseki's *Botchan*. The Southern Japan Seminar, Spring Meeting, Atlanta, Georgia, 28 April, available at [http://web.aall.ufl.edu/SJS/Nara.pdf] (2001)
[47] Kuno, S.: Functional Syntax. The University of Chicago Press, Chicago, London (1987)
[48] Numata, Y.: Toritateshi (focus particles). In et al., K.O., ed.: Iwayuru Nihongo Joshi no Kenkyu. Bonjinsha, Tokyo (1986)
[49] Numata, Y.: Toritate (focus). In et al., S.K., ed.: Toki, Hitei to Toritate. Iwanami, Tokyo (2000) 151–216
[50] Watanabe, A.: The genesis of negative concord: Syntax and morphology of negative doubling. Linguistic Inquiry **35** (2004) 559–612
[51] Sadanobu, T.: Shinteki prosesu kara mita toritateshi *Te*, *Demo*. (1997)
[52] Teramura, H.: Nihongo no Shintakusu to Imi III. Kuroshio, Tokyo (1984)
[53] Lambrecht, K.: Information Structure and Sentence Form. Cambridge University Press, Cambridge (1994)
[54] Hamblin, C.L.: Questions in montague grammar. Foundations of Language **10** (1973) 41–53
[55] Groenendijk, J., Stokhof, M.: Dynamic predicate logic. Linguistics and Philosophy **14** (1991) 39–100
[56] Veltman, F.: Defaults in update semantics. Journal of Philosophical Logic **25** (1996) 221–261

Conversational Implicatures Via General Pragmatic Pressures*

Christopher Potts

UMass Amherst, Amherst MA 01002, USA
potts@linguist.umass.edu
http://people.umass.edu/potts/

Abstract. This paper aims to show how general pragmatic pressures, interacting with the context of utterance, can produce specific conversational implicatures — as well as the lack thereof in nonsupporting environments. Inspired by the work of Merin (1997), Blutner (1998), van Rooy (2003a), Benz et al. (2005), and others, I use probabilities to represent speakers' belief states and the content of their utterances. These values determine an utterance's *quality rating* and *quantity rating*. I adapt Roberts' (1996) view of the question under discussion to define a *relevance ranking* of utterances. These values come together in a definition of *felicitous utterance*. This definition licenses certain inferences — relevance implicatures relating to the question under discussion (section 4.1) and a variety of quantity implicatures (sections 4.2–4.3).

1 Introduction

Conversational implicatures can be exquisitely sensitive to subtle changes in the context. An utterance might conversationally implicate a meaning M very robustly in one context, but a slight change to that context — a shift in the question under discussion, a revelation about the speaker's belief state, some additional linguistic material — might cause M to vanish.

My goal for this paper is to define a system of pragmatic pressures that can capture such extreme context sensitivity. Limitations of space and understanding lead me to concentrate on a specific subclass of conversational

* I've benefitted from an amazing amount of useful feedback on this general project. I'm indebted in particular to Jan Anderssen, Ash Asudeh, Rajesh Bhatt, Daniel Büring, Donna Byron, Shai Cohen, Regine Eckhardt, Kathryn Flack, Lyn Frazier, Hans-Martin Gärtner, Yurie Hara, Larry Horn, Gerhard Jäger, Pauline Jacobson, Mark Johnson, Ed Keenan, Angelika Kratzer, Manfred Krifka, Eric McCready, Barbara Partee, Tim Roeper, Tom Roeper, David Schueler, Florian Schwarz, Raj Singh, Tim Stowell, Anna Verbuk, Mike White and audiences at the UMass Amherst Semantics Group (Sep 6, 2005), Göttingen University (Sep 21, 2005), ZAS (Sep 23, 2005), UCLA (Dec 1, 2005), Yale (Nov 7, 2005), OSU (Feb 24, 2006), and Brown (Mar 6, 2006). Craige Roberts provided some crucial insights at a crucial moment in my development of these ideas.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 205–218, 2007.

implicatures, namely, those that we can trace back to a specific question under discussion. In this, I derive my inspiration primarily from Roberts (1996, 2004) and van Rooy (2003a), who regard the question under discussion as central both to understanding specific utterances and to following the general flow of conversation.

Section 2 gathers the requisite tools: I define a probabilistic perspective on propositions, then use the numerical values to define approximations of the Gricean maxims of quality, quantity, and relevance. In section 4, I show how this system of pragmatic pressures models conversational implicatures — and the lack thereof, where appropriate. In the closing section, I situate these results within the general theory of conversational implicatures, concentrating on the issue of whether generalized conversational implicatures are derived via the usual pragmatic mechanisms (Horn, 2005) or included in the semantico-pragmatic system as presumptive meanings that are present unless cancelled (Levinson, 2000; Chierchia, 2004).

Throughout, I have tried to give the central definitions as precisely as possible. This can make them appear somewhat complicated and involved, but I think this is just a consequence of the step-by-step format. It's my hope that the definitions are usable, more or less off the shelf, by researchers with computational[1] or mathematical aims as well as linguistic ones.

2 A System of Pragmatic Pressures

As a discourse participant, one could, in principle, say anything at all. The physical constraints are few; one could voice any number of lies and irrelevancies, at whatever length one wished. But this is not usually how things go. We (generally) limit ourselves to what is (basically) felicitous. And the set of pragmatically felicitous utterances at a given point in a discourse is highly constrained when compared with the utterance space at one's disposal.

What pressures shape the space of felicitous utterances? The Gricean maxims are a prominent, historically important answer. Grice identified pressures for truthfulness (quality), informativity (quantity), and relevance, as well as a general pressure to be clear and concise (manner). In this paper, I focus on the first three. The next section moves us from viewing propositions as sets to viewing them as probability distributions. This lays the groundwork for my particular take on the maxims of quality, quantity, and relevance.

2.1 Probabilities and Propositions

The foundational technical move of this paper is a shift in the usual perspective on propositions. We are accustomed to thinking of them as sets of intensional

[1] I have developed basic computational tools for solving problems within this theory and exploring its predictions. The CGI interface is linked from <http://people.umass.edu/potts/computation/>.

indices. But Merin (1997), van Rooy (2003b), Benz et al. (2005), and others have shown how we can fruitfully think of them instead as probability distributions when we are in the realm of utterances:

(1) For countable W, a function $P : \wp(W) \mapsto [0, 1]$ is a probability distribution iff:
  a. $P(W) = 1$; and
  b. if p and q are disjoint subsets of W, then $P(p \cup q) = P(p) + P(q)$.

These general probability distributions alone are not quite the right structures, however. We have a correspondence between, e.g., the proposition $\{w_1, w_2\}$ and the probability distribution P such that $P(\{w_1\}) = P(\{w_2\}) = .5$, but we also have probability distributions like $P'(\{w_1\}) = .6; P'(\{w_2\}) = .4$, which do not treat w_1 and w_2 equally and thus do not correspond to any proposition. To bring these two classes of structure more in line, we need to build in an additional requirement:

(2) The probability distribution P mimics the proposition q (a subset of W) iff:
  a. $P(\{w\}) = 0$ iff $w \notin q$
  b. $P(\{w\}) = P(\{w'\})$ for all $w, w' \in q$

Throughout this paper, I assume a dual view. In the semantics, propositions are sets of worlds (with all the important boolean structure this provides; Keenan and Falz 1985; Kamp and Partee 1995). In the pragmatics, propositions are probability distributions (with all the measuring techniques that these facilitate).

2.2 Quality Thresholds

The primary imperative of Grice (1975) is one of truthfulness: the maxim of quality enjoins us to be truthful, to say only those things for which we have adequate evidence. For the purposes of this paper, I factor out the difficult concepts of knowledge and evidence. The approach is instead framed in terms of belief. But this is, as far as I can tell, a move of convenience rather than necessity or substance; the theory of knowledge defined by Kratzer (2002) seems well suited to combination with the present approach.

From this simplified perspective, we can read the quality maxim as an injunction to confine oneself to utterances whose content is entailed by one's belief state (Groenendijk, 1999). Given the probabilistic perspective described above, this would play out as follows. Let P_S be the probability distribution modeling speaker S's belief state. Then we associate *quality ratings* with probabilities for the speaker:

(3) **Quality ratings**
  The *quality rating* of an utterance U by speaker S in context C is $P_S([\![U]\!])$.

The Gricean imperative would then be that a speaker S should confine himself to utterances U such that $P_S([\![U]\!]) = 1$.

In practice, though, we are not nearly this strict. We can be lax on quality, as when we brainstorm new ideas or participate in bull sessions (Frankfurt, 1986). Conversely, we can be quite strict on quality, as when we maneuver to land rockets on the moon or instruct our students (perhaps). This movement towards the ideal — a maximal probability — might never quite finish; if we are extremely skeptical, we might allow that we don't fully believe (know) anything. An element of doubt creeps in, the probabilities drop. Therefore, I propose that each context comes with a *quality threshold* C_τ. This is a numerical value in the real interval $[0, 1]$:

(4) **Quality thresholds**
An utterance U by speaker S in context C satisfies quality iff its quality rating is above the quality threshold C_τ for C.

We make use of these thresholds in a very simple way: I, as a speaker in context C, am forbidden from saying anything with a probability (according to my beliefs) that is below C_τ. The practical effect of this principle is to remove from consideration a great many things that, informative and relevant though they may be, are simply not supported by my epistemic state.

2.3 Quantity Ratings

A simplified version of Grice's maxim of quantity might read: Be informative! The actual statement is more complex than this, including as it does some information about what is required in the current context. But that work is, in the present system, achieved by relevance, so we do not skimp on coverage by purifying the quantity maxim to an unqualified demand for information.

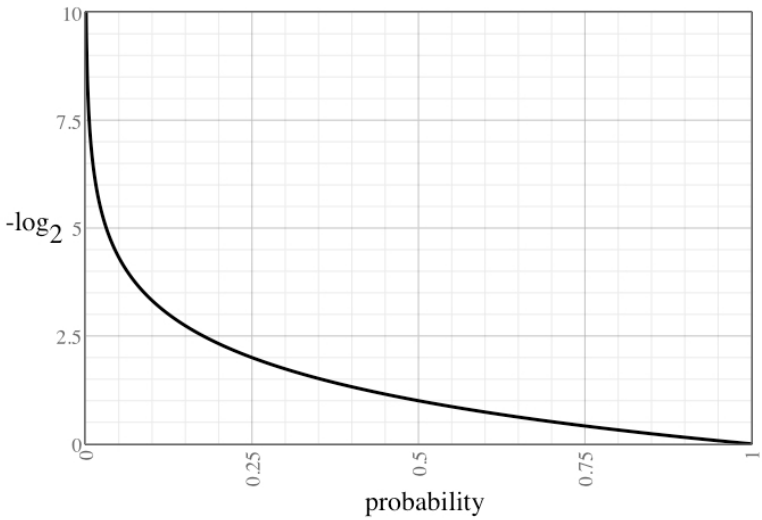
As with quality, I translate quantity into a setting in which we can take measurements, and the probabilities of section 2.1 are again the starting point. We use them to derive a standard value from information theory: informativity values, as defined in (5).

(5) **Information value of p for individual a**

$$\inf_a(p) = -\log_2 P_a(p)$$

I've chosen this particular scaling of probability values simply because it meets the two central demands in this area: (i) high probabilities correspond to low inf values, and (ii) low probabilities correspond to high inf values:

(6) **As probabilities drop,** inf $(-\log_2)$ **values rise**



The translation into pragmatic theory is direct: the inf values derived from the hearer's probabilities are the *quantity ratings*.

(7) **Quantity rating**
The *quantity rating* of an utterance U by speaker S to hearer H in context C is

$$\text{Quantity}_C\, U = \inf_H([\![U]\!])$$

Quantity measurements are taken relative to the hearer's belief state, P_H. We cannot define these values directly in terms of the speaker's information state, as this would place quantity in unwanted conflict with the quality maxim: quantity values would favor low probabilities while quality favored high probabilities. The result would likely be silence.

In practice, this hearer orientation means that a speaker must guess about the belief state of his addressee. But quantity values should be measured in terms of the actual belief state of the hearer, not in terms of the speaker's guesses about that belief state. Suppose I guess that you do not know the finer points of training for a marathon, but my guess is wrong: you are an expert on the topic. I begin to tell you about long training runs, proper hydration, and the misery of mile 18, and you grow agitated. What I am saying has very low quantity in this context. You are understandably annoyed: my behavior is infelicitous, and (7) tells us why.

2.4 Relevance Ranking

My approach to relevance brings to the fore two salient features of relevance as an intuitive notion: it is a matter of degree, and it is highly dependent upon the current context. The current system is built up from a gradient foundation (the probabilities), and all its values are derived by integrating information from the context. So it is well suited to providing a formal notion of relevance. (Here, we needn't worry about slighting Grice, as his maxim of relevance simply says "Be relevant".)

In this system, relevance is given as a ranking. Arriving at the ranking turns out to be a bit complex, but the procedure is derived from the insights of Groenendijk and Stokhof (1982) and van Rooy (2003b), and its motivations are intuitive (despite its involved sorting and selection procedure).

The notion of relevance defined here is relevance to a question. I assume a partition semantics for questions: a question contains a set of mutually exclusive alternatives, and asking a question can be construed as asking for an identification of the true alternative. So we get a good measure of the degree to which a proposition p answers a question Q by counting the number of cells in Q with which p has a non-null intersection. This count is the Ans value, as defined in (8).

(8) a. $p_Q = \{q \in Q \mid q \cap p \neq \emptyset\}$ (for p an answer to question Q)
b. $\mathrm{Ans}(p, Q) = |p_Q|$

This is a very partial ordering. One fact about the ordering is particularly important for getting at relevance: if A is a proper subset of B, then A and B have the same Ans values for any question that contains B or one of its supersets as a member. In practice, this means that we cannot yet distinguish a complete answer from a complete and overly informative answer. But overly informative answers are reliably judged to have degraded relevance values (Sperber and Wilson, 1995).

But the system has within it the potential to distinguish A from B in this example: the quantity rating of A will be at least as high as (probably higher than) the quantity rating of B. So we can use quantity to tighten up the ordering of answers relative to a given question:

(9) **Relevance-ranking**

i. Sort the space of utterances with quality ratings above the threshold into equivalence classes based on Ans-values.
ii. For each Ans-equivalence class, get the utterances with the lowest quantity ratings in that class. Keep them, and throw out the rest.
iii. The Ans ordering of the remaining set is the relevance ranking.

The representation in (10) might help convey the sense of (9). (For any number n, $\mathrm{Ans}\, n = \{p \mid \mathrm{Ans}(p) = n\}$.)

(10) **Strikeouts indicate relevance-based eliminations.**

Question Q

$$\left\{ \begin{array}{c} \{w_1, w_2\} \\ \{w_3, w_4\} \end{array} \right\} \Longrightarrow$$

Ans ordering for Q

$$\begin{array}{l} \mathrm{Ans}\,1 = \left\{ \begin{array}{cc} \{w_1, w_2\} & \{w_3, w_4\} \\ \cancel{\{w_1\}}\ \cancel{\{w_2\}} & \cancel{\{w_3\}}\ \cancel{\{w_4\}} \end{array} \right\} \\ \mathrm{Ans}\,2 = \left\{ \begin{array}{c} \{w_1, w_2, w_3, w_4\} \\ \cancel{\{w_1, w_2, w_3\}}\ \cancel{\{w_1, w_2, w_4\}}\ \cancel{\{w_2, w_3, w_4\}} \\ \cancel{\{w_1, w_3\}}\ \cancel{\{w_1, w_4\}}\ \cancel{\{w_2, w_3\}}\ \cancel{\{w_2, w_4\}} \end{array} \right\} \end{array} \Longrightarrow$$

Relevance ranking for Q

$$\begin{array}{ccc} \{w_1, w_2\} & & \{w_3, w_4\} \\ & \searrow \quad \swarrow & \\ & \{w_1, w_2, w_3, w_4\} & \end{array}$$

So if a speaker knows that $\{w_1\}$ is the actual world, then he will answer Q with $\{w_1, w_2\}$ — a relevance ranked utterance that scores a 1 on quality. Of course, $\{w_1\}$ is also a 1 on quality. But this overly informative answer is not relevance ranked. (Section 4.1 contains a fuller discussion of overly informative answers.)

3 Felicitous Utterances

It's time now to pool together the above concepts into a single definition of *felicitous utterance*. The backdrop for it is a view of context as tuples:

(11) A *context* is a tuple

$$\left\langle P_S, P_H, Q, C_\tau, U \right\rangle$$

where P_S is the speaker's belief state (probability function), P_H is the hearer's belief state, Q is a question under discussion, C_τ is a quality threshold, and U is an utterance.

The definition of utterance felicity takes the form of a ranking of utterances:

(12) **Felicitous utterances**
The set of felicitous utterances for a partial context $C = \langle P_S, P_H, Q, C_\tau \rangle$ is obtained as follows:

i. From the set of all propositions, eliminate those that have quality ratings at or below C_τ. (See (4).)
ii. With the resulting set, determine relevance rankings and throw out all utterances without such rankings. (That is, throw out every utterance that is not among the least informative members of its Ans-equivalence class. See (9).)

iii. From the resulting set of relevance-ranked utterances, extract the utterances with the lowest Ans values.
iv. From the resulting set, select the utterances with the highest quantity ratings. These are the felicitous utterances for C.

Once again, a slightly more visual representation might help to articulate the workings of this definition:

(13) Let W be the set of all possible worlds, and let $P = \wp(W)$ be the set of all propositions. The partial context is $C = \langle P_S, P_H, Q, C_\tau \rangle$, and we are determining which utterances U are felicitous in C.

Quality elimination
$P_Q = P - \{p \mid P_S(p) \leqslant C_\tau\}$ $\Longrightarrow$

Relevance elimination and Ans**-minimization**
$P_R = \min_{\mathrm{Ans}}(P_Q - \{p \mid p \text{ is not relevance ranked}\})$ $\Longrightarrow$

Quantity maximization
$P_{\text{felicitous}} = \max_{\text{Quantity}}(P_R)$

So P_Q is the set of utterance-contents that satisfy quality. P_R is the set of utterance-contents that satisfy relevance. And $\max_{\text{Quantity}}(P_R)$ is the set of all maximally informative things in the remaining set.

The next section is devoted to showing how we can use this definition to understand a range of conversational implicatures.

4 Conversational Implicatures

The algorithm described in section 3 is fruitfully thought of as a way of solving for the utterance element in our view of contexts:

(14) $P_S, P_H, Q, C_\tau \Rightarrow$ a ranking of utterances

It is illuminating to ask what happens when we fix the utterance U and instead solve for one (or more) of the other elements. The next few subsections do just that, pausing along the way to highlight the positive work that is done at each turn.

4.1 Relevance-Based Implicatures

What happens when we fix the elements P_S, P_H, C_τ, and U, and then ask about the question under discussion Q? The definition in (12) licenses at least one important inference about Q:

(15) Suppose a speaker S has uttered U in context C. Then the set of potential questions under discussion is not larger than the set of all Q such that U has a relevance ranking according to Q.

To make this concrete, imagine the following situation: you are a cooperative speaker. You say, "I live in New York". Then I know that the question under discussion (at least in your mind) is not Q_{country} = *Which country do you live in?*. For suppose it were. Then "I live in New York" would share an Ans value with "I live in the U.S.". By (12ii), it would be ruled out. But you are, by assumption, obeying this restriction. Hence the supposition is false — Q_{country} is not the question under discussion.

Though limited, (15) can make sense of the following dialogue, based on an example from Grice 1975:

(16) A: Is Smith happy at his new job?
B: Yes, and he hasn't even been to jail yet.

B's answer here is not relevance ranked if the question under discussion is given by A's utterance. It is eliminated in virtue of the less informative but complete answer "Yes". Let's assume that B is being truthful, and that his utterance is informative to A (which is basically guaranteed by A's question; see section 4.2). Then B has failed only on relevance. In Grice's terms, he has *flouted* it.

If B is in fact playing by the rules, then it must be because he is speaking relative to a different question under discussion. Some questions that do allow B's reply to have a relevance ranking are "Is Smith both happy at his new job and a free man?", as well as "Is Smith happy at his new job and is his job illegal". And so forth. B's utterance demands that we change the question away from A's. It does not tell us which question to switch to, so there is some indeterminacy. In practice, this seems right: it is clear that B intends something extra with his utterance. But what exactly that something extra is — that is typically unclear.

4.2 Quantity-Based Implicatures of Questions

Quite generally, if a speaker knows the answer to question Q, then it is highly marked for him to ask Q. The present system captures this infelicity by looking to the addressee — the answerer:

(17) If the speaker already knows the answer to his question, then the quantity rating of any felicitous answer, as defined in (12), will be disastrously low.

Quantity ratings are defined in terms of the addressee's belief state. If the addressee assigns p the probability 1, then the quantity rating for p (its inf value, as in (5)) is 0. So if I know p and yet persist in asking you whether-p, then I put you in a very bad spot indeed.

Of course, the answerer could in principle boost the quantity rating of his answer: he could give a false answer, which would have a pathological (∞) quantity rating, or he could say something informative but irrelevant. However, neither move is permitted, by (12). Unsupported or unsupportable claims are eliminated at (12i), and irrelevancies are eliminated at (12ii).

The system has a bit more to say on the matter. At step (12iii), we extract the elements with the lowest relevance ranking. The quantity maximization step, (12iv), is then taken in terms of this small set of elements with equivalent Ans values. One might wonder whether we could simplify the system by maximizing on the full set of relevance-ranked utterances, not just those with the lowest Ans-values, thereby removing step (12iii). However, situations in which the questioner knows the answer to his question help justify the multi-step process.

For instance, suppose A knows that Barbara lives in Moscow but nonetheless asks me "Which city does Barbara live in?". The utterances in (18) are all relevance ranked and have identical quantity values in this context.

(18) a. "Barbara lives in Moscow." quantity $= 0$
b. "Barbara lives in Russia." quantity $= 0$
c. "Barbara lives on earth." quantity $= 0$

But, intuitively, my felicitous move is (18a) (assuming I assign it a probability above the quality threshold). This follows from (12) and the fact that it has the lowest Ans value of the three.

Potts (2006) offers a fuller discussion of (17) and addresses some prima facie counterexamples to it.

4.3 Quantity-Based Implicatures of Answers

Using (12), we can infer a lot about the speaker's belief state. When we do this, we are essentially asking how much we can infer about the output set in (19).

(19) $Q, C_\tau, U \Rightarrow$ a set of potential belief states for the speaker

That is, we know what the question is, we know what the speaker said, and we know where the quality threshold is. What can we infer about the speaker's beliefs in such a situation? The following is the primary tool for answering this question:

(20) Suppose the speaker S uttered U in a context C with question Q. The set of potential belief states for S is the set of all P_S such that
a. the speaker's utterance is above the quality threshold according to P_S (i.e., $P_S([\![U]\!]) > C_\tau$); and
b. the speaker could not have answered A more completely with P_S (i.e., there is no utterance U' such that $P_S([\![U']\!]) > C_\tau$ and $\mathrm{Ans}_Q[\![U]\!] > \mathrm{Ans}_Q[\![U']\!]$).

This is just to say that the hearer can assume that the speaker will do his best to answer the question under discussion — up to the quality threshold. It determines a set of probability distributions — potential belief states for the speaker as far as the hearer knows based on what the speaker said and what the question under discussion is. These sets encode epistemic indeterminacy, and thus they tell us about the extent to which we draw clausal conversational implicatures.

To illustrate, I again turn to a variation on one of Grice's (1975) examples:

(21) A: Which city does Barbara live in?
B: Barbara lives in Russia.

As Grice observes, we typically draw the implicature from B's utterance that B does not know which city Barbara lives in. To show how the system predicts this, let's fix some details, as in (22) and (23).

(22) $W = \{w_1 \dots w_4\}$

$$\begin{aligned} [\![\textit{Barbara lives in Russia}]\!] &= \{w_1, w_2\} \\ [\![\textit{Barbara lives in Moscow}]\!] &= \{w_1\} \\ [\![\textit{Barbara lives in Petersburg}]\!] &= \{w_2\} \end{aligned}$$

$$\begin{aligned} [\![\textit{Barbara lives in Germany}]\!] &= \{w_3, w_4\} \\ [\![\textit{Barbara lives in Berlin}]\!] &= \{w_3\} \\ [\![\textit{Barbara lives in Cologne}]\!] &= \{w_4\} \end{aligned}$$

(23) a. $C_\tau = .9$
b. $Q = \big\{\{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}\big\}$
c. U (said by B) = "Barbara lives in Russia" relevance ranking = 2

Principle (20b) licenses A to infer from B's utterance that the quality threshold prevented B from saying any of the utterances with lower relevance rankings ("Barbara lives in Moscow", "Barbara lives in Petersburg", "Barbara lives in Berlin", and "Barbara lives in Cologne", which each have relevance rankings of 1). Thus, given this limited space, the only belief state B could be in is the following:

(24) a. $P_B([\![\textit{Barbara lives in Moscow}]\!]) = .5$
b. $P_B([\![\textit{Barbara lives in Petersburg}]\!]) = .5$

This is exactly the quantity implicature that Grice identified.

Minimally different contexts fail to support this inference, though. For instance, consider the following scenario, also based on the model in (22):

(25) a. $P_A(\{w_i\}) = .25$ for all $1 \leqslant i \leqslant 4$
b. A: "Does Barbara live in Russia?" $\big\{\{w_1, w_2\}, \{w_3, w_4\}\big\}$

In this scenario, according to the above system, B will answer the polarity question with "Yes, (she lives in Russia)". Here is a summary of the measurements:

(26)

utterance	quality	relevance	quantity
$[\![\textit{In Russia}]\!] = \{w_1, w_2\}$	1	1	1
$[\![\textit{On earth}]\!] = \{w_1 \dots w_4\}$	1	2	0
$[\![\textit{In Moscow}]\!] = \{w_1\}$	1	1	2

Though "In Moscow" is more informative, it is not even relevance ranked, hence not even a contender: it shares a relevance ranking with the less-informative "In Russia", and thus it is eliminated at step ii of (12).

And, in turn, we do not make the standard quantity implicature that Grice articulated for (21). That is, we do not infer that B lacks more specific knowledge than he offers here with his "Yes" answer. We can use (20) to see that the context underdetermines B's belief state. As far as A is concerned, B might be in any of the following and still be playing by the rules of the pragmatic game:

(27) Contenders for B's belief state $= \{P_a, P_b, P_c\}$
a. $P_a(w_1) = P_a(w_2) = .5$
b. $P_b(w_1) = 1$
c. $P_c(w_2) = 1$

And this is just to say that we do not infer that B lacks more specific knowledge. We are open to the possibility that he doesn't have it ($= P_a$) and we are open the possibility that he does ($= P_b$ or P_c).

5 Summary and Prospects

The foundational move of this paper is the shift to probabilities in the pragmatics (section 2.1). The resulting values form the basis for statements of the pressures that arrive via versions of quality, quantity, and relevance (section 2). With these values integrated (section 3), we are able to derive a range of conversational implicatures, as well as the lack of them in nonsupporting contexts (section 4).

In a sense, the results reduce to the following basic equations:

(28) a. In (12), we solve for U:

$$P_S, P_H, Q, C_\tau \Rightarrow \text{ a set of ranked utterances}$$

b. In (15), we learn something about Q:

$$P_H, U \Rightarrow \text{ a set of potential questions under discussion}$$

c. In (20), we solve for P_S:

$$Q, C_\tau, U \Rightarrow \text{ a set of potential belief states for the speaker}$$

It might be that I have not yet teased out the full set of implications of the theory for the size and composition of the output sets in (28). And it might be that a different set of pragmatic values, or a different method for combining them, will lead to even more nuanced inferences. It's my hope that this paper has at least provided a useful perspective on these matters.

I close by drawing out another consequence of this approach. A central question in pragmatic research these days is the nature of generalized conversational implicatures. Horn (2005) argues that they are implicatures like any other, and thus that they must be derived like any other — via appeal to semantic denotations, contextual information, and pragmatic pressures. They appear to have the status of default inferences only because they are supported by a wide range of contexts.

However, in the hands of Levinson (2000) and Chierchia (2004), generalized conversational implicatures are effectively lexicalized. They are, in Levinson's terms, presumptive meanings — default inferences, always present unless cancelled.

The present theory cannot decide the issue. It is consistent with both perspectives. But I hope it does show that Horn's very purely pragmatic approach is feasible. General pragmatic pressures can derive the relevant inferences. It is not necessary to stipulate them as part of the lexicon or semantic/pragmatic system. Moreover, I think results like those of section 4.3 point towards Horn's conception. There, we saw the conversational implicatures come and go as the question under discussion changed. There was no need to talk of cancellation or denial of implicatures. Rather, they arose where the context supported them, and they were simply absent where the context did not support them.

References

Benz, Anton, Gerhard Jäger, and Robert van Rooij. 2005. An introduction to game theory for linguists. In Anton Benz, Gerhard Jäger, and Robert van Rooij, eds., *Game Theory and Pragmatics*. Houndsmills, Basingstoke, Hampshire: Palgrave Macmillan.

Blutner, Reinhard. 1998. Lexical pragmatics. *Journal of Semantics* 15(2):115–162.

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In Adriana Belletti, ed., *Structures and Beyond: The Cartography of Syntactic Structures*, Volume 3, 39–103. New York: Oxford University Press. [The manuscript began circulating in 2001].

Frankfurt, Harry G. 1986. On bullshit. *Raritan* 6(2):81–100.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole and Jerry Morgan, eds., *Syntax and Semantics*, Volume 3: Speech Acts, 43–58. New York: Academic Press.

Groenendijk, Jeroen. 1999. The logic of interrogation. In Tanya Matthews and Devon Strolovitch, eds., *Proceedings of SALT IX*, 109–126. Ithaca, NY: Cornell University.

Groenendijk, Jeroen and Martin Stokhof. 1982. Semantic analysis of wh-complements. *Linguistics and Philosophy* 5(2):175–233.

Horn, Laurence R. 2005. The border wars. In Klaus von Heusinger and Ken P. Turner, eds., *Where Semantics Meets Pragmatics*. Oxford: Elsevier.

Kamp, Hans and Barbara H. Partee. 1995. Prototype theory and compositionality. *Cognition* 57(2):129–191.

Keenan, Edward L and Leonard M. Falz. 1985. *Boolean Semantics for Natural Language*. Dordrecht: D. Reidel.

Kratzer, Angelika. 2002. Facts: Particulars or information units? *Linguistics and Philosophy* 25(5–6):655–670.

Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Merin, Arthur. 1997. If all our arguments had to be conclusive, there would be few of them. Arbeitspapiere SFB 340 101, University of Stuttgart, Stuttgart. URL `http://semanticsarchive.net/Archive/jVkZDI3M/`.

Potts, Christopher. 2006. How far can pragmatic mechanisms take us? Commentary on truckenbrodt. *Theoretical Linguistics* URL `http://semanticsarchive.net/Archive/jViODAyZ/`.

Roberts, Craige. 1996. Information structure: Towards an integrated formal theory of pragmatics. In Jae Hak Yoon and Andreas Kathol, eds., *OSU Working Papers in Linguistics*, Volume Volume 49: Papers in Semantics, 91–136. Columbus, OH: The Ohio State University Department of Linguistics. Revised 1998.

Roberts, Craige. 2004. Context in dynamic interpretation. In Laurence R. Horn and Gregory Ward, eds., *Handbook of Contemporary Pragmatic Theory*, 197–220. Oxford: Blackwell.

van Rooy, Robert. 2003a. Questioning to resolve decision problems. *Linguistics and Philosophy* 26(6):727–763.

van Rooy, Robert. 2003b. Relevance and bidirectional OT. In Reinhard Blutner and Henk Zeevat, eds., *Pragmatics in Optimality Theory*, 173–210. Oxford: Palgrave Macmillan.

Sperber, Dan and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell, 2 ed.

Dake-wa: Exhaustifying Assertions*

Yurie Hara

Japan Society for Promotion of Science/Kyoto University/University of Massachusetts, Amherst
yhara@ling.bun.kyoto-u.ac.jp

1 Introduction

This paper shows that the use of the Japanese exhaustive particle *dake* 'only' in a Contrastive-marked sentence results in exhaustification over potential literal acts of assertion in the sense of [1], rather than exhaustification over propositions. Also, the data supports the idea that the exceptive meaning denoted by *dake* contributes to an expressive level of meaning.

2 Contrastive-Marking

As noted by [2], Japanese Contrastive-marking involves a morphological marker -*wa* and a prosodic peak in the intonation (indicated by capitals).

(1) a. Among John and Mary, who came to the party?
  b. JOHN-wa kita.
  John-Con came.
  'John came. (Mary didn't come, or I don't know about Mary.)'

Following the structure meaning approach (c.f. [3,4]), I argued in [5] that the prosodic peak of Contrastive-marking creates a partition of the asserted proposition into B (background) and F (Focus). The morphological *wa*-marking then introduces the CON operator that takes the structured meaning as its argument and yields a conventional implicature. The Contrastive-marked sentence presupposes that there exists a stronger alternative to the asserted proposition (2–b), and conventionally implicates that the speaker considers the possibility that the stronger alternative is false (2–c).

(2) Let F be the focus-marked elements, B the background, R the restriction. The interpretation of $\textsc{Con}(B)(F)$
  a. asserts: $B(F)$
  b. presupposes: $\exists F'[[F' \in R] \,\&\, [B(F') \Rightarrow B(F)] \,\&\, [B(F) \nRightarrow B(F')]]$
  c. implicates: $\diamond(\neg(B(F')))$

* This work is supposed by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists under Grant No. 02162. I am grateful to Chris Potts, Satoshi Tomioka, and the audience at LENS2006 for helpful comments.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 219–231, 2007.

The interpretation of (1-b) is depicted in (3).

(3) a. B= $\lambda x.\ x$ came. F=John F'= John and Mary
b. assertion: John came.
c. implicates: the speaker considers the possibility that 'John and Mary came' is false.
d. assertion+implicature: the speaker considers the possibility that 'Mary came' is false.

This treatment of Contrastive-marking predicts that if a Contrastive-marked proposition, i.e., B(F), is the strongest among its alternatives, the sentence causes a presupposition failure. This prediction is borne out by the following example. In (4), the asserted proposition, 'everyone came.' is the strongest among its alternatives. Namely, it entails all of its scalar alternatives, 'someone came.' 'most people came.' etc., and none of the alternatives entail it. As a consequence, Contrastive-marking is not compatible with the asserted proposition.

(4) #ZEN'IN-wa kita.
Everyone-Cont came
(no implicatures)

3 Puzzle

An interesting puzzle arises when a Contrastive-marked sentence contains the exhaustive particle *dake* as in (5).[1]

(5) JOHN-dake-wa kita.
John-dake-Con came.

Let us try to calculate the meaning of (5) with our current tools. If we take our F to be 'John' and F′ to be 'John and Mary', we obtain the implicature (6–c), which is incongruent with the native speakers' intuition.

(6) a. B= λx. Only x came. F=John F′=John and Mary
b. asserts:'Only John came.'
c. implicature: the speaker considers the possibility that it is not the case that only John and Mary came.
d. Unavailable reading→wrong prediction

Second, let us consider the case where we take our F to be 'only John'. (5) is predicted to cause a presupposition failure.

(7) a. B= $\lambda x.\ x$ came. F=only John
b. No implicature possible (presupposition failure)→wrong prediction

This is so for the following reason. 'Only John came.' entails 'no one else came.' as illustrated in (8).

[1] Satoshi Tomioka (p.c. to [6]) also notes that the contribution of *dake* in (5) is not clear.

(8) #JOHN-dake-ga kita. Mary-mo kita kamoshirenai.
John-dake-Nom came. Mary-Add came might
'Only John came. Mary might have come, too.'

Hence, the speaker has already given the strongest answer; there is no scalar alternative B(F′) that asymmetrically entails 'only John came' with respect to the question 'Who came?'. Therefore, (5) should fail to satisfy the presupposition requirement of Contrastive-marking (2-b).

Nonetheless, (5) is grammatical. The informal approximation of its interpretation I pursue in this paper is as follows:

(9) I make an assertion only about John with respect to the question 'Who came?' and I assert that John came.

This intuition is attested by the following examples. When *dake* is absent, the implicature of *wa* can be overtly expressed (10–a) or strengthened (10–b).

(10) Did John and Mary come?
a. JOHN-wa kita. Mary-mo kita kamoshirenai.
John-Con came. Mary-Add came might
'At least John came. Mary might have come, too.'
b. JOHN-wa kite, Mary-wa ko-nakat-ta.
John-Con came. Mary-Con come-Neg-Past
'At least John came, and Mary didn't come.'

On the other hand, when *dake* is present, the continuation is perceived as incongruent (11).[2]

(11) Did John and Mary come?
a. #JOHN-dake-wa kita. Mary-mo kita kamoshirenai.
John-dake-Con came. Mary-Add came might
'At least John came. Mary might have come, too.'
b. #JOHN-dake-wa kite, Mary-wa ko-nakat-ta.
John-dake-Con came. Mary-Con come-Neg-Past
'At least John came, and Mary didn't come.'

---

[2] This contrast between *dake-wa* and *dake* is first noted in [6]:

(i) a. #Taro-wa EEGO-dake-wa hanas-e-ru ga FURANSUGO-wa
Taro-Top English-dake-Con speak-can-Pres but French-Con
hanas-e-nai
speak-can-Neg
'English is the only langauge Taro can speak, but he cannot speak French.'
b. Taro-wa EEGO-wa hanas-e-ru ga FURANSUGO-wa
Taro-Top English-Con speak-can-Pres but French-Con
hanas-e-nai
speak-can-Neg
'Taro can speak English, but he cannot speak French.' [6, p.158]

I argue that this incongruence arises because the speaker by using *dake-wa* indicates 'John came' is the only assertion she can make with respect to the question under discussion 'Who came?', but then she continues to mention the alternative individual 'Mary'. In order to implement this intuition, I follow Yoshimura's [7] analysis which treats *dake* as an expressive item that generates a conventional implicature. I further propose that when *dake* is used in a Contrastive-marked sentence, *dake* exhaustifies over assertion potential literal acts in the sense of [1].

4 Levels of Meaning Contributed by *dake*

[8] observes that "*dake* primarily asserts the affirmative proposition while secondarily asserting the negative one."

For example, (12–b) would be infelicitous if the exceptive meaning is embedded under a conditional as in (12-b-ii). (12–b) is felicitous only under the interpretation where the affirmative proposition 'you can speak English' is embedded.

(12) a. In order to make an around-world trip,
b. EIGO-dake hanas-er-eba ii
English-dake speak-able-if good
(i) 'It's enough if you can speak English.'
(ii) #'It's enough if you cannot speak any other languages.' (Yoshimura 2005)

In contrast, if the context prefers that the negative proposition to be an argument, the use of *dake* turns out to be infelicitous as in (13).[3]

(13) #Nihongo-dake dekiru node, shuushoku deki-nakat-ta.
Japanese-dake capable because, getting.employed capable-Neg-Past
a. #'I couldn't get a job because I can speak Japanese.'
b. Intended (unavailable): 'I couldn't get a job because I cannot speak any other languages.' (Satoshi Tomioka, p.c.)

[7] provides an explanation for Kuno's [8] observation, modeling her analysis after Horn's [9] analysis of English *only*. [7] proposes that Japanese *dake* asserts the prejacent (affirmative) proposition and entails the exceptive meaning[4]

[3] Following is Yoshimura's example that makes the same point:

(i) #Ie-no roon-dake zeikin menjo nano-wa zannenna koto-da.
home-Gen loan-dake tax deductable Comp-Top too.bad thing-Cop
a. #'It is too bad that home loans are tax-deductable.'
b. Intended (unavailable): 'It is too bad that nothing else is tax-deductable.' (Yoshimura 2005 adapted from Horn 2002)

[4] In Horn (2002), the meaning contribution of the English *only* is the reverse of *dake*, i.e., it asserts the negative proposition and entails the affirmative proposition.

I equate the notion of 'entailment' in [9] and [7] to 'conventional implicature' in the sense of [10]. Hence, (14) is analyzed as having two independent meanings, one is an assertion 'John came' and the other is a conventional implicature 'no one else came'.

(14) JOHN-dake-ga kita.
John-only-Nom came.

a. Assertion: John came.
b. conventional implicature ('entailment' in Horn 2002 and Yoshimura 2005): No one else came.

Yoshimura's (2005) proposal is based on Horn's [9] assumption:

(15) Only the assertional content can be a complement of a higher functor.

Hence, in (12), only the affirmative proposition 'you can speak English' is embedded under the conditional. Similarly, (13) shows that the conventional implicature 'I cannot speak any other languages' takes scope outside of the *because*-clause.

To derive the correct interpretation for (5), I follow Yoshimura's (2005) proposal that the exhaustive meaning of Japanese *dake* contributes to a meaning level which is different from its prejacent proposition.

5 Scope of *Dake*

Given that the exceptive meaning of *dake* contributes to an expressive meaning, let us consider its meaning contribution itself. I propose here that *dake* takes different scopes depending on whether the sentence is Contrastive-marked or not. More specifically, *dake* normally takes the prejacent proposition as its argument and generates the exceptive meaning as a conventional implicature. When it is used in a Contrastive-marked sentence, *dake* takes scope over the assertion potential literal act [1] and implicates that the asserted proposition is the only assertion that the speaker makes with respect to the question under discussion.

5.1 Exhaustification over Propositions

We start with the case without Contrastive-marking. (14), repeated here as (16), has a LF structure as in (17).

(16) John-dake-ga kita.
John-dake-Nom came

a. Assertion: John came. $(= \alpha)$
b. conventional implicature: No one else came.

(17) [SpeechActP Assert [IP **dake** α]]

We obtain the conventional implicature 'No one else came' by negating all the alternative propositions $p \in Alt(\alpha)$ as depicted in (18).

(18) $\forall p[[p \in Alt(\alpha) \& p \neq \alpha] \rightarrow \neg p]$

5.2 Quantification at Non-propositional Levels

Now, let us turn to the case with Contrastive-marking. It has been observed that a Contrastive-marked element takes scope higher than the propositional level. In [11], I claimed that the use of *wa* introduces the operator CON that must be linked to an attitude holder.[5]

The claim is motivated by the following fact. The implicature induced by Contrastive *wa* can be relativized to an attitude-holder other than the speaker if *wa* is embedded under an attitude predicate:

(19) PETER-**wa** kita-to John-ga **shinjite**-iru
Peter-**Con** come-Comp John-Nom **believe**-Prog

a. Global: The speaker knows [John believes Peter came]
Implicature: The speaker doesn't know [whether John knows that Mary came]

b. Local: The speaker knows [John believes Peter came]
Implicature: John doesn't know whether Mary came

Also, [13] observe that in (20–a), *JOHN-dake* 'only John' with a nominative marker receives a narrow scope interpretation with respect to the attitude verb, while with the Contrastive marker *-wa* in (20–b), it receives wide-scope.

(20) a. JOHN-dake-**ga** kuru to omotte-ita.
John-only-Nom come Comp thought
'I thought that only John would come.' (thought > only)

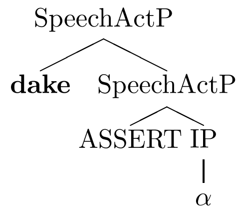
b. JOHN-dake-**wa** kuru to omotte-ita.
John-only-Con come Comp thought
'Only John, I thought that he would come.' (only > thought)

[13] claims that the *wa*-marked element serves as a link to the information expressed by the sentence. Hence, it takes wider scope with respect to everything else in the sentence. Given these observations, I assume here that Contrastive-marking forces *dake* to take scope higher than a mere propositional level. We start by considering two possibilities. One is quantification over speech acts and the other is quantification over potential literal acts.

Speech Acts. First, let us assume that if *dake* is used in a Contrastive-marked sentence, the sentence has a LF structure as in (21), and *dake* exhaustifies over assertion speech acts as formulated in (22).

[5] See also [12] who claims that contrastiveness operates on speech acts, not propositions.

(21)



(22) $\forall a[[a \in Alt(\text{ASSERT}(\alpha))\&a \neq \text{ASSERT}(\alpha)] \rightarrow \neg a]$

However, the computation in (22) causes a theoretical problem. Namely, (22) involves a negation over speech acts, which is not a valid operation in the computation of speech acts as argued by [14].

[14] considers speech acts as moves in conversational games in the sense of [15]. In other words, speech acts lead from one set of social commitments to another set. Given this assumption, [14] claims that the only operation involved in speech acts is conjunction.

[14] motivates his proposal by the fact that a pair-list reading of a *wh*-question is possible only with a universal quantifier. The pair-list reading of (23) is derived by universal quantification over the question act, which is possible since universal quantification is reduced to conjunction.

(23) Which dish did every guest make?
⇔For every guest x: Which dish did x make?
⇔Which dish did Al make, **and** which dish did Bill make, **and** which did Carl make?

On the other hand, other quantifiers like *most*, which involve disjunction, cannot operate over question acts; hence, fail to have a pair-list reading (24).

(24) #Which dish did most guests make?
⇔For most guests x: Which dish did x make?
⇔Which dish did Al make and which dish did Bill make, **or** which dish did Al make and which dish did Carl make, **or** which dish did Bill make and which dish did Carl make?

[14] gives the following explanation for why the only operation allowed for speech act computation is conjunction.

Conversational games are characterized by a set of states, and transitions between those states. If s is the current state in a conversational game, then the performance of an appropriate act A leads to a new state, s'.

Performing conjoined acts [A & A'](s) results in the union of the commitments that the consecutive acts of A(s) and A'(s) would have led to, namely A(s) ∪A(s'). For example, the resulting state of a consecutive utterance of acts in (25-a) is equivalent to the resulting state of the conjoined acts in (25-b). Hence, we can maintain the same semantic type of commitment states.

(25) a. Which dish did Al make? –The pasta.
Which dish did Bill make? –The salad.

b. Which dish did Al make? And which dish did Bill make?
Al (made) the pasta, and Bill the salad. [14]

On the other hand, a disjunction of A and A' at the state s would result in a **set** of commitment states, i.e., {A(s), A(s')}, which is of a higher type than the initial commitment state. Another operation of disjunction would result in an even higher type. Thus, [14] concludes that there is no simple way to disjoint question acts. To illustrate, if (26) were a case of disjoint questions, the addressee would have a choice of answering one of the questions. Thus, the addressee could choose the first question and answer *I have been to Sweden*, even if the addressee has been to both Sweden and Germany. According to [14], however, this is an incomplete answer. Rather, the questioner asks whether the addressee has been to Sweden or to Germany; and hence, it should be answered by *yes* or *no*.

(26) Have you ever been to Sweden or have you ever been to Germany? [14]

Krifka (2001) further argues that negation is not involved in the algebra of speech acts, since negation would allow us to derive disjunction from the combination of conjunction and negation by De Morgan's law ($\neg[A\&A'] = \neg A \cup \neg A'$).[6]

Going back to Japanese exhaustification, as we have seen in Section 4, the use of *dake* involves two commitments: a positive one (assertion) and a negative one (conventional implicature). If the exhaustification took scope over speech acts, then the negation would also take scope over speech acts. This is not a legal operation on speech acts, hence we need to seek for another object that correctly characterizes the intuition in (9) and can take scope under negation.

Potential Literal Acts (Siegel, To appear). [1] introduces a notion of *potential literal acts* in order to make a correct paraphrase of so-called *Biscuit Conditionals* (or *Relevance Conditionals*). Since [16], it has been noted that there exist conditional sentences whose consequences are not literally restricted by the the propositional content of the antecedents. For example, in (27), the presence of the pizza in the fridge does not depend on whether the addressee is hungry or not.

---

[6] Indeed, it is not possible to negate question act. In Hara (in progress), I argue that the use of *-wa* forces the exhaustification by *dake* to take place over question acts, and triggers negation of alternative acts, which is not a valid move in terms of conversational games as in (i–b).

(i) a. JOHN-dake-**ga** nani-o kai-mashi-ta-ka?
John-only-Nom what-Acc buy-Hon-Past-Q
'What did only John buy?'
b. *JOHN-dake-**wa** nani-o kai-mashi-ta-ka?
John-only-Con what-Acc buy-Hon-Past-Q

(ii) Intended Interpretation of (i-b)
a. As for John, what did he buy and
b. #It is not the case that as for other people, what did they buy?

(27) If you're hungry, there's pizza in the fridge. [1]

In the various literature [17,18,19,20,21,22,23] it has been suggested that Biscuit Conditionals restrict the speech act performed by the main clause as illustrated in (28).[7]

(28) If you're hungry, ASSERT (There's pizza in the fridge).

According to [1], however, this analysis leads us to an incorrect paraphrase for a Biscuit Conditional like (29).

(29) Whenever you get hungry, there's pizza in the fridge. (Chris Potts p.c. to [1])

If the conditions expressed by the antecedent of (29) actually applied to the speech act, (29) should be paraphrased as follows:

(30) at any time t at which you get hungry (PERFORMED ASSERTION) there's pizza in the fridge.

As [1] notes, "[t]he speaker certainly will not be performing the assertion at any time t at which the listener gets hungry." Hence, the speech act of assertion is not an appropriate object to be conditioned by the antecedent of a Biscuit Conditional.

Instead, [1] proposes that Biscuit Conditionals involve existential quantification over potential literal acts:

(31) [Potential literal acts] are abstract objects consisting only of propositional content and
whatever illocutionary force potential can be read directly from their morphosyntactic form, not necessarily the actual illocutionary act that might be performed. [1]

[1] also assumes that variables for potential literal acts (assertions, questions, commands, etc.) are introduced by a context-sensitive meaning-shift rule ((32)) in order to accommodate the cases where interpretation of a sentence does not converge.

(32) If B is a sentence of English with the morphosyntactic shape of an assertion and $^\wedge\beta$ is its translation,
then a **is an assertion of** $p \wedge p = {}^\wedge\beta$, is also a possible translation of B, where a varies over assertions, p varies over propositions, and **is an**

[7] There is another approach to this type of construction, i.e., Performative Hypothesis [24,25,26,27,28,29,30]. In this approach, (27) is understood as having an implicit performative predicate as in (i).

(i) If you're hungry, I say to you there's pizza in the fridge.

See [1] for discussion against Performative Hypothesis.

> **assertion of** is the relation between assertions and propositions such that if x is an assertion of y, then y is the propositional component of x. [1]

After the meaning-shift rule, existential closure applies to these variable, hence (27) is paraphrased as in (33).

(33) If you're hungry, there is a (relevant) assertion that there's pizza in the fridge. [1]

[1] argues for the necessity of potential literal acts by pointing out that her analysis can make an adequate paraphrase for (29) as in (34). Potential literal acts are not acts that are actually performed and they do not specify the contextual variables (speaker, addressee, etc.) of actual speech acts. Therefore, the analysis does not tell us that in (29), there is a performed act at each time when the addressee gets hungry. Rather, (29) expresses that there is an abstract potential literal act at each time when the addressee gets hungry.

(34) At any time t at which you get hungry, there is/will be a (relevant) assertion that there's pizza in the fridge.

In summary, potential literal acts are abstract semantic objects that contain propositional content and illocutionary force potential.

In the next section, I adopt this notion of potential literal acts and analyze *dake* in a Contrastive-marked sentence as a quantifier which operates over alternative potential literal acts.

5.3 Exhaustification over Potential Literal Acts

As mentioned in section 5.2, a Contrastive-marked element takes wide scope over the entire proposition. Accordingly, when *dake* is used in a Contrastive-marked sentence, it attempts to negate alternative speech acts. However, as argued by [14], negation cannot operate over speech acts, hence the meaning shift rule (32) applies. To illustrate, (5) (repeated here as (35)) is translated into a potential literal act as (36-a). In addition, the focus-marking on *John* generates an alternative potential literal act as in (36-b).

(35) JOHN-dake-wa kita.
John-dake-Con came.

(36) a. a is an assertion of $p \wedge$ p=came(John)
b. alternative potential literal act:
a' is an assertion of $p' \wedge p'$=came(Mary)

After the meaning-shift rule, the conventional implicature denoted by *dake* negates the generated alternatives as in (37) ('assertion(B(F))' is a shorthand for 'a is an assertion of $p \wedge p$=B(F)').

(37) $\forall a'[[a' \in Alt(\text{assertion}(\text{B}(\text{F})))\&a' \neq \text{assertion}(\text{B}(\text{F}))] \rightarrow \neg a']$

Given this, the interpretation of **dake**(B)(F) in a Contrastive-marked sentence is summarized as follows:

(38) Let F be the focus-marked elements, B the background
The interpretation of **dake**(B)(F)
  a. asserts: B(F)
  b. implicates: There is no assertion of individuals other than F with respect to the question B.

Now, remember that both Contrastive *wa* and *dake* are expressive items that induce conventional implicatures. As a consequence, when both *wa* and *dake* are used as in (5), two conventional implicatures are generated independently.

(39) JOHN-dake-wa came. (5)
  a. B= $\lambda x.\ x$ came. F=John
  b. assertion: John came.
  c. conventional implicature 1, Con(B)(F): The speaker considers the possibility that 'Mary came' is false.
  d. conventional implicature 2, **dake**(B)(F): There is no assertion of Mary with respect to the question $\lambda x.\ x$ came.

Let us go back to the contrast between *wa* and *dake-wa* repeated here as (40) and (41).

(40) Did John and Mary come?
  a. JOHN-wa kita. Mary-mo kita kamoshirenai.
  John-Con came. Mary-Add came might
  'At least John came. Mary might have come, too.'
  b. JOHN-wa kite, Mary-wa ko-nakat-ta.
  John-Con came. Mary-Con come-Neg-Past
  'At least John came, and Mary didn't come.'

Since in (40), *dake* is absent, the first sentence only generates conventional implicature 1, 'the speaker considers the possibility that 'Mary came' is false.', which is compatible with the subsequent sentences.

(41) Did John and Mary come?
  a. #JOHN-dake-wa kita. Mary-mo kita kamoshirenai.
  John-dake-Con came. Mary-Add came might
  'At least John came. Mary might have come, too.'
  b. #JOHN-dake-wa kite, Mary-wa ko-nakat-ta.
  John-dake-Con came. Mary-Con come-Neg-Past
  'At least John came, and Mary didn't come.'

On the other hand, in (41), *dake* generates an additional implicature 'There is no assertion about Mary with respect to the question $\lambda x.\ x$ came.' Hence, any continuation that asserts the speaker's knowledge of any other individuals with respect to the question under discussion turns out to be infelicitous.

6 Conclusion

The compositional contribution of the Japanese exhaustive particle *dake* within a Contrastive-marked sentence is puzzling if only one dimension of meaning is considered. If the exhaustive meaning denoted by *dake* contributed to an assertive level of meaning, the sentence would cause a presupposition failure, or compute an implicature which does not match the native speakers' intuitions. Instead, I propose that the use of *dake-wa* indicates the exhaustification over potential literal acts of assertion, rather than the exhaustification over propositions.

References

1. Siegel, M.: Biscuit conditionals: Quantification over potential literal acts. Linguistics and Philosophy (To appear)
2. Kuno, S.: The Structure of the Japanese Language. MIT Press, Cambridge, Mass (1973)
3. von Stechow, A.: Focusing and backgrounding operators. In Abraham, W., ed.: Discourse Particles. Benjamins, Amsterdam (1991) 37–84
4. Krifka, M.: For a structured meaning account of questions and answers. In Fery, C., Sternefeld, W., eds.: Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow. Akademie Verlag, Berlin (2001) 287–319
5. Hara, Y.: Contrastives and gricean principles. In Dekker, P., Franke, M., eds.: Fifteenth Amsterdam Colloquium, Universiteit van Amsterdam (2005) 101–106
6. Futagi, Y.: Japanese Focus Particles at the Syntax-Semantics Interface. PhD thesis, New Brunswick Rutgers, The State University of New Jersey (2004)
7. Yoshimura., K.: Only: Presuppose, entail or assert? Presented at LASSO (34th Annual Meeting of the Linguistic Association of the Southwest), (2005)
8. Kuno, S.: The syntax and semantics of the dake and sika constructions. In: Harvard Working Papers in Linguistics 7. (1999) 144–172
9. Horn, L.: Assertoric inertia and npi-licensing. In: Proceedings of CLS. (2002) 55.82
10. Potts, C.: The Logic of Conventional Implicatures. Oxford Studies in Theoretical Linguistics. Oxford University Press, Oxford (2005) [Revised 2003 UC Santa Cruz PhD thesis].
11. Hara, Y.: Grammar of Knowledge Representation: Japanese Discourse Items at Interfaces. PhD thesis, University of Delaware, Newark, DE (2006)
12. Tomioka, S.: Grammar of contrastiveness. A Lecture Note for 'Informationstruktur- Grammar of Contrastiveness', University of Potsdam (2006)
13. Portner, P., Yabushita, K.: The semantics and pragmatics of topic phrases. Linguistics and Philosophy **21**(2) (1998) 117–157
14. Krifka, M.: Quantifying into question acts. Natural Language Semantics **9** (2001) 1–40
15. Wittgenstein, L.: Philosophische Untersuchungen. Basil Blackwell, London (1958)
16. Austin, J.L.: Ifs and cans. In: Philosophical Papers. Oxford University Press, Oxford (1961)
17. Grice, H.P.: Logic and conversation. In Cole, P., Morgan, J., eds.: Syntax and Semantics. Volume 3: Speech Acts. Academic Press, New York (1975) 43–58

18. Dummet, M.: Frege: Philosophy of Language. Duckworth, London (1973)
19. Bach, K., Harnish, R.: Linguistic Communication and Speech Acts. MIT Press, Cambridge, MA (1979)
20. Horn, L.R.: A Natural History of Negation. University of Chicago Press, Chicago (1989) Reissued 2001 by CSLI.
21. Wakker, G.: Conditionals in the layered structure of functional grammar. In M. Fortescue, P.H., Kristoffersen, L., eds.: Layered Structure and Reference in a Functional Perspective. John Benjamins, Amsterdam (1992) 369–386
22. Geis, M.: Speech Acts and Conversational Interaction. Cambridge University Press, Cambridge (1995)
23. Bach, K.: The myth of conventional implicature. Linguistics and Philosophy **22**(4) (1999) 367–421
24. Ross, J.R.: Constraints on Variables in Syntax. PhD thesis, MIT, Cambridge, MA (1970)
25. Lakoff, G.: Linguistics and natural logic. In Davidson, D., Harmon, G., eds.: Semantics of Natural Language. Reidel, Dordrechtz (1972)
26. Sadock, J.: Towards a Linguistic Theory of Speech Acts. Academic Press, New York (1974)
27. Comrie, B.: Conditionals: A typology. In E. Traugott, A. ter Meulen, J.S.R., Ferguson, C., eds.: On Conditionals. Cambridge University Press, Cambridge (1986) 77–99
28. Van der Auwera, J.: Conditionals and speech acts. In E. C. Traugott, A. ter Meulen, J.S.R., Ferguson, C.A., eds.: On Conditionals. Cambridge University Press, Cambridge (1986)
29. Sweetser, E.: From Etymology to Pragmatics: Metaphorical and cultural aspects of semantic structure. Cambridge University Press, Cambridge (1990)
30. Iatridou, S.: Topics in Conditionals. PhD thesis, MIT, Cambridge, MA (1991)

Unembedded ‘Negative’ Quantifiers

Yukio Furukawa

Department of Linguistics, McGill University, 1085 Dr. Penfield, Montréal, Québec, H3A 1A7 Canada
yukio.furukawa@mail.mcgill.ca

Abstract. Unembedded ‘negative’ quantifiers, which are isolated in conversation initial position, provide evidence for the debate between the direct interpretation analysis of an isolated quantifier and the ellipsis analysis. In fact, unembedding of a Greek ‘negative’ quantifier *KANENAS* supports the former, while unembedding of a Russian ‘negative’ quantifier *ni-kogo* supports the latter.

1 Introduction: Unembedded Quantifiers

A widely held view about a quantifier phrase (QP) in natural language is that it is a function from a predicate to a truth value. However, natural language has several cases where it is less obvious that a QP is construed in such a way. Consider (1), where a QP is isolated in the conversation initial position. Following Stainton [19], let us call a quantifier isolated in such a way an unembedded quantifier. Despite the absence of an overt predicate, native speakers find that (1) meaningfully conveys a proposition such as ‘the unoccupied seat is reserved for an editor of Natural Language Semantic’. Assuming that meaningfulness is determined only in terms of propositions, one question arises: in what way is its propositional meaning obtained though it apparently lacks a predicate to be applied? Is its QP meaningful by itself?

(1) *situation: There are some empty seats around a table, and then, pointing at one, I say,*
An editor of Natural Language Semantics (Stainton [19]: 315)

There exist two opposing views in the philosophical literature. Stainton’s [19] view is that a QP is by itself meaningful enough to convey a proposition. Although (1) is not uttered as a sentence, it can be falsified, e.g. if the seat is reserved for Emmon Bach. On the other hand, Stanley’s [21] view is that a QP has no meaning in isolation. Although, apparently, it is not uttered as a sentence, this type of analysis would posit that its missing piece (perhaps a predicate) exists covertly. Stanley claims that the propositional meaning that native speakers find in (1) is not due to the meaningfulness of the unembedded QP, but rather is due to ellipsis.[1]

[1] Perhaps, ellipsis has two dominant views. In one view, ellipsis is phonological reduction that is claimed e.g. by Sag [17]. The other view is that ellipsis is semantic enrichment or recovery as claimed e.g. by Williams [24]. I am basically neutral to those views. Hence, I am using the term ‘ellipsis’ in a loose way; throughout this article (cf. Note 10), ellipsis means association of covert constituents in order for a fragment to be construed as a proposition.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 232 – 245, 2007.

To effectively argue that (1) is a piece of evidence for the ellipsis analysis (Stanley's analysis), we have to show that (1) really has an elided constituent. Contrary to, e.g., verb phrase (VP) ellipsis, however, (1) has no strong indicator of ellipsis such a stranded auxiliary. On the other hand, in order to argue that (1) is a piece of evidence for the direct interpretation analysis (Stainton's analysis), we have to show that (1) has no elided constituent. Once we admit that language has covert constituents, however, to show the absence of a covert constituent seems to be an especially difficult task; theories do not predict whether or not grammar constrains ellipsis strictly enough to prohibit its occurrence in (1) (see Ludlow [14]; cf. Lobeck [15]).

Given this dilemma, (1) does not seem to provide any decisive evidence for either one of the two analyses. I suggest that unembedded 'negative' quantifiers provide decisive evidence for the debate. In fact, in one case, the ellipsis analysis loses since postulating an elided constituent makes the sentence ungrammatical. On the other hand, in another case, the direct interpretation analysis loses since an unembedded 'negative' quantifier fails to convey its intended negative meaning without postulating an elided constituent.

2 Setups

Before going into the discussion of unembedded 'negative' quantifiers, we need a few assumptions about (i) the 'negative' QPs that I use in the discussion, (ii) the contexts behind the conversation initial utterances and (iii) anaphoricity. I will justify these assumptions in this section.

2.1 *KANENAS* in Greek and *ni-kogo* in Russian

The term, *'negative' quantifiers*, that I use throughout this article does not necessarily mean these quantifiers entail negation (therefore, I use quotation marks). I rather use the term to refer to a quantificational expression that participates in a negative sentence with a particular type of negative meaning, i.e., 'total negation'.

We know at least one thing about 'negative' quantifiers: unless there is double negation, the meaning of a sentence with a 'negative' quantifier is normally total negation. Total negation is a cover term to refer to the meanings expressed by the sentences in (2). The idea behind total negation is that the relation between the two relevant sets in (2) (i.e., the set of people and the set denoted by the predicate) is total (see also Lappin [13]). Thus, it is not connected to a particular logical form.

(2) a. For everyone, I did not see her/him.
b. It is not the case that I saw someone.
c. I saw no one.
d. I did not see even one person.
e. I saw zero people.

The following discussion focuses its attention on two quantificational expressions: *KANENAS*/*KANENAN* in Greek and *ni-kogo* in Russian.[2] Both of them indicate total

---

[2] *KANENAN* is the Accusative form of *KANENAS*. *Ni-kogo* is the Genitive form of *ni-kto*.

negation: neither (3a) nor (3b) allows the existence of people seen by the subject. Remarkably, both of them require clause-mate sentential negation, as shown in (3).[3]

(3) a. Greek (Giannakidou [6])
O Petros *(*dhen*) idhe **KANENAN**.
the Peter not saw.3sg n-person
'Peter did not see anyone.'
b. Russian (Watanabe [23]: 592)
Ja *(*ne*) videl **ni-kogo**.
I neg saw no one
'I saw nobody.'

While Giannakidou [5], [6] uses 'n-person' as the gloss of *KANENAS*, Watanabe [23] uses 'nobody' as the gloss of *ni-kogo* (and *KANENAS*). In terms of negativity, their claims are in competition with one another. Giannakidou claims that *KANENAS* is not a negative quantifier but a (nonnegative) universal quantifier corresponding to 'everybody',[4] while Watanabe claims that both *KANENAS* and *ni-kogo* are negative quantifiers. It seems to me, however, that the debate as to whether or not *KANENAS* and *ni-kogo* are negative has not yet been settled.[5] Although I reproduced the examples in (3) with no modification, this does not mean that I assume that *KANENAS* and *ni-kogo* are expressions corresponding to a person n-word or 'nobody'. (Moreover, I claim later that *ni-kogo* should not be glossed as 'nobody'.)[6]

[3] The clause-mate sentential negation is obligatory even if they are located in the subject positions (see (ia-b)). This is different from e.g. an Italian negative quantifier *nessuno*. The sentential negation is absent if *nessuno* occupies the subject position, as shown in (ic).

(i) a. KANENAS *(dhen) ipe TIPOTA.
n-person not said.3sg n-thing
'Nobody said anything.' (Giannakidou [5]: 461)
b. Ni-kto *(ne) zvonil.
No one neg called
'No one called.' (Watanabe [23]: 592)
c. *Nessuno* (*non) ha visto Mario.
nobody neg has seen Mario
'Nobody has seen Mario.' (Ladusaw [12]: 216)

[4] Giannakidou [5] claims that the total negation in (3a) is due to the obligatory wide scope taking nature of *KANENAS* (over negation).

[5] Watanabe argues against Giannakidou's claim, observing that *KANENAS* can be isolated as a fragmental answer. He claims that, if it is assumed that a fragmental answer involves ellipsis whose elided constituent finds a morph-syntactic antecedent in its question, its usage as a fragmental answer provides counterevidence to Giannakidou's claim; since a nonnegative question has no syntactic constituent that contains sentential negation, negation must be located in (isolated) *KANENAS*. Otherwise, its fragment fails to express its intended negative meaning. However, whether or not his assumption (i.e., a fragmental answer involves ellipsis whose elided constituent finds a syntactic antecedent in the question) is warranted seems less clear, and hence, I do not think that he provides a knockdown argument against Giannakidou.

[6] To avoid any confusion, I do not put any gloss on *KANENAS*, *ni-kogo* and *dare-mo* in the examples that belong to me.

2.2 Context

I crucially use two situations as contexts for the conversation initial utterances. In both situations, the sentences with *KANENAS* and *ni-kogo*, whose meanings are similar to 'there is nobody', are felicitously uttered, as shown in (4) and (5). These data suggest that neither the propositional content of *dhen iparxei KANENAS* nor that of *ni-kogo net* induces infelicity in the two contexts.

(4) *situation: Mary is a TA. Since she had to teach today, she went to the classroom. Since it was her first tutorial, she expected every student's participation. When she opened the door, however, she found that no one/student was there. Then, she said,*

Greek: a. Dhen iparxei **KANENAS**!
neg exist.3sg

Russian: b. **Ni-kogo** net!
neg

English: c. There is nobody!

Japanese: d. (Dare-mo) i-nai!
exist-neg

(5) *situation: Basically the same as the situation in (4), but, this time, her expectation was weak since it happened in the middle of the semester. She expected that not every student would be present in her tutorial, in other words, she expected some student's absence.*

Greek: a. Dhen iparxei **KANENAS**!

Russian: b. **Ni-kogo** net!

English: c. There is nobody!

Japanese: d. (Dare-mo) i-nai!

Note that, the crucial difference between the context of (4) and the context of (5) is that only in the latter is the inference 'there is at least one person/student such that s/he is not present' available.

2.3 Anaphoricity

If an unembedded quantifier involves ellipsis, I assume that its elided constituent has to find an antecedent, as is generally considered to be required. For example, an elided VP in (6b) must find a VP in the preceding text (6a), i.e. *believe that the earth is larger than it is*, and take it as an antecedent.

(6) a. John believes that the earth is larger than it is.
b. … but Joan doesn't. (Hankamer & Sag [9]: 419)

Since unembedded quantifiers are conversation initial, however, no preceding text exists. How, then, can an elided constituent within a conversation initial utterance find an antecedent? In order to account for the meaningfulness of an unembedded quantifier, I appeal to the notion of deep anaphora by Hankamer & Sag [9] though this is a descriptive term. I assume that the elided constituent of an unembedded quantifier is a kind of deep anaphora, which finds its antecedent in the previous context.

Note that, I assume that any elided constituent of an unembedded quantifier would be a predicate, following the inventory of the generalized quantifier (see Barwise & Cooper [1]), but it is normally impossible to define a predicate/property in the context: perhaps, either individuals or propositions exist in the context (see Stanley & Szabó [22]). To avoid any conflict about recoverability, I simply assume that its elided predicate finds an antecedent predicate through lambda abstraction of a proposition in the context.

We also need two more assumptions about negation and anaphoricity. I assume that no covert sentential negation exists in natural language. If natural language had a covert sentential negation, negative sentences would be phonologically indistinguishable from nonnegative sentences. At the same time, due to identity between an anaphoric element and its antecedent, I assume that a nonnegative antecedent does not recover the content of a negative anaphora. As suggested by its infelicity, it is implausible to think that (7b) has the same content as that of (6b) by recovering *not believe that the earth is larger than it is* in its elided VP.

(7) a. John believes that the earth is larger than it is.
    b. #… but Joan does.

3 Unembedded *KANENAS* and *ni-kogo*

As shown in (8), both *KANENAS* and *ni-kogo* can be unembedded.[7] According to my informants, these unembedded quantificational expressions are meaningful, and their meanings are equivalent to the meanings of their fully sentential counterparts in (5). For the moment, let us assume that the propositional contents of unembedded *KANENAS* and *ni-kogo* are equivalent to the propositional contents of their fully sentential counterparts, i.e., *dhen iparxei KANENAS* and *ni-kogo net*, respectively.

(8) *situation: the same as (5)*
    Greek: a. KANENAS!
    Russian: b. Ni-kogo!
    English: c. Nobody!
    Japanese: d. #Dare-mo!

However, it is not the case that such unembedding is always possible. While *KANENAS* can be unembedded under the context of (4), *ni-kogo* cannot. Note that, if the propositional meaning intended by (9b) is equivalent to the content of *ni-kogo net*, infelicity in (9b) is not due to its propositional content, as observed in (4b).

---

7 According to the studies of negative concord, both *KANENAS* and *ni-kogo* can be isolated as fragmental answers, too (see Giannakidou [5], [6]; Watanabe [23]; see also Note 6).

(9) *situation: the same as (4)*
Greek: a. KANENAS!
Russian: b. #Ni-kogo![8]
English: c. Nobody!
Japanese: d. #Dare-mo!

Interestingly, if unembedded *ni-kogo* is uttered as a polar (or a yes/no) question, it becomes felicitous not only under the context of (5) but also under the context of (4), as shown in (10a) and (11a).

(10) *situation: the same as (5)*

Russian:	a. Ni-kogo?	a'. Ni-kogo net?
English:	b. Nobody?	b'. Is there nobody/is nobody here?
Japanese:	c. Dare-mo?	c'. (Dare-mo) i-nai no?

(11) *situation: the same as (4)*

Russian:	a. Ni-kogo?	a'. Ni-kogo net?
English:	b. Nobody?	b'. Is there nobody/is nobody here?
Japanese:	c. #/??Dare-mo?	c'. (Dare-mo) i-nai no?

Then, why is (9b) infelicitous? The following section addresses unembedded *ni-kogo*, first.

3.1 Unembedded *ni-kogo*

The direct interpretation analysis cannot predict the infelicity of (9b). Due to the felicity and meaningfulness of (8b), let us suppose that unembedded *ni-kogo* conveys a proposition whose content is equivalent to the content of its fully sentential counterpart *ni-kogo net*. Felicity in (9b) is wrongly predicted since, as shown in (4b), this content never induces infelicity under the context of (4).

Can the ellipsis analysis predict its infelicity? If it can, then what is the source of the infelicity in (9b)? Moreover, does unembedded *ni-kogo* have any reason for the presence of an elided constituent?

Consider what happens if unembedded *ni-kogo* involves ellipsis. As a general convention on ellipsis, the elided predicate in unembedded *ni-kogo* must find an antecedent. Then, what kind of antecedent is available under the contexts of (4) and (5)? As mentioned in Section 2.3, only a negative antecedent can recover the content of a negative anaphora. Since the context of (5) provides a negative inference to the effect of 'there is a person/student x such that s/he is not present', the negative predicate that can be an antecedent of the elide predicate is defined via lambda abstraction of this proposition. Hence, felicity in (8b) and (10a) is predictable.

[8] The judgments about (9b) are very subtle, and one of my informants finds no infelicity. In fact, there exists a difficulty in asking its judgment: it seems hard to show that those who find it felicitous consider it as an isolated QP or as shorthand (see Stainton [20]; Stanley [21]).

Next, consider the context in (4), which is about Mary's expectation that everybody (or every student) is present. Since her expectation implies that every member of the relevant students is present, it may not imply any negative proposition such as 'there is a person/student x such that s/he is not present'. Thus, only nonnegative antecedents are available under the context. However, a conflict arises between what is recoverable from the antecedent and what must be recovered as the elided predicate of unembedded *ni-kogo*. As observed in (1), *ni-kogo* requires the clause-mate sentential negation, if it occurs in a sentence. If underlyingly (9b) and (11a) have fully sentential forms by the association of elided predicates, then their elided predicates must contain sentential negation. If the elided predicate in (9b) and (11a) finds a nonnegative predicate in the context of (4), the recovered sentences necessarily violate this clause-mate condition. Therefore, it is impossible to recover the content of the elided predicates in (9b) and (11a) from the context of (4).

At this intermediate stage, the ellipsis analysis can correctly predict infelicity of (9b) but may wrongly predict infelicity of (11a).[9] Can the ellipsis analysis manage to predict felicity of (11a)? I claim that it can. (11a) is a polar question; I suggest that a key to the solution lies in the semantics of such polar questions. It is widely assumed that a polar question denotes a set consisting of its yes-answer and its no-answer (see (12a)). Moreover, since normally its no-answer is defined through negating its yes-answer and vice-versa, I propose (12b) as the defining condition on its semantic object (12a). Then, suppose that a polar question has two defining conditions, namely defining its morph-syntactic object and defining its semantic object. I suspect that the definability of its semantic object is relevant to (11a).[10]

(12) a. $?p = \{p=1, \neg p=1\}$ where ? is a polar question operator and p is a proposition.
b. A polar question $?p = \{p=1, \neg p=1\}$ is defined if and only if either $p=1$ or $\neg p=1$ is defined.

Then, how is the set denoted by the polar question (10a) and (11a) defined? According to Haspelmath [10], the *ni*-prefix of *ni-kogo* is a scalar focus item, similar to 'even'.[11] Furthermore, Guerzoni [7] nicely observes that the set denoted by a polar

[9] One might think that the direct interpretation of *ni-kogo* takes place in the case of a polar question; if (11a) is by itself interpreted as a polar question equivalent to (11a'), its felicity and meaningfulness are predicted. As concluded earlier, however, we have a reason to believe that its direct interpretation is absent in assertion (or in non-question). Given that, it seems hard to restrict its direct interpretation to questions (especially if the semantics of a polar question is defined as (12a)). Hence, I do not think that felicity in (11a) provides evidence in support of the direct interpretation analysis.

[10] In order for (12b) to go through, I have to assume that ellipsis involved here is not phonological reduction (see Sag [17]), but semantic recovering (see Williams [24]).

[11] According to Haspelmath [10], Russian has several scalar focus particles, and the *ni*-prefix of *ni-kogo* is one of them. Interestingly, the *ni*-prefix is legitimate only in negative environments.

question containing a scalar focus item has to be defined in two ways, depending on the scope relations between the scalar focus item and the polar question operator.[12]

Following these observations, I assume that the set denoted by the question (10a) and (11a) is defined either as the set (13a) or as the set (13b). Note that, in (13), I use P to illustrate the elided predicate. For presentational convenience, suppose for the time being that P is λx.not.being.present(x). Then, ¬[[*ni*]]([[*kogo*]]([[P]]))=1 in (13a) and [[*ni*]](¬[[*kogo*]]([[P]]))=1 in (13b) roughly mean 'the proposition that it is not the case that even one person is such that s/he is not present is true' and 'the proposition that even one person is such that it is not that case that s/he is not present is true', respectively.[13]

(13) a. {[[*ni*]]([[*kogo*]]([[P]]))=1, ¬[[*ni*]]([[*kogo*]]([[P]]))=1} (for Q>*ni*)
b. {[[*ni*]]([[*kogo*]]([[P]]))=1, [[*ni*]](¬[[*kogo*]]([[P]]))=1} (for *ni*>Q)

The content of the proposition [[*kogo*]]([[P]])=1 is recoverable from the context of (5) but not from the context of (4): only the context of (5) has a negative inference 'there is a person x such that s/he is not present'. Hence, both [[*ni*]]([[*kogo*]]([[P]]))=1 and ¬[[*ni*]]([[*kogo*]]([[P]]))=1 are definable under the context of (5). By (12b), the questions (13a) and (13b) are both definable. Therefore, (10a) is felicitous and meaningful.

On the other hand, the content of the proposition ¬[[*kogo*]]([[P]])=1 is recoverable from the context of (4): it roughly means 'it is not the case that x is not present', and this is equivalent to 'x is present'. Clearly, this proposition is available in the context of (4). Since [[*ni*]](¬[[*kogo*]]([[P]]))=1 is definable, the question (13b) is definable due to (12b). The felicity and meaningfulness of (11a) follow from the definability of (13b).

We observed that the ellipsis analysis but not the direct interpretation analysis makes the right predictions in the felicity and meaningfulness of unembedded *ni-kogo*. However, I have not presented yet any evidence that indicates the presence of an elided constituent in unembedded *ni-kogo*. I suggest that the data on unembedded *ni-kogo* in (8)-(11) also show that this construction really has an elided constituent.

Consider the clause-mate condition observed in (1b). If the sentential negation is necessary for the purpose of syntactic licensing such as agreement, more specifically,

[12] It is well-known that a polar question with a scalar focus item sometimes induces and sometimes does not induce a bias toward its negative answer. Guerzoni observes that the presence and absence of bias are reducible to the scope relations between the scalar focus item and the question operator. Assuming that a proposition associated with *even* presupposes that it is the most unlikely proposition, what is presupposed by the yes-answer is contradictory to what is presupposed by the no-answer if *even* takes wider scope than the polar question operator. Since, in this case, both p=1 and ¬p=1 are presupposed as the most unlikely proposition, one of them is necessarily false. Hence, the question has a bias to the answer that has a true presupposition. On the other hand, if *even* takes narrower scope than the question operator, the question has no bias: since what is presupposed by the yes-answer is the same as what is presupposed by the no-answer, both the yes-answer and the no-answer are valid as long as the presupposition is true.

[13] I assume that *kogo* is a general person indefinite pronoun 'one'. Note that, according to Haspelmath [10], *kogo* is also used as a person wh-phrase.

if *ni-kogo* must be syntactically licensed by the sentential negation, the acceptability of (11a) is not predicted. As observed above, if its elided constituent contains negation, the constituent's content is not recoverable from the context of (4). Nevertheless, the native speakers find that (11a) is acceptable.

Why does *ni-kogo* need a clause-mate sentential negation? I claim that sentential negation is semantically necessary in (1b) to convey the intended negative meaning. Why is this the case? I claim that sentential negation is necessary because *ni-kogo* does not entail negation. If *ni-kogo* is negative, felicity of (9b) is wrongly predicted; since the necessity of the sentential negation is not for a syntactic reason, recovering a nonnegative predicate in its elided constituent does not result in a violation of a syntactic condition. Moreover, (9b) would not fail to convey its intended negative meaning since *ni-kogo* could be a provider of negativity.

Because (9b) is infelicitous, I claim that *ni-kogo* does not entail negation.[14] Since *ni-kogo* does not entail negation, negativity must be supplied from an independent source. Assuming that natural language has no covert sentential negation, the only thing that can provide negativity in unembedded *ni-kogo* is its elided constituent. I conclude that the presence of an elided constituent is indicated by the fact that negation belongs to the elided constituent, which is the only possible provider of the negativity for the intended negative meaning by unembedded *ni-kogo*.

3.2 Unembedded *KANENAS*

Now, consider unembedded *KANENAS*. Contrary to *ni-kogo*, *KANENAS* can be unembedded under the context of (4), as shown in (9a). The felicity of (9a) presents a problem for the ellipsis analysis; I will outline this problem below, and show that the direct interpretation analysis better predicts the felicity judgments for *KANENAS*.

Suppose for a moment that the propositional meaning of unembedded *KANENAS* is due to ellipsis. As observed in the previous section, the context of (4) does not recover the content of an elided constituent containing negation; from the context of (4), a nonnegative predicate is necessarily recovered in the elided predicate of (9a). Once such a nonnegative predicate is recovered, however, the sentence becomes ungrammatical because *KANEANS* cannot be a clause-mate of the sentential negation, despite the clause-mate condition observed in (1a). Given that, the elided predicate in (9a) necessarily fails to find an antecedent, contrary to its felicity and meaningfulness. Therefore, the ellipsis analysis fails to predict its felicity.[15]

Can the direct interpretation analysis correctly predict the felicity of (9a)? According to Haspelmath [10], *enas* of *KANENAS* is a general person indefinite

[14] Although the *ni*-prefix of *ni-kogo* apparently contains a negative morphology *n* (cf. Haspelmath [10]), I do not assume that *n* is a morphological realization of entailed negation.

[15] As far as I can see, the only case where the ellipsis analysis works to account for the felicity in (9a) is that *KANENAS* entails negation, and at the same time, it needs a clause-mate sentential negation for a semantic reason. However, these two demands may not be satisfied simultaneously: as long as *KANENAS* entails negation, it can be a provider of negativity, and hence it is implausible to assume that it must have a clause-mate sentential negation for a semantic reason. Moreover, I do not find cases where it can be in a sentence without negation. Therefore, I do not think that the ellipsis analysis has a plausible account for the felicity in (9a).

pronoun corresponding to 'one'. Then, it is not unreasonable to assume that, e.g. just like *everyone* in English, *KANENAS* has the person property. If the meaning intended by (9a) is described as a negative existential proposition (14a), it may not be so difficult to claim that *KANENAS* is by itself interpreted as such a proposition by claiming that *KAN* is decomposed as negation and an existential quantifier.[16]

(14) a. $\neg\exists x.\text{one}(x)$
b. $\forall x[\text{one}(x) \rightarrow Q(x)]$

On the other hand, if the intended meaning of (9a) is described as a universal negative proposition (14b), it is difficult to claim that *KANENAS* is by itself interpreted as such a proposition.[17] To describe its intended meaning in terms of a universal proposition, *KANENAS* must have at least two predicates; in addition to $\lambda x.\text{one}(x)$, it must entail another predicate $\lambda y.Q(y)$. Since the direct interpretation analysis does not assume any supplement, $\lambda y.Q(y)$ must be found within *KANENAS*.[18] However, it is not obvious that *KANENAS* has such a predicate.

What, then, is the description of the meaning of (9a)? Through the observation in (15), Giannakidou [5] claims that *KANENAS* by itself has a universally quantificational force: *KANENAN* (the Accusative form of *KANENAS*) accommodates modification by an absolutely-type adverb, which often diagnoses universality of a modified noun phrase. If her claim is correct, the problem of how to find $\lambda y.Q(y)$ within *KANENAS* remains.[19] Since it seems hard to test absolutely modification in unembedded quantifiers, I leave further issues about unembedded *KANENAS* as future research.

(15) a. Kim didn't see (*absolutely) anyone. (Horn [11]: 160)
b. Dhen idha (apolitos) KANENAN (apolitos).
neg saw.1sg absolutely n-person absolutely
'I saw absolutely nobody.' (Giannakidou [5]: 472)

---

16 According to Haspelmath [10], historically, *kan* comes from kaì eán 'even if'.

17 Since (i) roughly means 'for everything, it is not a person', it is not the intended meaning by (9a).
(i) $\forall x.\text{not.being.one}(x)$

18 One might think that, if unembedded *KANENAS* conveys a universal negative proposition (14b), the syntactic ellipsis analysis provides more convincing account for its meaningfulness, for example, by assuming that a covert predicate such as $\lambda x.\text{not.being.present}(x)$ is associated with it. As mentioned earlier, however, the ellipsis analysis faces a problem in felicity of (9a), in that it is impossible to recover the content of an elided constituent containing negation.

19 In fact, *KANENAS* has another intriguing aspect: it has a nonemphatic counterpart *kanenas*. According to Giannakidou [5], they are completely different from each other. For example, *KANENAS* exhibits similar behaviors to those of n-words or negative concord items. On the other hand, *kanenas* never exhibits negative concord, but shows similar properties to negative polarity items (and its isolation is not allowed). Note that, in Greek, the emphatic/nonemphatic paradigm is not restricted to the person 'negative' quantifiers: it is observed in 'negative' quantifiers with other restrictions such as the thing-restriction (*TIPOTA*/*tipota*), the place-restriction (*PUTHENA*/*puthena*) and so on (see Giannakidou [5]). Since, at this stage, the role of the emphatic/nonemphatic distinction plays is not clear, I postpone my proposal about the semantics of *KANENAS*.

4 Conclusion and Remaining Issues

Through the observations regarding felicity and meaningfulness of unembedded *KANENAS* and *ni-kogo*, this article claimed that the propositional meaning of unembedded *ni-kogo* is due to ellipsis, since unembedded *ni-kogo* fails to convey its intended negative meaning without postulating an elided constituent. Furthermore, this article also claimed that unembedded *KANENAS* is by itself interpreted as a proposition, since postulating an elided constituent makes its underlying sentence ungrammatical.

However, at least two problems are remaining. One problem, which has already been mentioned in Section 3.2, is about the description of the meaning of unembedded *KANENAS*. The other problem is about unembedded *ni-kogo* as a polar question. When Section 3.1 claimed that *ni-kogo* needs a clause-mate sentential negation for a semantic reason, it appealed to the failure to define (11a) morph-syntactically. This might have a strong prediction: if its polar question is semantically definable, *ni-kogo* does not have to co-occur with the sentential negation. It seems to me, however, that this prediction is not borne out. As far as I can see, *ni-kogo* must co-occur with sentential negation even in a question. I leave these two issues as future research.

Acknowledgments

Materials discussed in this paper were partly presented at the 7th Seoul International Conference on Generative Grammar held at Konkuk University, Seoul in August 2005 and as an electronic poster for Polarity from Different Perspectives held at New York University in March 2005. I would like to thank Anastasia Giannakidou, Brendan Gillon, Kyle Johnson, Masakazu Kuno, Jason Merchant, Jon Nissenbaum, Michel Paradis, Miyuki Sawada, Bernhard Schwarz, Junko Shimoyama and Akira Watanabe for helpful comments and suggestions. I am also grateful to Christina Manouilidou and Athanasios Tsiamas for the discussion of Greek data, and Alyona Belikova and Larissa Nossalik for the discussion of Russian data. Any remaining errors belong to me.

References

1. Barwise, J., Cooper, R.: Generalized quantifiers and natural language. Linguistics and Philosophy **4** (1981) 159-219
2. Elugardo, R., Stainton, R. J.: Shorthand, syntactic ellipsis and the pragmatic determinants of 'What is said'. Mind & Language **19** (2004) 442-471
3. Elugardo, R., Stainton, R. J.: Ellipsis and Nonsentential Speech. Springer, Dordrecht (2005)
4. Furukawa, Y.: Indeterminates+MO is neither an NPI nor an NCI. In press in Proceedings of the Workshop on Altaic Formal Linguistics 2. MIT Working Papers in Linguistics, Cambridge, MA (2006) 157-172
5. Giannakidou, A.: Negative … Concord? Natural Language and Linguistic Theory **18** (2000) 457-523

6. Giannakidou, A.: N-words and negative concord. In Everaert, M., van Riemsdijk, H., eds.: The Blackwell Companion to Syntax, Volume III. Blackwell, Oxford (2005)
7. Guerzoni, E.: Even-NPIs in yes/no questions. Natural Language Semantics **12** (2004) 319-343
8. Hamblin, C. L.: Questions in Montague Grammar. Foundations of Language **10** (1973) 41-53
9. Hankamer, J., Sag, I.: Deep and surface anaphora. Linguistics Inquiry **7** (1976) 391-426
10. Haspelmath, M.: Indefinite Pronouns. Oxford University Press, Oxford (1997)
11. Horn, L. R.: Pick a Theory (Not Just *Any* Theory): Indiscriminative and the Free-Choice Indefinite. In Horn, L. R., Kato, Y., eds.: Negation and Polarity. Oxford University Press, Oxford (2000) 147-192
12. Ladusaw, W. A.: Configurational expression of negation. In van der Does, J., van Eijck, J., eds.: Quantifiers, Logic and Language. CSLI Publications, Stanford, CA (1995) 203-223
13. Lappin, S.: Generalized quantifiers, exception phrases and logicality. Journal of Semantics **13** (1996) 197-220
14. Ludlow, P.: A note on alleged cases of nonsentential assertion. In Elugardo & Stainton [3]. (2005) 95-108
15. Lobeck, A.: Ellipsis. Oxford University Press, Oxford (1995)
16. Merchant, J.: Fragments and Ellipsis. Linguistics and Philosophy **27** (2004) 661-738
17. Sag, I.: Deletion and Logical Form. Doctoral dissertation, MIT. (1976)
18. Shimoyama, J.: Wide Scope Universal NPIs in Japanese. Paper presented at Georgetown University Round Table 2004 (2004)
19. Stainton, R. J.: Quantifier phrases, meaningfulness 'in isolation' and ellipsis. Linguistics and Philosophy **21** (1998) 311-340
20. Stainton, R. J.: In defense of nonsentential assertion. In Szabó, Z., ed.: Semantics vs. Pragmatics. Oxford University Press, Oxford (2004) 383-457
21. Stanley, J. Context and logical form. Linguistics and Philosophy **23** (2000) 391-433
22. Stanley, J., Szabó, Z.: On quantifier domain restriction. Mind & Language **15** (2000) 219-261
23. Watanabe, A.: The genesis of negative concord. Linguistic Inquiry **35** (2004) 559-612
24. Williams, E.: Discourse and logical form. Linguistic Inquiry **8** (1977) 101-139
25. van der Wouden, T.: Negative Contexts. Routledge, London (1997)

Appendix. Unembedded *dare-mo*

Japanese has an apparently similar expression to *KANENAS* and *ni-kogo*, i.e. *dare-mo*. In a sentence, *dare-mo* has to be a clause-mate of the sentential negation, and a sentence with *dare-mo* has a meaning of total negation, as shown in (16).

(16) boku-wa *dare-mo* mi-*(nakat-)ta.
I-top see-neg-past
'For all people, I did not see them.'

In the paradigmatic examples of unembedded 'negative' QPs above, I also put the cases of unembedded *dare-mo*. This is not only to display varieties of felicity, but also to note intriguing behavior of unembedded *dare-mo*. As observed in (8)-(11), contrary to unembedded *KANENAS* and *ni-kogo*, unembedded *dare-mo* is felicitous only in one of the four cases; the case where it is uttered as a question under the context of (5). Then, why does unembedded *dare-mo* exhibit such a limited felicity?

First of all, (8d) and (9d) appear to suggest that unembedding of *dare-mo* is impossible unless it is a question. However, this is not the case. As shown in (17c), its non-question unembedding is possible under the context of (17). Note that, the context of (17) crucially differs from the contexts of (4) and (5) in that Mary's expectation was everyone's absence.

(17) *situation: basically the same as (5), but, this time, her expectation was no student's participation, since, at the end of the semester, students were too busy to attend her tutorial.*
Russian: a. Ni-kogo. a'. Ni-kogo net.
English: b. Nobody. b'. There is nobody.
Japanese: c. Dare-mo. c'. (Dare-mo) i-nai.

How is the meaning of unembedded *dare-mo* in (17c) described? Shimoyama [18] nicely observes that total negation with *dare-mo* is described as a wide scope universal reading over negation. [20] Due to her observation, I assume that the propositional content of the sentence (17c'), *dare-mo i-nai*, is a universal negative proposition (18). Furthermore, just as I did in the cases of unembedded *KANENAS* and *ni-kogo*, let us assume that the propositional meaning intended by unembedded *dare-mo* is equivalent to the meaning of this sentence. Given these, the intended meaning of (17c) is described as (18).

(18) $\forall x[\text{student}(x) \rightarrow \neg\text{present}(x)]$

The direct interpretation analysis does not predict the limited felicity of unembedded *dare-mo*: despite the felicity of the fully sentential utterances, only (10c) and (17c) are felicitous. Then, why is unembedded *dare-mo* felicitous under the context of (17) but not in the other contexts? The answer seems to lie in the difference between the three contexts. As mentioned earlier, the context of (17) crucially differs from the other two contexts in that the inference 'for everyone, s/he was not present' is available only in the context of (17). I suspect that this is the source of felicity in (17c). More precisely, I suspect that the elided constituent in unembedded *dare-mo* is not a predicate, but a proposition only recoverable from the context of (17), i.e. the universal negative proposition (18).

The contrast between (10c) and (11c) supports the proposition recovering analysis. Suppose that the polar question (10c) and (11c) denotes the set (19). Since its no-answer (i.e. $\neg\forall x[\text{student}(x) \rightarrow \neg\text{present}(x)]$) is not recoverable from the context of (4) but from the context of (5), only (10c) is felicitous.[21]

(19) $\{\forall x[\text{student}(x) \rightarrow \neg\text{present}(x)], \neg\forall x[\text{student}(x) \rightarrow \neg\text{present}(x)]\}$

[20] Due to the space limitation, I do not present details of the discussions in Shimoyama [18], and I refer readers to Shimoyama [18] and Furukawa [4].

[21] However, there is a problem in (11c): $\neg\forall x[\text{person}(x) \rightarrow \neg\text{present}(x)]$ is true whenever $\forall x[\text{person}(x) \rightarrow \text{present}(x)]$ is true. I have to prevent $\forall x[\text{person}(x) \rightarrow \text{present}(x)]$ from being an antecedent for the no-answer in (19). I do not have any satisfactory explanation, and I leave this for future research.

To claim that what is elided in (17c) is the proposition (18), I have to address one question: what kind of role does *dare-mo* have? The data about unembedding in (8)-(11) and (17) suggest that *dare-mo* has no universally quantificational force. If it were a universal QP corresponding to ‘everyone’, felicity in (8d) would be wrongly predicted; since the context of (5) clearly has a proposition, namely ‘(there is a) y (such that it) is not present’, the predicate λy.not.being.present(y) is recoverable from the context. Felicity in (8d) would be wrongly predicted, too, if it were considered as a universal determiner ‘every’. The context of (5) has two propositions, namely, ‘(there is an) x (such that it) is a student’ and ‘(there is a) y (such that it) is not present’, and as a result, two predicates, λx.student(x) and λy.not.being.present(y), are recoverable from the context.

In Furukawa [4], I suggest that *dare-mo* is not a QP but a selector of a particular type of proposition, namely $\forall\alpha\neg\beta$.[22] The nature of *dare-mo* as a selector is consistent with the data from unembedding: since *dare-mo* in (17c) is not an unembedded QP but an unembedded selector, its elided constituent is a universal negative proposition (18). Unless unembedded *dare-mo* is a question, the content of its elided proposition is recoverable from no other context but the context of (17).

[22] Due to the space limitation, I do not present details of the discussions in Furukawa [4], and I simply refer readers to Furukawa [4].

The Fourth Workshop on Learning with Logics and Logics for Learning (LLLL2006)

Akihiro Yamamoto[1], Kouichi Hirata[2], and Ken Satoh[3]

[1] Graduate School of Informatics, Kyoto University, Japan
akihiro@i.kyoto-u.ac.jp
[2] Department of Artificial Intelligence, Kyushu Institute of Technology, Japan
hirata@ai.kyutech.ac.jp
[3] National Institute of Informatics and Sokendai, Japan
ksatoh@nii.ac.jp

1 The Workshop

The Fourth Workshop on Learning with Logics and Logics for Learning (LLLL2006) was held at June 5 and 6, 2006 in Tokyo, as one of the co-located workshops with the 20th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2006), and also as a meeting of the series SIG-FPAI of JSAI.

The workshop is proposed to bring together researchers who are interested in both of the areas of machine learning and computational logic, and to have intensive discussions on various relations between the two with making their interchange more active. We called for papers which are concerned with the following topics (but not exclusive): learning and knowledge discovery using logics, algorithmic aspects of learning based on logics, logics for machine learning and knowledge discovery, machine learning as a foundation of mathematics/mathematical procedures, amalgamation of logic-based learning and statistical/information theoretical learning, and learning and knowledge discovery from relational data, structured/semi-structured data or real-valued data. Also we organized the program committee for the workshop consisting of 16 excellent researchers to the area in logic and/or learning.

Furthermore, the workshop included two invited talks by Prof. György Turán and Dr. Tamás Horváth. By reviewing every submitted paper by three PC members, we selected 11 papers for the contributed talks, and arranged them into the following 4 tracks in the program:

1. Logical Foundations of Inductive Inference: 3 talks.
2. Learning with Logical Formulae: 2 talks.
3. Knowledge Discovery from Structured Data: 3 talks.
4. N-gram Analysis: 3 talks.

The proceedings was published by JSAI. More information about the workshop and past workshops are available at the LLLL homepage:

http://www.i.kyoto-u.ac.jp/~akihiro/LLLL.html

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 249–250, 2007.

2 The Post-workshop Proceedings

After the workshop, we asked the contributors to revise the workshop papers and submit them to the post-workshop proceedings, this volume of LNAI. We received nine revised papers and they were reviewed by three PC members again. Finally, we have selected the following 4 papers for this volume.

Akama and Zeugmann investigate the power of *consistent* learning under the framework of inductive inference of recursive functions. They also introduce and discuss several variations of consistent learning, for example, *coherent* learning.

De Brecht, Kobayashi, Tokunaga and Yamamoto generalize their previous results on inductive inference of algebraic structures from positive data, by introducing *closed set systems*. They also give some relationship with closed set systems and *approximate learning.*

Haraguchi and Okubo introduce a new method of a branch and bound search algorithm for finding only *top N number of extents of formal concepts*. Then, they give the experimental results showing that the method can effectively find potentially interesting documents.

Kurai, Minato and Zeugmann give a new method of n-gram analysis for sequence data based on *ZBDDs* (*Zero-suppressed BDDs*) that are standard tools for logical synthesis and give a compact representation of combinatorial itemsets. Also they give its potential power for knowledge discovery from sequence data.

Acknowledgments

We have finished LLLL2006 successfully. We would like to show our thankfulness to the invited speakers, all speakers of contributed papers, and all audiences who attended the workshop. We are grateful for the great support that we received from the program committee members:

Yoji Akama, Marta Arias, Hiroki Arimura, Eiju Hirowatari, Tamás Horváth, Katsumi Inoue, Roni Khardon, Eric Martin, Shin-ichi Minato, Tetsuhiro Miyahara, Luc de Raedt, M. R. K. Krishna Rao, Taisuke Sato, Joe Suzuki, György Turán, and Hiroaki Watanabe.

In particular, we would express our great thankfulness to Prof. Hiroki Arimura, the chair of SIG-FPAI, for his support to LLLL2006 with the Grant-in-Aid for Specially Promoted Research. We finally express our great thankfulness to the committee of JSAI2006 for their arrangement, in particular, Prof. Takashi Washio and Dr. Akihiro Inokuchi, who are the organizers of the co-located international workshops of JSAI2006.

Consistency Conditions for Inductive Inference of Recursive Functions

Yohji Akama[1] and Thomas Zeugmann[2]

[1] Mathematical Institute, Tohoku University,
Sendai Miyagi Japan, 980-8578
akama@math.tohoku.ac.jp
[2] Division of Computer Science
Hokkaido University, N-14, W-9, Sapporo 060-0814, Japan
thomas@ist.hokudai.ac.jp

Abstract. A consistent learner is required to correctly and completely reflect in its actual hypothesis all data received so far. Though this demand sounds quite plausible, it may lead to the unsolvability of the learning problem.

Therefore, in the present paper several variations of consistent learning are introduced and studied. These variations allow a so-called δ–delay relaxing the consistency demand to all but the last δ data.

Additionally, we introduce the notion of *coherent* learning (again with δ–delay) requiring the learner to correctly reflect only the last datum (only the $n - \delta$th datum) seen.

Our results are threefold. First, it is shown that all models of coherent learning with δ–delay are exactly as powerful as their corresponding consistent learning models with δ–delay. Second, we provide characterizations for consistent learning with δ–delay in terms of complexity. Finally, we establish strict hierarchies for all consistent learning models with δ–delay in dependence on δ.

1 Introduction

Algorithmic learning has attracted much attention of researchers in various fields of computer science. Inductive inference addresses the question whether or not learning problems may be solved algorithmically at all. There has been huge progress since the pioneering paper of Gold [8] but several questions still deserve attention, in particular from the viewpoint of potential applications.

A main problem of algorithmic learning theory is to synthesize "global descriptions" for the objects to be learned from examples. Thus, one goal is the following. Let f be any computable function from $\mathbb{N}$ into $\mathbb{N}$. Given more and more examples $f(0), f(1), ..., f(n), \dots$ a learning strategy is required to compute a sequence of hypotheses $h_0, h_1, \dots, h_n, \dots$ the limit of which is a correct global description of the function f, i.e., a program that computes f. Since at

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 251–264, 2007.

any stage n of this learning process the strategy knows exclusively the examples $f(0), f(1), ..., f(n)$, one may be tempted to require the strategy to produce only hypotheses h_n such that for any $x \leq n$ the "hypothesis function" g described by h_n is defined and computes the value $f(x)$. Such a hypothesis is called *consistent.* If a hypothesis does not completely and correctly encode all information obtained so far about the unknown object it is called *inconsistent.* A learner exclusively outputting consistent hypotheses is called *consistent.* Requiring a consistent learner looks quite natural at first glance. Why a strategy should output a conjecture that is falsified by the data in hand?

But this is a misleading impression. One of the surprising phenomena discovered in inductive inference is the inconsistency phenomenon (cf., e.g., Barzdin [2], Blum and Blum [4], Wiehagen and Liepe [21], Jantke and Beick [12] as well as Osherson, Stob and Weinstein [17] and the references therein). That is, there are classes of recursive functions that can only be learned by inconsistent strategies.

Naturally, the inconsistency phenomenon has been studied subsequently by many researchers. The reader is encouraged to consult e.g., Jain *et al.* [11], Fulk [7], Freivalds, Kinber and Wiehagen [6] and Wiehagen and Zeugmann [22, 23] for further investigations concerning consistent and inconsistent learning.

In the present paper we introduce and study several variations of consistent learning that have not been considered in the literature. First, we introduce the notion of *coherent* learning. A learner is said to be coherent if it correctly reflects the last datum received (say $f(x_n)$), i.e., if every h_n output satisfies the requirement that the "hypothesis function" g described by h_n is defined on input x_n and $g(x_n) = f(x_n)$. Furthermore, we introduce the notion of δ–delay, where $\delta \in \mathbb{N}$. Then, *coherent* learning with δ–delay means that every h_n output satisfies that $g(x_n \dot{-} \delta)$ is defined and $g(x_n \dot{-} \delta) = f(x_n \dot{-} \delta)$ (cf. Definition 5).

Furthermore, we adopt the notion of δ–delay to the consistent learning types mainly studied so far, i.e., to $\mathcal{CONS}$ (defined by Barzdin [2]), $\mathcal{R}$-$\mathcal{CONS}$ (introduced by Jantke and Beick [12]) and $\mathcal{T}$-$\mathcal{CONS}$ (defined by Wiehagen and Liepe [21]) (cf. Definitions 2, 3 and 4, respectively).

Our results are threefold. First, it is shown that all models of coherent learning with δ–delay are exactly as powerful as their corresponding consistent learning models with δ–delay, see Theorem 1. Second, we provide characterizations for consistent learning with δ–delay in terms of complexity (cf. Theorems 2 and 3). Finally, we establish strict hierarchies for all consistent learning models with δ–delay in dependence on δ, see Theorem 5 and Corollary 6.

The paper is structured as follows. Section 2 presents notation and definitions. Then we show the equivalence of coherent and consistent learning for all variants defined (cf. Section 3). The announced characterizations are shown in Section 4. In Section 5 we prove three new infinite hierarchies for consistent learning with δ–delay. In Section 6 we discuss the results obtained and present open problems. The bibliography is provided in the References.

2 Preliminaries

Unspecified notations follow Rogers [18]. $\mathbb{N} = \{0, 1, 2, ...\}$ denotes the set of all natural numbers. The set of all finite sequences of natural numbers is denoted by $\mathbb{N}^*$. For $a, b \in \mathbb{N}$ we define $a \dot{-} b$ to be $a - b$ if $a \geq b$ and 0, otherwise.

By $\mathfrak{P}$ and $\mathfrak{T}$ we denote the set of all partial and total functions of one variable over $\mathbb{N}$, respectively. The classes of all partial recursive and recursive functions of one, and two arguments over $\mathbb{N}$ are denoted by $\mathcal{P}$, $\mathcal{P}^2$, $\mathcal{R}$, and $\mathcal{R}^2$, respectively. $\mathcal{R}_{0,1}$ denotes the set of all $0-1$ valued recursive functions (recursive predicates). Sometimes it will be suitable to identify a recursive function with the sequence of its values, e.g., let $\alpha = (a_0, ..., a_k) \in \mathbb{N}^*$, $j \in \mathbb{N}$, and $p \in \mathcal{R}_{0,1}$; then we write $\alpha j p$ to denote the function f for which $f(x) = a_x$, if $x \leq k$, $f(k+1) = j$, and $f(x) = p(x-k-2)$, if $x \geq k+2$.

Every function $\psi \in \mathcal{P}^2$ is said to be a *numbering*. Furthermore, let $\psi \in \mathcal{P}^2$, then we write ψ_i instead of $\lambda x \psi(i, x)$ and set $\mathcal{P}_\psi = \{\psi_i \mid i \in \mathbb{N}\}$ as well as $\mathcal{R}_\psi = \mathcal{P}_\psi \cap \mathcal{R}$. Consequently, if $f \in \mathcal{P}_\psi$, then there is a number i such that $f = \psi_i$. If $f \in \mathcal{P}$ and $i \in \mathbb{N}$ are such that $\psi_i = f$, then i is called a ψ–program for f. A numbering $\varphi \in \mathcal{P}^2$ is called a Gödel numbering (cf. Rogers [18]) iff $\mathcal{P}_\varphi = \mathcal{P}$, and for any numbering $\psi \in \mathcal{P}^2$, there is a $c \in \mathcal{R}$ such that $\psi_i = \varphi_{c(i)}$ for all $i \in \mathbb{N}$. *Göd* denotes the set of all Gödel numberings. Furthermore, we write (φ, Φ) to denote any complexity measure as defined in Blum [5]. That is, $\varphi \in$ *Göd*, $\Phi \in \mathcal{P}^2$ and (1) $dom(\varphi_i) = dom(\Phi_i)$ for all $i \in \mathbb{N}$ and (2) the predicate $\Phi_i(x) = y$ is uniformly recursive for all $i, x, y \in \mathbb{N}$.

Furthermore, let $\mathcal{NUM} = \{U \mid (\exists \psi \in \mathcal{R}^2)\, [U \subseteq \mathcal{P}_\psi]\}$ denote the family of all subsets of all recursively enumerable classes of recursive functions.

Moreover, using a fixed encoding $\langle ... \rangle$ of $\mathbb{N}^*$ onto $\mathbb{N}$ we write f^n instead of $\langle (f(0), ..., f(n)) \rangle$, for any $n \in \mathbb{N}$, $f \in \mathcal{R}$.

The quantifier $\overset{\infty}{\forall}$ stands for "almost everywhere" and means "all but finitely many." Finally, a sequence $(j_n)_{j \in \mathbb{N}}$ of natural numbers is said to *converge* to the number j iff all but finitely many numbers of it are equal to j. Next we define some concepts of learning.

Definition 1 (Gold [8]). *Let $U \subseteq \mathcal{R}$ and let $\psi \in \mathcal{P}^2$. The class U is said to be* learnable in the limit with respect to ψ *iff there is a strategy $S \in \mathcal{P}$ such that for each function $f \in U$,*

(1) *for all $n \in \mathbb{N}$, $S(f^n)$ is defined,*
(2) *there is a $j \in \mathbb{N}$ such that $\psi_j = f$ and the sequence $(S(f^n))_{n \in \mathbb{N}}$ converges to j.*

If U is learnable in the limit with respect to ψ by a strategy S, we write $U \in \mathcal{LIM}_\psi(S)$. Let $\mathcal{LIM}_\psi = \{U \mid U$ is learnable in the limit w.r.t. $\psi\}$, and let $\mathcal{LIM} = \bigcup_{\psi \in \mathcal{P}^2} \mathcal{LIM}_\psi$.

As far as the semantics of the hypotheses produced by a strategy S is concerned, whenever S is defined on input f^n, then we always interpret the number $S(f^n)$ as a ψ–number. This convention is adopted to all the definitions below. Furthermore, note that $\mathcal{LIM}_\varphi = \mathcal{LIM}$ for any Gödel numbering φ. In the above definition $\mathcal{LIM}$ stands for "limit." Note that within Definition 1 no requirement is made concerning the intermediate hypotheses output by the strategy S.

Definition 2. *Let $U \subseteq \mathcal{R}$, let $\psi \in \mathcal{P}^2$ and let $\delta \in \mathbb{N}$. The class U is called* consistently learnable in the limit with δ–delay with respect to ψ *iff there is a strategy $S \in \mathcal{P}$ such that*

(1) *$U \in \mathcal{LIM}_\psi(S)$,*
(2) *$\psi_{S(f^n)}(x) = f(x)$ for all $f \in U$, $n \in \mathbb{N}$ and $x \leq n \dot{-} \delta$.*

$\mathcal{CONS}^\delta_\psi(S)$, $\mathcal{CONS}^\delta_\psi$ and $\mathcal{CONS}^\delta$ are defined analogously as above.

Note that for $\delta = 0$ we get Barzdin's [2] original definition of $\mathcal{CONS}$. We therefore usually omit the upper index δ if $\delta = 0$. This is also done for all other versions of consistent learning defined below. Moreover, we use the term *δ–delay*, since a consistent strategy with δ–delay correctly reflects all but at most the last δ data seen so far. If a strategy does not always works consistently with δ–delay we call it *δ–delay inconsistent*.

Next, we modify $\mathcal{CONS}^\delta$ in the same way Jantke and Beick [12] changed $\mathcal{CONS}$, i.e., we add the requirement that the strategy is defined on every input.

Definition 3. *Let $U \subseteq \mathcal{R}$, let $\psi \in \mathcal{P}^2$ and let $\delta \in \mathbb{N}$. The class U is called* $\mathcal{R}$–consistently learnable in the limit with δ–delay with respect to ψ *iff there is a strategy $S \in \mathcal{R}$ such that $U \in \mathcal{CONS}^\delta_\psi(S)$.*
$\mathcal{R}$-$\mathcal{CONS}^\delta_\psi(S)$, $\mathcal{R}$-$\mathcal{CONS}^\delta_\psi$ and $\mathcal{R}$-$\mathcal{CONS}^\delta$ are defined analogously as above.

Note that in the latter definition consistency with δ–delay is only demanded for inputs that correspond to some function f from the target class. Therefore, in the following definition we incorporate Wiehagen and Liepe's [21] requirement to a strategy to work consistently on all inputs into our scenario of consistency with δ–delay.

Definition 4. *Let $U \subseteq \mathcal{R}$, let $\psi \in \mathcal{P}^2$ and let $\delta \in \mathbb{N}$. The class U is called* $\mathcal{T}$–consistently learnable in the limit with δ–delay with respect to ψ *iff there is a strategy $S \in \mathcal{R}$ such that*

(1) *$U \in \mathcal{CONS}^\delta_\psi(S)$,*
(2) *$\psi_{S(f^n)}(x) = f(x)$ for all $f \in \mathcal{R}$, $n \in \mathbb{N}$ and $x \leq n \dot{-} \delta$.*

$\mathcal{T}$-$\mathcal{CONS}^\delta_\psi(S)$, $\mathcal{T}$-$\mathcal{CONS}^\delta_\psi$ and $\mathcal{T}$-$\mathcal{CONS}^\delta$ are defined in the same way as above.

Next, we introduce *coherent* learning (again with δ-delay). While our consistency with δ-delay demand requires a strategy to correctly reflect all but at most the last δ data seen so far, the coherence requirement only demands to correctly reflect the value $f(n \dot{-} \delta)$ on input f^n.

Definition 5. *Let $U \subseteq \mathcal{R}$, let $\psi \in \mathcal{P}^2$ and let $\delta \in \mathbb{N}$. The class U is called* coherently learnable in the limit with δ–delay with respect to ψ *iff there is a strategy $S \in \mathcal{P}$ such that*

(1) $U \in \mathcal{LIM}_\psi(S)$,
(2) $\psi_{S(f^n)}(n \dot{-} \delta) = f(n \dot{-} \delta)$ *for all* $f \in U$ *and all* $n \in \mathbb{N}$.

$\mathcal{COH}^\delta_\psi(S)$, $\mathcal{COH}^\delta_\psi$ *and* $\mathcal{COH}^\delta$ *are defined analogously as above.*

Now, performing the same modifications to coherent learning with δ–delay as we did in Definitions 3 and 4 to consistent learning with δ–delay results in the learning types $\mathcal{R}\text{-}\mathcal{COH}^\delta$ and $\mathcal{T}\text{-}\mathcal{COH}^\delta$, respectively. We therefore omit the formal definitions of these learning types here.

Using standard techniques one can show that for all $\delta \in \mathbb{N}$ and all learning types $LT \in \{\mathcal{CONS}^\delta,\ \mathcal{R}\text{-}\mathcal{CONS}^\delta,\ \mathcal{T}\text{-}\mathcal{CONS}^\delta,\ \mathcal{COH}^\delta,\ \mathcal{R}\text{-}\mathcal{COH}^\delta,\ \mathcal{T}\text{-}\mathcal{COH}^\delta\}$ we have $LT_\varphi = LT$ for every Gödel numbering φ.

Note that in the following $\subseteq$ denotes subset and $\subset$ denotes *proper* subset. Finally, incomparability of sets is denoted by $\#$.

3 Coherence and Consistency of Learning Strategies

In this section we study the problem whether or not the relaxation to learn coherently with δ–delay instead of demanding consistency with δ–delay does enhance the learning power of the corresponding learning types introduced in Section 2. The negative answer is provided by the following theorem.

Theorem 1. *Let* $\delta \in \mathbb{N}$ *be arbitrarily fixed. Then we have*

(1) $\mathcal{CONS}^\delta = \mathcal{COH}^\delta$,
(2) $\mathcal{R}\text{-}\mathcal{CONS}^\delta = \mathcal{R}\text{-}\mathcal{COH}^\delta$,
(3) $\mathcal{T}\text{-}\mathcal{CONS}^\delta = \mathcal{T}\text{-}\mathcal{COH}^\delta$.

Proof. By definition, we obviously have $\mathcal{CONS}^\delta \subseteq \mathcal{COH}^\delta$, $\mathcal{R}\text{-}\mathcal{CONS}^\delta \subseteq \mathcal{R}\text{-}\mathcal{COH}^\delta$ and $\mathcal{T}\text{-}\mathcal{CONS}^\delta \subseteq \mathcal{T}\text{-}\mathcal{COH}^\delta$.

For showing the opposite directions we can essentially use in all three cases the same idea. Let $\delta \in \mathbb{N}$, $\varphi \in \textit{Göd}$, $U \subseteq \mathcal{R}$ and any strategy $\hat{S}$ be arbitrarily fixed such that $U \in LT_\varphi(\hat{S})$, where $LT \in \{\mathcal{COH}^\delta,\ \mathcal{R}\text{-}\mathcal{COH}^\delta,\ \mathcal{T}\text{-}\mathcal{COH}^\delta\}$. Next, we define a strategy S as follows. Let $f \in \mathcal{R}$ and let $n \in \mathbb{N}$. On input f^n do the following.

1. Compute $\hat{S}(f^0), \dots, \hat{S}(f^n)$ and determine the largest number n^* such that $\hat{S}(f^{n^*-1}) \neq \hat{S}(f^{n^*})$.
2. Output the canonical φ–program i computing the following function g:
   $g(x) = f(x)$ for all $x \leq n^*$, and
   $g(x) = \varphi_{\hat{S}(f^{n^*})}(x)$ for all $x > n^*$.

First, we show that S learns U consistently with δ–delay.

By construction, we have $\varphi_{S(f^n)}(x) = f(x)$ for all $x \leq n^*$, and thus S is consistent on all data $f(0), \dots, f(n^*)$. If $n - n^* \leq \delta$, we are already done. Finally, if $n - n^* > \delta$, then we exploit the fact that $\hat{S}$ works coherently with δ–delay and that $\hat{S}(f^{n^*+k}) = \hat{S}(f^{n^*})$ for all $k = 1,\dots,n-n^*$. Thus, for all $k \in \{1,\dots,n-n^*-\delta\}$ we get

$$\varphi_{S(f^n)}(n^*+k) = \varphi_{\hat{S}(f^{n^*})}(n^*+k) = \varphi_{\hat{S}(f^{n^*+\delta+k})}(n^*+k) = f(n^*+k)\ . \quad \text{(i)}$$

Since in this case $\hat{S}(f^n)$ is defined for all $f \in U$ and all $n \in \mathbb{N}$, we can directly conclude that $S(f^n)$ is defined for all $f \in U$ and all $n \in \mathbb{N}$, too. This proves Assertion (1).

If $\hat{S} \in \mathcal{R}$, then so is S and thus Assertion (2) follows.

Finally, if $\hat{S} \in \mathcal{R}$ and $\hat{S}$ works $\mathcal{T}$–coherently, then we directly get $S \in \mathcal{R}$ and S is $\mathcal{T}$–consistent, since now (i) is true for all $f \in \mathcal{R}$. This completes the proof. □

4 Characterizations

Within this section, we characterize consistent learning with δ–delay in terms of complexity.

First, we recall the definitions of recursive and general recursive operator. Let $(F_x)_{x\in\mathbb{N}}$ be the canonical enumeration of all finite functions.

Definition 6 (Rogers [18]). *A mapping $\mathfrak{O}\colon \mathfrak{P} \mapsto \mathfrak{P}$ from partial functions to partial functions is called a* partial recursive operator *iff there is a recursively enumerable set $W \subset \mathbb{N}^3$ such that for any $y, z \in \mathbb{N}$ it holds that $\mathfrak{O}(f)(y) = z$ iff there is $x \in \mathbb{N}$ such that $(x, y, z) \in W$ and f extends the finite function F_x.*

Furthermore, $\mathfrak{O}$ is called a general recursive operator *iff $\mathfrak{T} \subseteq dom(\mathfrak{O})$, and $f \in \mathfrak{T}$ implies $\mathfrak{O}(f) \in \mathfrak{T}$.*

A mapping $\mathfrak{O}\colon \mathcal{P} \mapsto \mathcal{P}$ is called an effective operator *iff there is a function $g \in \mathcal{R}$ such that $\mathfrak{O}(\varphi_i) = \varphi_{g(i)}$ for all $i \in \mathbb{N}$. An effective operator $\mathfrak{O}$ is said to be* total effective *provided that $\mathcal{R} \subseteq dom(\mathfrak{O})$, and $\varphi_i \in \mathcal{R}$ implies $\mathfrak{O}(\varphi_i) \in \mathcal{R}$.*

For more information about general recursive operators and effective operators the reader is referred to [10, 15, 24]. If $\mathfrak{O}$ is an operator which maps functions to functions, we write $\mathfrak{O}(f, x)$ to denote the value of the function $\mathfrak{O}(f)$ at the argument x. Any computable operator can be realized by a 3-tape Turing machine T which works as follows: If for an arbitrary function $f \in dom(\mathfrak{O})$, all pairs $(x, f(x))$, $x \in dom(f)$ are written down on the input tape of T (repetitions are allowed), then T will write exactly all pairs $(x, \mathfrak{O}(f, x))$ on the output tape of T (under unlimited working time).

Let $\mathfrak{O}$ be a general recursive or total effective operator. Then, for $f \in dom(\mathfrak{O})$, $m \in \mathbb{N}$ we set: $\Delta\mathfrak{O}(f, m) =$"the least n such that, for all $x \le n$, $f(x)$ is defined and, for the computation of $\mathfrak{O}(f, m)$, the Turing machine T only uses the pairs $(x, f(x))$ with $x \le n$; if such an n does not exist, we set $\Delta\mathfrak{O}(f, m) = \infty$."

For $u \in \mathcal{R}$ we define Ω_u to be the set of all partial recursive operators $\mathfrak{O}$ satisfying $\Delta\mathfrak{O}(f, m) \le u(m)$ for all $f \in dom(\mathfrak{O})$. For the sake of notation, below we shall use $id + \delta$, $\delta \in \mathbb{N}$, to denote the function $u(x) = x + \delta$ for all $x \in \mathbb{N}$.

Note that in the following we use mainly ideas and techniques from Wiehagen [20] who proved theses theorems for the case $\delta = 0$. Variants of these characterizations for $\delta = 0$ can also be found in Wiehagen and Liepe [21] as well as in Odifreddi [16].

Furthermore, in the following we always assume that learning is done with respect to any fixed $\varphi \in Göd$.

As in Blum and Blum [4] we define operator complexity classes as follows. Let $\mathfrak{O}$ be any computable operator; then we set

$$\mathcal{R}_{\mathfrak{O}} = \{f \mid \exists i[\varphi_i = f \wedge \overset{\infty}{\forall} x[\Phi_i(x) \le \mathfrak{O}(f,x)]]\} \cap \mathcal{R} \ .$$

First, we characterize $\mathcal{T}\text{-}\mathcal{CONS}^{\delta}$.

Theorem 2. *Let $U \subseteq \mathcal{R}$ and let $\delta \in \mathbb{N}$; then we have: $U \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ if and only if there exists a general recursive operator $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $U \subseteq \mathcal{R}_{\mathfrak{O}}$.*

Proof. Necessity. Let $U \in \mathcal{CONS}^{\delta}(S)$, $S \in \mathcal{R}$. Then for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$ we define $\mathfrak{O}(f,n) = \Phi_{S(f^{n+\delta})}(n)$.

Since $\varphi_{S(f^{n+\delta})}(n)$ is defined for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$, by Condition (2) of Definition 4, we directly get from Condition (1) of the definition of a complexity measure that $\Phi_{S(f^{n+\delta})}(n)$ is defined for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$, too. Moreover, for every $t \in \mathfrak{T}$ and $n \in \mathbb{N}$ there is an $f \in \mathcal{R}$ such that $t^n = f^n$. Hence, we have $\mathfrak{O}(\mathfrak{T}) \subseteq \mathcal{R} \subseteq \mathfrak{T}$. Moreover, in order to compute $\mathfrak{O}(f,n)$ the operator $\mathfrak{O}$ reads only the values $f(0), \dots, f(n+\delta)$. Thus, we have $\mathfrak{O} \in \Omega_{id+\delta}$.

Now, let $f \in U$. Then the sequence $(S(f^n))_{n\in\mathbb{N}}$ converges to a correct φ–program i for f. Consequently, $\mathfrak{O}(f,n) = \Phi_i(n)$ for almost all $n \in \mathbb{N}$. Therefore, we conclude $U \subseteq \mathcal{R}_{\mathfrak{O}}$.

Sufficiency. Let $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $U \subseteq \mathcal{R}_{\mathfrak{O}}$. We have to define a strategy $S \in \mathcal{R}$ such that $U \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}(S)$. By the definition of $\mathcal{R}_{\mathfrak{O}}$ we know that for every $f \in U$ there exist i and k such that $\varphi_i = f$ and $\Phi_i(x) \le \max\{k,\ \mathfrak{O}(f,x)\}$ for all x. Thus, the desired strategy S searches for the first pair (i,k) in the canonical enumeration c_2 of $\mathbb{N} \times \mathbb{N}$ and converges to i provided it has been found. Until this pair (i,k) is found, the strategy S outputs auxiliary consistent hypotheses. For doing this, we choose $g \in \mathcal{R}$ such that $\varphi_{g(\langle\alpha\rangle)}(x) = y_x$ for every tuple $\alpha \in \mathbb{N}^*$, $\alpha = (y_0, \dots, y_n)$ and all $x \le n$.

$S(f^n) =$ “Compute $\mathfrak{O}(f,x)$ for all $x \le n \dot{-} \delta$. Search for the least $z \le n$ such that for $c_2(z) = (i,k)$ the conditions

(A) $\Phi_i(x) \le \max\{k,\ \mathfrak{O}(f,x)\}$ for all $x \le n \dot{-} \delta$, and

(B) $\varphi_i(x) = f(x)$ for all $x \le n \dot{-} \delta$

are fulfilled. If such a z is found, set $S(f^n) = i$.

Otherwise, set $S(f^n) = g(f^n)$.”

Since $\mathfrak{O} \in \Omega_{id+\delta}$, the strategy can compute $\mathfrak{O}(f,x)$ for all $x \le n \dot{-} \delta$ and since $c_2 \in \mathcal{R}$, it also can perform the desired search effectively. By Condition (2) of the definition of a complexity measure, the test in (A) can be performed effectively, too. If this test has succeeded, then Test (B) can also be effectively executed by Condition (1) of the definition of a complexity measure. Thus, we get $S \in \mathcal{R}$. Finally, by construction S is always consistent with δ-delay, and if $f \in U$ it converges to a correct φ–program for f. $\square$

Theorem 3. *Let $U \subseteq \mathcal{R}$ and let $\delta \in \mathbb{N}$; then we have: $U \in \mathcal{CONS}^\delta$ if and only if there exists a partial recursive operator $\mathfrak{O} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(U) \subseteq \mathcal{R}$ and $U \subseteq \mathcal{R}_{\mathfrak{O}}$.*

Proof. The necessity is proved *mutatis mutandis* as in the proof of Theorem 2 with the only modification that $\mathfrak{O}(f, x)$ is now defined for all $f \in U$ instead of $f \in \mathcal{R}$. This directly yields $\mathfrak{O} \in \Omega_{id+\delta}$, $\mathfrak{O}(U) \subseteq \mathcal{R}$ and $U \subseteq \mathcal{R}_{\mathfrak{O}}$.

The only modification for the sufficiency part is to leave $S(f^n)$ undefined if $\mathfrak{O}(f, x)$ is not defined for $f \notin U$. We omit the details. $\square$

We finish this section by using Theorem 2 to show that $\mathcal{T}$-$\mathcal{CONS}^\delta$ is closed under enumerable unions. Looking at applications this is a favorable property, since it provides a tool to build more powerful learners from simpler ones.

Theorem 4. *Let $\delta \in \mathbb{N}$ and let $(S_i)_{i\in\mathbb{N}}$ be a recursive enumeration of strategies working $\mathcal{T}$-consistently with δ-delay. Then there exists a strategy $S \in \mathcal{R}$ such that $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}(S_i) \subseteq \mathcal{T}\text{-}\mathcal{CONS}(S)$.*

Proof. The proof of the necessity of Theorem 2 shows that the construction of the operator $\mathfrak{O}$ is effective provided a program for the strategy is given. Thus, we effectively obtain a recursive enumeration $(\mathfrak{O}_i)_{i\in\mathbb{N}}$ of operators $\mathfrak{O}_i \in \Omega_{id+\delta}$ such that $\mathfrak{O}_i(\mathcal{R}) \subseteq \mathcal{R}$ and $\mathcal{T}\text{-}\mathcal{CONS}(S_i) \subseteq \mathcal{R}_{\mathfrak{O}_i}$.

Now, we define an operator $\mathfrak{O}$ as follows. Let $f \in \mathcal{R}$ and $x \in \mathbb{N}$. We set $\mathfrak{O}(f, x) = \max\{\mathfrak{O}_i(f, x) \mid i \leq x\}$.

Thus, we directly get $\mathfrak{O} \in \Omega_{id+\delta}$, $\mathfrak{O}(\mathcal{R}) \subseteq \mathcal{R}$ and $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}(S_i) \subseteq \mathcal{R}_{\mathfrak{O}}$. Thus, by Theorem 2 we can conclude $\bigcup_{i\in\mathbb{N}} \mathcal{T}\text{-}\mathcal{CONS}(S_i) \subseteq \mathcal{T}\text{-}\mathcal{CONS}(S)$. $\square$

On the other hand, $\mathcal{CONS}^\delta$ and $\mathcal{R}$-$\mathcal{CONS}$ are not even closed under finite union. This is a direct consequence of a more general result Barzdin [1] showed, i.e., there are classes $U = \{f \mid f \in \mathcal{R},\ \varphi_{f(0)} = f\}$ and $V = \{\alpha 0^\infty \mid \alpha \in \mathbb{N}^*\}$ such that $U \cup V \notin \mathcal{LIM}$. Now, it is easy to verify U, $V \in \mathcal{R}\text{-}\mathcal{CONS}^\delta$ and thus U, $V \in \mathcal{CONS}^\delta$ for every $\delta \in \mathbb{N}$, but since $U \cup V \notin \mathcal{LIM}$ we clearly have $U \cup V \notin \mathcal{R}\text{-}\mathcal{CONS}^\delta$ and $U \cup V \notin \mathcal{CONS}^\delta$ for all $\delta \in \mathbb{N}$.

5 Hierarchy Results

Within this section we study the problem whether or not the introduction of δ–delay to consistent learning yields an advantage with respect to the learning power of the defined learning types.

For answering this problem it is advantageous to recall the definition of reliable learning introduced by Blum and Blum [4] and Minicozzi [14]. Intuitively, a learner M is reliable provided it converges if and only if it learns.

Definition 7 (Blum and Blum [4], Minicozzi [14]). *Let $U \subseteq \mathcal{R}$, let $\mathcal{M} \subseteq \mathfrak{T}$ and let $\varphi \in$ Göd; then U is said to be* reliably learnable on $\mathcal{M}$ *if there is a strategy $S \in \mathcal{R}$ such that*

(1) *$U \in \mathcal{LIM}_\varphi(S)$, and*
(2) *for all functions $f \in \mathcal{M}$, if the sequence $(S(\varphi^n))_{n\in\mathbb{N}}$ converges, say to j, then $\varphi_j = f$.*

By $\mathcal{M}$-$\mathcal{REL}$ we denote the family of all function classes that are reliably learnable on $\mathcal{M}$.

In particular, we shall consider the cases where $\mathcal{M} = \mathfrak{T}$ and $\mathcal{M} = \mathcal{R}$, i.e., reliable learnability on the set of all total functions and all recursive functions, respectively. Furthermore, in the following, for any set M, we use $\wp(M)$ to denote the power set of M.

Theorem 5. *The following statements hold for all $\delta \in \mathbb{N}$:*

(1) $\mathcal{T}$-$\mathcal{CONS}^\delta \subset \mathcal{T}$-$\mathcal{CONS}^{\delta+1} \subset \mathfrak{T}$-$\mathcal{REL}$,
(2) $\mathcal{NUM} \cap \wp(\mathcal{R}_{0,1}) = \mathcal{T}$-$\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{0,1}) = \mathcal{T}$-$\mathcal{CONS}^{\delta+1} \cap \wp(\mathcal{R}_{0,1}) = \mathfrak{T}$-$\mathcal{REL} \cap \wp(\mathcal{R}_{0,1})$,
(3) $\mathcal{T}$-$\mathcal{CONS}^\delta \cap \wp(\mathcal{R}_{0,1}) \subset \mathcal{R}$-$\mathcal{REL} \cap \wp(\mathcal{R}_{0,1})$.

Proof. Let $\delta \in \mathbb{N}$ be arbitrarily fixed. Then by Definition 4 we obviously have $\mathcal{T}$-$\mathcal{CONS}^\delta \subseteq \mathcal{T}$-$\mathcal{CONS}^{\delta+1}$. For showing $\mathcal{T}$-$\mathcal{CONS}^{\delta+1} \setminus \mathcal{T}$-$\mathcal{CONS}^\delta \neq \emptyset$ we use the following class. Let (φ, Φ) be any complexity measure; we set

$$U_{\delta+1}^{(\varphi,\Phi)} = \{f \mid f \in \mathcal{R},\ \varphi_{f(0)} = f,\ \forall x[\Phi_{f(0)}(x) \le f(x+\delta+1)]\} .$$

Claim 1. $U_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}$-$\mathcal{CONS}^{\delta+1}$.

The desired strategy S is defined as follows. Let $g \in \mathcal{R}$ be the function defined in the sufficiency proof of Theorem 2. For all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$ we set

$$S(f^n) = \begin{cases} f(0), & \text{if } n \le \delta \text{ or } n > \delta \text{ and } \Phi_{f(0)}(y) \le f(y+\delta+1) \\ & \text{and } \varphi_{f(0)}(y) = f(y) \text{ for all } y \le n \dot{-} \delta \dot{-} 1 \\ g(f^n), & \text{otherwise.} \end{cases}$$

Now, by construction one easily verifies $U_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}$-$\mathcal{CONS}^{\delta+1}(S)$. This proves Claim 1.

Claim 2. $U_{\delta+1}^{(\varphi,\Phi)} \notin \mathcal{T}$-$\mathcal{CONS}^\delta$.

Suppose the converse. Then there must be a strategy $S \in \mathcal{R}$ such that $U_{\delta+1}^{(\varphi,\Phi)} \in \mathcal{T}$-$\mathcal{CONS}^\delta(S)$. We continue by constructing a function φ_{i^*} belonging to $U_{\delta+1}^{(\varphi,\Phi)}$ but on which S fails.

Furthermore, let $r \in \mathcal{R}$ be such that $\Phi_i = \varphi_{r(i)}$ for all $i \in \mathbb{N}$ and r is strongly monotone growing, i.e., $r(i) < r(i+1)$ for all $i \in \mathbb{N}$. Then *range*(r) is recursive (cf. Rogers [18]). Choose $s \in \mathcal{R}$ such that for all $j \in \mathbb{N}$ we have for all $x \le \delta$

$$\varphi_{s(j)}(x) = \begin{cases} i, & \text{if there is an } i \text{ with } r(i) = j\ , \\ 0, & \text{otherwise.} \end{cases}$$

For the further definition of $\varphi_{s(j)}$ we also use in every step $\delta+1$ arguments. For $x = 0,\ \delta+1,\ 2\delta+2,\ 3\delta+3, \ldots$ we set

$$\begin{aligned}\varphi_{s(j)}(x+\delta+1) &= \varphi_j(x)+1\\ &\ \,\vdots\\ \varphi_{s(j)}(x+2\delta+1) &= \varphi_j(x+\delta)+1\end{aligned}$$

provided $\varphi_j(x),\ \varphi_j(x+1), \ldots, \varphi_j(x+\delta)$ are all defined,$\varphi^{x+\delta}_{s(j)}$ is defined and

$$S\left(\varphi^{x+\delta}_{s(j)}\right) = S\left(\langle(\varphi_{s(j)}(0), \ldots, \varphi_{s(j)}(x+\delta), \varphi_j(x), \ldots, \varphi_j(x+\delta))\rangle\right)$$

and

$$\begin{aligned}\varphi_{s(j)}(x+\delta+1) &= \varphi_j(x)\\ &\ \,\vdots\\ \varphi_{s(j)}(x+2\delta+1) &= \varphi_j(x+\delta)\end{aligned}$$

provided $\varphi_j(x),\ \varphi_j(x+1), \ldots, \varphi_j(x+\delta)$ are all defined, $\varphi^{x+\delta}_{s(j)}$ is defined and

$$S\left(\varphi^{x+\delta}_{s(j)}\right) \neq S\left(\langle(\varphi_{s(j)}(0), \ldots, \varphi_{s(j)}(x+\delta), \varphi_j(x), \ldots, \varphi_j(x+\delta))\rangle\right) .$$

Otherwise, $\varphi_{s(j)}(x+\delta+1), \ldots, \varphi_{s(j)}(x+2\delta+1)$ remain undefined.

By the Fixpoint Theorem (cf. Rogers [18]) there exists a number i^* such that $\varphi_{s(r(i^*))} = \varphi_{i^*}$.

Next, we show that $\varphi_{i^*} \in U^{(\varphi,\Phi)}_{\delta+1}$. This is done inductively. For the induction base, by construction we have $\varphi_{i^*}(0) = \cdots = \varphi_{i^*}(\delta) = i^*$. Hence, $\Phi_{i^*}(0), \ldots, \Phi_{i^*}(\delta)$ are all defined, too. Therefore, we know that $\varphi^{\delta}_{s(r(i^*))}$ is defined and so either $\varphi_{s(r(i^*))}(\delta+1) = \Phi_{i^*}(0)+1,\ \ldots,\ \varphi_{s(r(i^*))}(2\delta+1) = \Phi_{i^*}(\delta)+1$ provided

$$S\left(\varphi^{\delta}_{s(r(i^*))}\right) = S\left(\langle(\varphi_{s((r(i^*))}(0), \ldots, \varphi_{s((r(i^*))}(\delta), \Phi_{i^*}(0), \ldots, \Phi_{i^*}(\delta))\rangle\right)$$

or $\varphi_{s(r(i^*))}(\delta+1) = \Phi_{i^*}(0),\ \ldots,\ \varphi_{s(r(i^*))}(2\delta+1) = \Phi_{i^*}(\delta)$ if

$$S\left(\varphi^{\delta}_{s(r(i^*))}\right) \neq S\left(\langle(\varphi_{s((r(i^*))}(0), \ldots, \varphi_{s((r(i^*))}(\delta), \Phi_{i^*}(0), \ldots, \Phi_{i^*}(\delta))\rangle\right) .$$

Note that one of these cases must happen, since otherwise S would not be $\mathcal{T}$-consistent with δ–delay.

Hence, $\Phi_{i^*}(0) \leq \varphi_{i^*}(\delta+1), \ldots, \Phi_{i^*}(\delta) \leq \varphi_{i^*}(2\delta+1)$, since $\varphi_{s(r(i^*))} = \varphi_{i^*}$. So we know that $\varphi_{i^*}(\delta+1), \ldots, \varphi_{i^*}(2\delta+1)$ as well as $\Phi_{i^*}(\delta+1), \ldots, \Phi_{i^*}(2\delta+1)$ are all defined. This completes the induction base.

Consequently, we have the induction hypothesis that for some $x = 0, \delta + 1,\ 2\delta+2,\ 3\delta+3, \ldots$ the values $\varphi_{i^*}(z)$ are defined and $\Phi_{i^*}(z) \le \varphi_{i^*}(z+\delta+1)$ for all $z \le x+\delta$. This of course implies $\varphi^{x+\delta}_{s(r(i^*))}$ is defined, too. The induction step is done from x to $x+\delta+1$. First, we either have $\varphi_{s(r(i^*))}(x+\delta+1) = \Phi_{i^*}(x)+1$, $\ldots$, $\varphi_{s(r(i^*))}(x+2\delta+1) = \Phi_{i^*}(x+\delta)+1$ provided

$$S\left(\varphi^{x+\delta}_{s(r(i^*))}\right) = S\left(\langle(\varphi_{s(r(i^*))}(0), \ldots, \varphi_{s(r(i^*))}(x+\delta), \Phi_{i^*}(x), \ldots, \Phi_{i^*}(x+\delta))\rangle\right)$$

or $\varphi_{s(r(i^*))}(x+\delta+1) = \Phi_{i^*}(x)$, $\ldots$, $\varphi_{s(r(i^*))}(x+2\delta+1) = \Phi_{i^*}(x+\delta)$ if

$$S\left(\varphi^{x+\delta}_{s(r(i^*))}\right) \neq S\left(\langle(\varphi_{s(r(i^*))}(0), \ldots, \varphi_{s(r(i^*))}(x+\delta), \Phi_{i^*}(x), \ldots, \Phi_{i^*}(x+\delta))\rangle\right) .$$

Note that one of these cases must happen, since otherwise S would not be $\mathcal{T}$-consistent with δ–delay.

Therefore, $\varphi_{i^*}(x+\delta+1), \ldots, \varphi_{i^*}(x+2\delta+1)$ are all defined and $\Phi_{i^*}(x) \le \varphi_{i^*}(x+\delta+1), \ldots, \Phi_{i^*}(x+\delta) \le \varphi_{i^*}(x+2\delta+1)$.

Now, we also know that $\Phi_{i^*}(x+\delta+1)$, $\ldots$, $\Phi_{i^*}(x+2\delta+1)$ are all defined. Therefore, we have shown that $\varphi_{i^*} \in U^{(\varphi,\Phi)}_{\delta+1}$. Finally, by construction we directly obtain that S performs infinitely mind changes when successively fed φ_{i^*}, a contradiction to $U^{(\varphi,\Phi)}_{\delta+1} \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}(S)$. This proves Claim 2.

Taking into account that a strategy working $\mathcal{T}$-consistently with δ–delay converges when successively fed any function f iff it learns f, we directly get $\mathcal{T}\text{-}\mathcal{CONS}^{\delta} \subseteq \mathfrak{T}\text{-}\mathcal{REL}$ for every $\delta \in \mathbb{N}$. Furthermore, as shown in Minicozzi [14], $\mathfrak{T}\text{-}\mathcal{REL}$ is closed under recursively enumerable union. Therefore, setting $U = \bigcup_{\delta\in\mathbb{N}} U^{(\varphi,\Phi)}_{\delta+1}$ we can conclude $U \in \mathfrak{T}\text{-}\mathcal{REL}$. But obviously $U \notin \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ for any δ. This proves Assertion (1).

Using Theorem 2 one easily sees that for every operator $\mathfrak{O} \in \Omega_{id+\delta}$ there is a monotone operator $\hat{\mathfrak{O}} \in \Omega_{id+\delta}$ such that $\mathfrak{O}(f,x) \le \hat{\mathfrak{O}}(f,x)$ for all $f \in \mathcal{R}$ and all $x \in \mathbb{N}$. Here, we call an operator monotone if for all $f, g \in \mathcal{R}$ and $\overset{\infty}{\forall} x[f(x) \le g(x)]$ implies $\overset{\infty}{\forall} x[\mathfrak{O}(f,x) \le \mathfrak{O}(g,x)]$.

When restricted to learn classes U of recursive predicates this directly implies that for every function $f \in U$ there is a φ–program i such that $\varphi_i = f$ and $\overset{\infty}{\forall} x[\Phi_i(x) \le \hat{\mathfrak{O}}(1^\infty, x)]$. Thus, by the Extrapolation Theorem we can conclude $U \in \mathcal{NUM}$ (cf. Barzdin and Freivalds [3]).

The same ideas can be used to show the remaining part for $\mathfrak{T}\text{-}\mathcal{REL}$ (cf. Grabowski [9]). Hence, Assertion (2) is shown.

Finally, Assertion (3) is an immediate consequence of Assertion (2) and Theorems 2 and 3 from Stephan and Zeugmann [19] which together show that $\mathcal{NUM} \cap \wp(\mathcal{R}_{0,1}) \subset \mathcal{R}\text{-}\mathcal{REL} \cap \wp(\mathcal{R}_{0,1})$. This completes the proof. $\square$

Together with Theorem 4 the latter proof allows a nice corollary.

Corollary 6. *For all $\delta \in \mathbb{N}$ we have:*

(1) $\mathcal{CONS}^{\delta} \subset \mathcal{CONS}^{\delta+1}$,
(2) $\mathcal{R}\text{-}\mathcal{CONS}^{\delta} \subset \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1}$.

Proof. We use $U^{(\varphi,\Phi)}_{\delta+1}$ from the proof of Theorem 5 and $V = \{\alpha 0^\infty \mid \alpha \in \mathbb{N}^*\}$. Clearly, $U^{(\varphi,\Phi)}_{\delta+1}$, $V \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}$ and hence, by Theorem 4 we also have $U^{(\varphi,\Phi)}_{\delta+1} \cup V \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1}$. Consequently, $U^{(\varphi,\Phi)}_{\delta+1} \cup V \in \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1}$ and $U^{(\varphi,\Phi)}_{\delta+1} \cup V \in \mathcal{CONS}^{\delta+1}$. It remains to argue that $U^{(\varphi,\Phi)}_{\delta+1} \cup V \notin \mathcal{CONS}^{\delta}$. This will suffice, since $\mathcal{R}\text{-}\mathcal{CONS}^{\delta} \subseteq \mathcal{CONS}^{\delta}$.

Suppose the converse, i.e., there is a strategy $S \in \mathcal{P}$ such that $U^{(\varphi,\Phi)}_{\delta+1} \cup V \in \mathcal{CONS}^{\delta}(S)$. By the choice of V we can directly conclude that then $S \in \mathcal{R}$ and that S has to work consistently with δ–delay on every f^n, $f \in \mathcal{R}$ and $n \in \mathbb{N}$. But this would imply $U^{(\varphi,\Phi)}_{\delta+1} \cup V \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}(S)$, a contradiction to $U^{(\varphi,\Phi)}_{\delta+1} \notin \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$. □

A closer look at the latter proof shows that we have even proved the following corollary shedding some light on the power of our notion of δ–delay.

Corollary 7. *$\mathcal{T}\text{-}\mathcal{CONS}^{\delta+1} \setminus \mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.*

The situation is comparable to Lange-Zeugmann's [13] bounded example memory learnability BEM_k of languages from positive data, where BEM_k yields an infinite hierarchy such that $\bigcup_{k\in\mathbb{N}} BEM_k$ is a proper subclass of the class of all indexed families that can be conservatively learned.

On the one hand, the latter corollary shows the strength of δ–delay. On the other hand, the δ–delay cannot compensate all the learning power that is provided by the different consistency demands on the domain of the strategies.

Theorem 8. *$\mathcal{R}\text{-}\mathcal{CONS} \setminus \mathcal{T}\text{-}\mathcal{CONS}^{\delta} \neq \emptyset$ for all $\delta \in \mathbb{N}$.*

Proof. The proof can be done by using the class $U = \{f \mid f \in \mathcal{R},\ \varphi_{f(0)} = f\}$ of self-describing functions. Obviously, $U \in \mathcal{R}\text{-}\mathcal{CONS}(S)$ as witnessed by the strategy $S(f^n) = f(0)$ for all $f \in \mathcal{R}$ and all $n \in \mathbb{N}$. Now, assuming $U \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ for some $\delta \in \mathbb{N}$ would directly imply that $U \cup V \in \mathcal{T}\text{-}\mathcal{CONS}^{\delta}$ for the same δ (here V is the class defined in the proof of Corollary 6) by Theorem 4. But this is a contradiction to $U \cup V \notin \mathcal{LIM}$ as shown in Barzdin [1]. □

Finally, putting Corollary 7 and Theorem 8 together, we get the following incomparabilities.

Corollary 9. *$\mathcal{T}\text{-}\mathcal{CONS}^{\delta} \# \mathcal{CONS}^{\mu}$ and $\mathcal{T}\text{-}\mathcal{CONS}^{\delta} \# \mathcal{R}\text{-}\mathcal{CONS}^{\mu}$ for all δ, $\mu \in \mathbb{N}$ provided $\delta > \mu$.*

6 Conclusions and Future Work

Looking for possible relaxations for the demand to learn consistently we have introduced the notions of coherent learning and of δ-delay. As our results show, coherent learning with δ-delay has the same learning power as consistent learning with δ-delay for all versions considered. Thus, coherence is in fact no weakening of the consistency demand.

On the other hand, we could establish three new infinite hierarchies of consistent learning in dependence on the delay δ.

The figure below summarizes the achieved separations and coincidences of the various coherent and consistent learning models investigated within this paper.

$$
\begin{array}{ccccccccccc}
\mathcal{T}\text{-}\mathcal{COH} & \subset & \mathcal{T}\text{-}\mathcal{COH}^1 & \subset \cdots \subset & \mathcal{T}\text{-}\mathcal{COH}^\delta & \subset & \mathcal{T}\text{-}\mathcal{COH}^{\delta+1} & \subset \cdots \subset & \mathfrak{T}\text{-}\mathcal{REL} \\
\shortparallel & & \shortparallel & & \shortparallel & & \shortparallel & & \shortparallel \\
\mathcal{T}\text{-}\mathcal{CONS} & \subset & \mathcal{T}\text{-}\mathcal{CONS}^1 & \subset \cdots \subset & \mathcal{T}\text{-}\mathcal{CONS}^\delta & \subset & \mathcal{T}\text{-}\mathcal{CONS}^{\delta+1} & \subset \cdots \subset & \mathfrak{T}\text{-}\mathcal{REL} \\
\cap & & \cap & & \cap & & \cap & & \cap \\
\mathcal{R}\text{-}\mathcal{COH} & \subset & \mathcal{R}\text{-}\mathcal{COH}^1 & \subset \cdots \subset & \mathcal{R}\text{-}\mathcal{COH}^\delta & \subset & \mathcal{R}\text{-}\mathcal{COH}^{\delta+1} & \subset \cdots \subset & \mathcal{R}\text{-}\mathcal{REL} \\
\shortparallel & & \shortparallel & & \shortparallel & & \shortparallel & & \shortparallel \\
\mathcal{R}\text{-}\mathcal{CONS} & \subset & \mathcal{R}\text{-}\mathcal{CONS}^1 & \subset \cdots \subset & \mathcal{R}\text{-}\mathcal{CONS}^\delta & \subset & \mathcal{R}\text{-}\mathcal{CONS}^{\delta+1} & \subset \cdots \subset & \mathcal{R}\text{-}\mathcal{REL} \\
\cap & & \mid\!\cap & & \mid\!\cap & & \mid\!\cap & & \cap \\
\mathcal{COH} & \subset & \mathcal{COH}^1 & \subset \cdots \subset & \mathcal{COH}^\delta & \subset & \mathcal{COH}^{\delta+1} & \subset \cdots \subset & \mathcal{LIM} \\
\shortparallel & & \shortparallel & & \shortparallel & & \shortparallel & & \shortparallel \\
\mathcal{CONS} & \subset & \mathcal{CONS}^1 & \subset \cdots \subset & \mathcal{CONS}^\delta & \subset & \mathcal{CONS}^{\delta+1} & \subset \cdots \subset & \mathcal{LIM}
\end{array}
$$

Fig. 1. Hierarchies of consistent learning with δ-delay

Moreover, we showed characterization theorems for $\mathcal{CONS}^\delta$ and $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ in terms of complexity. These theorems provide a first explanation for the increase in learning power caused by the δ-delay. On the other hand, the characterization for $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ proved to be very useful for showing the closure of $\mathcal{T}\text{-}\mathcal{CONS}^\delta$ under recursively enumerable unions. Thus, it would be nice to find also a characterization for $\mathcal{R}\text{-}\mathcal{CONS}^\delta$ in terms of complexity. This seems to be a challenging problem.

Finally, further work should solve the problem whether or not $\mathcal{CONS} \setminus \mathcal{R}\text{-}\mathcal{CONS}^\delta \neq \emptyset$ for all $\delta \in \mathbb{N}$. We conjecture the affirmative answer.

Acknowledgments. The authors are very grateful to Kouichi Hirata, Ken Satoh, and Akihiro Yamamoto for their help and support during the editing process of these proceedings and to the anonymous referees for their careful reading.

References

[1] J. M. Barzdin. Две теоремы о предельном синтезе функций. In J. M. Barzdin, editor, Теория Алгоритмов и Программ, volume I, pages 82 – 88. Latvian State University, 1974.

[2] J. M. Barzdin. Inductive inference of automata, functions and programs. In *Proc. of the 20-th International Congress of Mathematicians, Vancouver, Canada*, pages 455–460, 1974. (republished in *Amer. Math. Soc. Transl.* (2) 109, 1977, pp.107-112).

[3] J. M. Barzdin and R. V. Freivalds. On the prediction of general recursive functions. *Soviet Math. Dokl.*, 13:1224–1228, 1972.

[4] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Inform. Control*, 28(2):125–155, June 1975.

[5] M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14(2):322–336, 1967.

[6] R. Freivalds, E. B. Kinber, and R. Wiehagen. How inductive inference strategies discover their errors. *Inform. Comput.*, 118(2):208–226, 1995.

[7] M. A. Fulk. Saving the phenomenon: Requirements that inductive inference machines not contradict known data. *Inform. Comput.*, 79(3):193–209, 1988.

[8] E. M. Gold. Language identification in the limit. *Inform. Control*, 10(5):447–474, 1967.

[9] J. Grabowski. Starke Erkennung. In R. Linder and H. Thiele, editors, *Strukturerkennung diskreter kybernetischer Systeme*, volume 82, pages 168–184. Seminarberichte der Sektion Mathematik der Humboldt-Universität zu Berlin, 1986.

[10] J. Helm. On effectively computable operators. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik (ZML)*, 17:231–244, 1971.

[11] S. Jain, D. Osherson, J. S. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory, second edition.* MIT Press, Cambridge, Massachusetts, 1999.

[12] K. P. Jantke and H.-R. Beick. Combining postulates of naturalness in inductive inference. *Elektronische Informationsverarbeitung und Kybernetik*, 17(8/9):465–484, 1981.

[13] S. Lange and T. Zeugmann. Incremental learning from positive data. *J. of Comput. Syst. Sci.*, 53(1):88–103, 1996.

[14] E. Minicozzi. Some natural properties of strong identification in inductive inference. *Theoret. Comput. Sci.*, 2:345–360, 1976.

[15] P. Odifreddi. *Classical Recursion Theory.* North Holland, Amsterdam, 1989.

[16] P. Odifreddi. *Classical Recursion Theory, Vol. II.* North Holland, Amsterdam, 1999.

[17] D. N. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* MIT Press, Cambridge, Massachusetts, 1986.

[18] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967. Reprinted, MIT Press 1987.

[19] F. Stephan and T. Zeugmann. Learning classes of approximations to non-recursive functions. *Theoret. Comput. Sci.*, 288(2):309–341, 2002.

[20] R. Wiehagen. *Zur Theorie der Algorithmischen Erkennung.* Dissertation B, Humboldt-Universität zu Berlin, 1978.

[21] R. Wiehagen and W. Liepe. Charakteristische Eigenschaften von erkennbaren Klassen rekursiver Funktionen. *Elektronische Informationsverarbeitung und Kybernetik*, 12(8/9):421–438, 1976.

[22] R. Wiehagen and T. Zeugmann. Ignoring data may be the only way to learn efficiently. *J. of Experimental and Theoret. Artif. Intell.*, 6(1):131–144, 1994.

[23] R. Wiehagen and T. Zeugmann. Learning and consistency. In *Algorithmic Learning for Knowledge-Based Systems*, volume 961 of *Lecture Notes in Artificial Intelligence*, pages 1–24. Springer, 1995.

[24] T. Zeugmann. On the nonboundability of total effective operators. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik (ZML)*, 30:169–172, 1984.

Inferability of Closed Set Systems from Positive Data

Matthew de Brecht[1], Masanori Kobayashi[2], Hiroo Tokunaga[2], and Akihiro Yamamoto[1]

[1] Graduate School of Informatics, Kyoto University
Yoshida Honmachi, Sakyo-ku, Kyoto, Japan 606-8501
matthew@mbox.kudpc.kyoto-u.ac.jp, akihiro@i.kyoto-u.ac.jp
[2] Department of Mathematics and Information Science
Tokyo Metropolitan University
1-1 Minami-Ohsawa, Hachiohji-shi, Japan 192-0397
masanori@math.metro-u.ac.jp, tokunaga@comp.metro-u.ac.jp

Abstract. In this paper, we generalize previous results showing connections between inductive inference from positive data and algebraic structures by using tools from universal algebra. In particular, we investigate the inferability from positive data of language classes defined by closure operators. We show that some important properties of language classes used in inductive inference correspond closely to commonly used properties of closed set systems. We also investigate the inferability of algebraic closed set systems, and show that these types of systems are inferable from positive data if and only if they contain no infinite ascending chain of closed sets. This generalizes previous results concerning the inferability of various algebraic classes such as the class of ideals of a ring. We also show the relationship with algebraic closed set systems and approximate identifiability as introduced by Kobayashi and Yokomori [11]. We propose that closure operators offer a unifying framework for various approaches to inductive inference from positive data.

1 Introduction

Inferability of language classes from positive data in the limit, as originally proposed by Gold [8], has become a popular model for language learning after Angluin [1] showed its applications in learning various classes of languages such as pattern languages. Further investigation yielded many more important results and various properties of language classes to guarantee inferability. Applications in learning with logic has been shown to a large extent with work on the inferability of elementary formal systems [3]. For example, a class of elementary formal systems related to context sensitive languages has been shown to be inferable. More recently, Stephan and Ventsov [16] have shown important connections between learning from positive data and computational algebra. Another approach to learning, approximately identifying languages in the limit, proposed by Kobayashi and Yokomori [11], has shown connections with topological spaces. These above results have shown an important relationship between

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 265–275, 2007.

language learning, logic, and various branches of mathematics. One of the major goals of our research is to find a unifying framework for these various approaches using tools from universal algebra.

In this paper, we focus on the fact that various properties of language classes that are important in inferability from positive data have close connections with well known properties used to characterize algebraic structures. For example, the class of all ideals of a recursive ring is inferable from positive data if and only if the ring is Noetherian [16]. Kobayashi et al. [10] expanded on these results and showed the connection between finite elasticity and Noetherian rings and also discussed the inferability of unions of classes of ideals.

We generalize these results by investigating the inferability of language classes defined by closure operators. We show that several properties of classes of languages useful in inductive inference have close parallels in closed set systems, and that the equivalence between inferability from positive data and being Noetherian applies to a broad family of structures called algebraic closed set systems. In addition, we show the close relationship between Kobayashi and Yokomori's notion of upper-best approximate identifiability and algebraic closed set systems. Our results clarify the relationship between several previously proposed frameworks, and allow the results to be applied to a broad class of problems.

In Section 2 we review basic definitions and theorems concerning the inferability of classes of sets from positive data. In Section 3, we introduce closed set systems and give theorems about their inferability from positive data. Section 4 briefly discusses the inferability of unions of closed set systems. Section 5 focuses on algebraic closed set systems, and shows the connections with approximate identifiability. We discuss our results in Section 6.

2 Inferability from Positive Data

First we will introduce some basic definitions and theorems about inferring classes of recursive sets from positive data. In this paper we consider indexed classes of recursive sets over some universe U. We assume that for an indexed class of recursive sets $\mathcal{L}$, there is a recursive characteristic function f such that:

$$f(n,s) = \begin{cases} 0 \text{ if } s \notin L_n, \\ 1 \text{ if } s \in L_n \end{cases}$$

for $L_n \in \mathcal{L}$. Such a class $\mathcal{L}$ is sometimes called uniformly recursive. We use $\subseteq$ to represent set inclusion and $\subset$ to represent strict set inclusion.

A *language* over a set U is any subset of U. Given a language L, an infinite sequence $\sigma = s_0, s_1, s_2, \ldots$ such that $L = \{s_i \mid s_i \in \sigma\}$ is called a *positive presentation* of L. An *inference algorithm* M is an algorithm that incrementally receives elements of a positive presentation of a language and occasionally outputs a positive integer representing the index of a language. We say that the output of M *converges* to an integer j if M's output is infinite and all but finitely many integers equal j, or if M's output is finite and the last integer equals j. If for any positive presentation of a set L, the output of M converges to an integer

j such that $L_j = L$, then we say that M *infers* L *from positive data*. Note that j need not be the same for every presentation of L. If an inference algorithm M exists that infers from positive data every $L \in \mathcal{L}$, then we say that $\mathcal{L}$ *is inferable from positive data*[1].

A *finite tell-tale* of $L \in \mathcal{L}$ is a finite subset T of L such that for all $L' \in \mathcal{L}$, $T \subseteq L'$ implies that $L' \not\subset L$ (Angluin [1]). A *characteristic set* of $L \in \mathcal{L}$ is a finite subset F of L such that for all $L' \in \mathcal{L}$, $F \subseteq L'$ implies that $L \subseteq L'$ (Angluin [2], Kobayashi [12]). A class $\mathcal{L}$ of sets has *infinite elasticity* if and only if there exist infinite sequences $L_1, L_2, L_3, \ldots$ of sets in $\mathcal{L}$ and elements $s_0, s_1, s_2, \ldots$ such that $\{s_0, \ldots, s_{n-1}\} \subseteq L_n$ but $s_n \notin L_n$. $\mathcal{L}$ has *finite elasticity* if and only if it does not have infinite elasticity (Wright [17], Motoki et al. [13]).

In the following three theorems, $\mathcal{L}$ is an indexed class of recursive languages.

Theorem 1 (Angluin [1]). *$\mathcal{L}$ is inferable from positive data if and only if there exists a procedure to enumerate the elements of a finite tell-tale of every $L \in \mathcal{L}$.*

Theorem 2 (Angluin [2], Kobayashi [12]). *If every $L \in \mathcal{L}$ has a characteristic set, then $\mathcal{L}$ is inferable from positive data.*

Theorem 3 (Wright [17]). *If $\mathcal{L}$ has finite elasticity, then $\mathcal{L}$ is inferable from positive data.*

It can be shown that if $\mathcal{L}$ has finite elasticity, then every $L \in \mathcal{L}$ has a characteristic set. Furthermore, if $\mathcal{L}$ is an indexed class of recursive languages and $L \in \mathcal{L}$ has a characteristic set, then a procedure exists that enumerates the elements of a finite tell-tale of L. However, the existence of an algorithm that enumerates a finite tell-tale of all $L \in \mathcal{L}$ does *not* imply that all $L \in \mathcal{L}$ have a characteristic set, nor does each $L \in \mathcal{L}$ having a characterisitic set imply that $\mathcal{L}$ has finite elasticity.

3 Inferability of Closed Set Systems

A mapping $C\colon 2^U \to 2^U$ over some set U is called a *closure operator* if and only if for all $X \in 2^U$

1. $X \subseteq C(X)$,
2. $C(C(X)) = C(X)$, and
3. $X \subseteq Y$ implies $C(X) \subseteq C(Y)$.

If $X = C(X)$, then X is called a *closed set*. If a closed set $X = C(Y)$ for some finite set Y, then X is called a *finitely generated* closed set. The class of all closed sets of a closure operator is called a *closed set system*. It can be shown that a class of sets is a closed set system if and only if it is closed under arbitrary intersections. In the following, we will always assume that U is a countable (possibly infinite) set.

[1] Classes that fulfill this requirement for inferability from positive data are sometimes called *Ex-learnable* (see [9]).

A common example of a closed set system is the class of closed sets of some topological space. However, the definition above is more general than a topological closure operator, since it does not require the finite union of closed sets to be closed. Another example of a closure operator is Tarski's immediate consequence operator in logical inference (which is actually an *algebraic* closure operator, defined as below).

Theorem 4. *Given a closed set system $\mathcal{L}$, the following are equivalent for all closed sets $L \in \mathcal{L}$:*

1. *L is finitely generated.*
2. *L has a characteristic set.*
3. *L has a finite tell-tale.*

Proof. For $(1 \Rightarrow 2)$, assume $L = C(X)$ for some finite subset X. Then for any other closed set $L' \in \mathcal{L}$, $X \subseteq L'$ implies $L = C(X) \subseteq C(L') = L'$. Therefore, X is a characteristic set of L.

The $(2 \Rightarrow 3)$ part follows from the definitions of characteristic set and finite tell-tale.

For $(3 \Rightarrow 1)$, given any finite tell-tale T of L, the closure of T can not be a strict subset of L. This means that $C(T) = L$, and therefore L is finitely generated by T. □

This result gives us the following nice characterization of the learnability of closed set systems from positive data.

Theorem 5. *A closed set system of indexed recursive sets is inferable from positive data if and only if every closed set is finitely generated.*

Proof. Follows directly from Theorems 1, 2, and 4. □

Example 1. Let $U = \mathbf{N}$, where $\mathbf{N}$ is the set of all non-negative integers, and let $\mathcal{L} = \{L_i \,|\, i \geq 1\} \cup \{L_\omega\}$, where $L_i = \{x \,|\, 1 \leq x \leq i\}$ and $L_\omega = \{x \,|\, x \geq 0\}$. Then $\mathcal{L}$ is a closed set system with closure operator $C(\cdot)$, where for any subset X of U, $C(X)$ is defined to be the smallest $L \in \mathcal{L}$ containing X. Since $L_\omega = C(\{0\})$ and $L_i = C(\{i\})$ for $i \geq 1$, all of the closed sets are finitely generated. Therefore $\mathcal{L}$ is inferable from positive data.

Next we give another condition equivalent to the existence of a finite tell-tale within a closed set system, which will be useful in our discussion of algebraic closure operators below.

Theorem 6. *A closed set $L \in \mathcal{L}$ has a finite tell-tale if and only if there is no infinite family of closed sets $\{L_i \,|\, i \geq 0\}$ such that*

1. *$L_i \subset L_j$ for $i < j$, and*
2. *$L = \bigcup_{i=0}^{\infty} L_i$.*

Proof. Assume that T is a finite tell-tale of L, and that $\{L_i \mid i \geq 0\}$ is an infinite family of closed sets that fulfills conditions 1 and 2. Since T is finite, there must be some L_i such that $T \subseteq L_i$. However, since $L_i \subset L$ for all i, this contradicts the definition of a finite tell-tale.

Conversely, if L does not have a finite tell-tale, then we can show that L equals an infinite strictly ascending chain composed of closed strict subsets of L. Let $\{s_0, s_1, s_2, \ldots\}$ be a complete enumeration of the elements of L. Since $\{s_0\}$ is not a finite tell-tale tale of L, $C(\{s_0\}) = L_0 \subset L$. For the $(i+1)$-th step, let s_{i+1} be the element in the enumeration with the smallest index such that $s_{i+1} \in L - L_i$, then $L_i \subset C(\{s_0, \ldots, s_{i+1}\}) = L_{i+1} \subset L$. Repeating this process infinitely produces a strictly ascending chain of closed sets L_i $(i \geq 0)$ such that $L_i \subset L$ for all i and $L = \bigcup_{i=0}^{\infty} L_i$. □

The closed sets of a closed set system form a complete lattice ordered by set inclusion, with joins and meets defined by

$$\bigvee_{i \in I} C(X_i) = C\left(\bigcup_{i \in I} X_i\right)$$

and

$$\bigwedge_{i \in I} C(X_i) = \bigcap_{i \in I} C(X_i).$$

In fact, every complete lattice is isomorphic to some closed set system (see Burris et al. [5] for a proof).

An element a in a lattice L is *compact* if and only if whenever $a \leq \bigvee A$ $(A \subseteq L)$ there exists a finite subset F of A such that $a \leq \bigvee F$. This definition is equivalent to the definition of compact sets within the complete lattice of open sets of a topological space. It is also used extensively in domain theory where compact elements are elements that are way below themselves.

Corollary 1. *Let $\mathcal{L}$ be a closed set system of indexed recursive sets. If every $L \in \mathcal{L}$ is compact, then $\mathcal{L}$ is inferable from positive data.*

Proof. If a closed set equals the union of an infinite strictly ascending chain of closed sets, then it will not be compact. Therefore, the corollary follows directly from Theorems 4, 5, and 6. □

Note that this is not a necessary requirement for inferability. For example, L_ω in Example 1 is not compact.

We will call a closed set system *Noetherian* if and only if it contains no infinite strictly ascending chains of closed sets.

Theorem 7. *A closed set system $\mathcal{L}$ is Noetherian if and only if it has finite elasticity.*

Proof. Assume that the closed sets $L_1, L_2, L_3, \ldots$ and elements $s_0, s_1, s_2, \ldots$ demonstrate the infinite elasticity of $\mathcal{L}$. From the definition of infinite elasticity,

$\{s_0, \ldots, s_{n-1}\} \subseteq C(\{s_0, \ldots, s_{n-1}\}) \subseteq L_n$, but since $s_n \notin L_n$, $C(\{s_0, \ldots, s_{n-1}\}) \neq C(\{s_0, \ldots, s_n\})$. Therefore, $C(\{s_0, \ldots, s_{n-1}\}) \subset C(\{s_0, \ldots, s_n\})$. Since n is arbitrary, $C(\{s_0, \ldots, s_i\})$ $(i \geq 1)$ is an infinite strictly ascending chain, and thus $\mathcal{L}$ is not Noetherian.

Next, assume that $L_1 \subset L_2 \subset \cdots$ is an infinite strictly ascending chain of closed sets in $\mathcal{L}$. Let s_0 be any element of L_1, and let s_n be an element of $L_{n+1} - L_n$ for $n \geq 1$. Then obviously the closed sets $L_1, L_2, L_3, \ldots$ and elements $s_0, s_1, s_2, \ldots$ show that $\mathcal{L}$ has infinite elasticity. □

Corollary 2. *If a closed set system of indexed recursive sets is Noetherian, then it is inferable from positive data.*

4 Inferability of Unions of Closed Set Systems

In this section we will consider the inferability of unions of closed set systems from positive data. First, for closed set systems $\mathcal{L}_1$ and $\mathcal{L}_2$, we define the union $\mathcal{L}_1 \tilde{\cup} \mathcal{L}_2$ to be:

$$\mathcal{L}_1 \tilde{\cup} \mathcal{L}_2 = \{L_1 \cup L_2 \mid L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2\}.$$

To avoid confusion, in this section we will refer to $\mathcal{L}_1 \tilde{\cup} \mathcal{L}_2$ as the "union" of two closed set systems, and refer to the typical union of two sets $X \cup Y$ as the "set union". Note that in general $\mathcal{L}_1 \tilde{\cup} \mathcal{L}_2$ will not be a closed set system.

Theorem 8 (Wright [17]). *If $\mathcal{L}_1$ and $\mathcal{L}_2$ have finite elasticity, then $\mathcal{L}_1 \tilde{\cup} \mathcal{L}_2$ has finite elasticity.*

Corollary 3. *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be Noetherian closed set systems of indexed recursive sets. Then $\mathcal{L}_1 \tilde{\cup} \mathcal{L}_2$ is inferable from positive data.*

Given a closed set system $\mathcal{L}$, we follow Shinohara and Arimura [15] and define the unbounded union $\mathcal{L}^*$ to be:

$$\mathcal{L}^* = \{\bigcup_{i \in I} L_i \mid L_i \in \mathcal{L}, I \subset \mathbf{N}, 1 \leq \#I < \infty\},$$

where $\#I$ denotes the cardinality of the set I, and $\mathbf{N}$ is the set of all non-negative integers.

In some cases, such as the class of closed sets of a topology, the set union of two closed sets is closed, and so $\mathcal{L} = \mathcal{L}^*$. However, in general the set union of two closed sets is not necessarily closed.

Theorem 9. *Let $\mathcal{L}$ be a closed set system of indexed recursive sets. Then $\mathcal{L}^*$ is inferable from positive data if and only if every closed set $L \in \mathcal{L}$ is equal to a union of closed sets generated from a single element.*

Proof. Assume that all closed sets of $\mathcal{L}$ are equal to a finite union of closed sets generated from a single element. Then for any $L \in \mathcal{L}^*$, $L = C(\{s_0\}) \cup \cdots \cup$

$C(\{s_m\})$, where $C(\{s_i\})(1 \leq i \leq m)$ are closed sets in $\mathcal{L}$. Then $\{s_0, \ldots, s_m\}$ is a characteristic set of L, for assume that some $L' = C(\{t_1\}) \cup \cdots \cup C(\{t_n\}) \in \mathcal{L}^*$ exists and $\{s_0, \ldots, s_m\} \subseteq L'$. Then for any element $x \in L$, $x \in C(\{s_i\})$ for some s_i, and $s_i \in C(\{t_j\})$ for some t_j. Since $\{s_i\} \subseteq C(\{t_j\})$ implies $C(\{s_i\}) \subseteq C(C(\{t_j\})) \subseteq C(\{t_j\})$, it follows that $x \in L'$, and since x was arbitrary, $L \subseteq L'$.

Next, assume that some closed set $L \in \mathcal{L}$ is not equal to a finite union of closed sets generated from a single element. For any finite subset $\{s_0, \ldots, s_n\}$ of L, $C(\{s_0\}) \cup \cdots \cup C(\{s_n\}) \subseteq L$, but since L does not equal such a union of closed sets, $C(\{s_0\}) \cup \cdots \cup C(\{s_n\}) \subset L$. However, since both L and $C(\{s_0\}) \cup \cdots \cup C(\{s_n\})$ are in $\mathcal{L}^*$, L does not have a finite tell-tale within $\mathcal{L}^*$. □

A class of languages is said to have *finite thickness* if no element is contained within an infinite number of languages [1]. It is easy to see that, if a class of languages has finite thickness, then it has finite elasticity. An *infinite anti-chain with respect to set inclusion* is an infinite family of sets $\mathcal{S}$ such that for any $X \in \mathcal{S}$ and $Y \in \mathcal{S}$ such that $X \neq Y$, X and Y are incomparable, meaning neither $X \subseteq Y$ nor $Y \subseteq X$ holds. Shinohara and Arimura[15] showed that if a class of languages $\mathcal{L}$ has finite thickness and contains no infinite anti-chains, then $\mathcal{L}^*$ is inferable from positive data. The following theorem shows the connection between their result and Theorem 8.

Theorem 10. *Let $\mathcal{L}$ be a Noetherian closed set system that contains no infinite anti-chains with respect to set inclusion. Then every closed set in $\mathcal{L}$ is equal to a finite union of closed sets generated from a single element.*

Proof. Let $\mathcal{L}$ be such a closed set system and assume $L \in \mathcal{L}$ is not equal to a finite union of closed sets generated from a single element. Let s_0, s_1, ... be an enumeration of the elements of L. $C(\{s_0\}) \subset L$, so find the next element in the enumeration with the smallest index that is not contained in $C(\{s_0\})$ (without loss of generality, assume this element is s_1). Therefore, $C(\{s_1\})$ is either a strict superset of $C(\{s_0\})$, or else they are incomparable. Since $C(\{s_0\}) \cup \cdots \cup C(\{s_n\}) \subset L$ for any $s_0, \ldots, s_n \in L$, this process of finding supersets or incomparable sets can be continued infinitely. However, this implies that there is either an infinitely increasing chain of closed sets or an infinite anti-chain in $\mathcal{L}$, and in either case there is a contradiction. □

5 Inferability of Algebraic Closed Set Systems

In this section we will investigate the inferability from positive data of a special case of closed set systems called algebraic closed set systems. A closure operator $C(\cdot)$ is called *algebraic* if and only if for every closed set $C(X)$,

$$C(X) = \bigcup \{C(Y) \mid Y \subseteq X, Y \text{ is finite}\}.$$

Algebraic closure operators are also sometimes referred to as *finitary* closure operators. An *algebraic closed set system* is a class of sets which are precisely the closed sets of some algebraic closure operator. The closed sets of an algebraic closed set system form an *algebraic lattice*, which is a complete lattice in which every element is the supremum of compact elements. In algebraic closed set systems, the compact closed sets are precisely the finitely generated closed sets.

Example 2. Let R be a ring and for $X \subseteq R$, let $I(X)$ be the ideal generated from X. Then $I(\cdot)$ is a closure operator. Furthermore, for any $a \in I(X)$, a is equal to some finite sum $r_1x_1s_1+\cdots+r_nx_ns_n$ ($r_i, s_i \in R, x_i \in X$), so $a \in I(\{x_1, \ldots, x_n\})$. Therefore $I(\cdot)$ is an algebraic closure operator. Similarly, for any group G and $X \subseteq G$, define $S(X)$ to be the subgroup generated by X. Then $S(\cdot)$ is an algebraic closure operator.

Algebraic closed set systems are quite prominent in universal algebra. It is well known that for any (finitary) algebra $\mathcal{A}$, the subalgebras of $\mathcal{A}$ form an algebraic closed set system. It can also be shown that for any algebraic closure operator $C(\cdot)$, an algebra $\mathcal{A}$ exists such that the closed sets of $C(\cdot)$ are precisely the subalgebras of $\mathcal{A}$. See, for example, Birkhoff [4], Burris et al. [5], and Cohn [6] for a proof of the above, and for more examples of algebraic closed set systems. The system defined in Example 1 is a good example of a closed set system that is *not* algebraic.

Theorem 11 (see Birkhoff [4], Cohn [6]). *A closed set system is algebraic if and only if the union of any ascending chain (with respect to $\subseteq$) of closed sets is closed.*

It follows that any Noetherian closed set system is algebraic, although the converse does not necessarily hold. However, we have the following interesting result concerning the inferability of algebraic closed set systems from positive data.

Theorem 12. *An algebraic closed set system of indexed recursive sets is inferable from positive data if and only if it is Noetherian.*

Proof. The if part is obvious, so for the only if part, assume that an algebraic closed set system $\mathcal{L}$ is not Noetherian. Then there is an infinite class of languages L_i ($i \geq 0$) such that $L_i \subset L_j$ for $i < j$. However, from Theorem 11, $\bigcup_{i=0}^{\infty} L_i$ is a closed set, and by Theorem 6 it does not have a finite tell-tale. □

Since the class of ideals of a ring form an algebraic closed set system, this theorem generalizes the results from Stephan and Ventsov [16], which showed that the class of ideals of a recursive ring is inferable from positive data if and only if the ring is Noetherian[2].

Theorem 12 is also related to results about upper-best approximate identifiability of concept classes as defined in Kobayashi and Yokomori [11]. Let $\mathcal{L}$ be

[2] More precisely speaking, Stephan and Ventsov showed that a recursive ring is BC-learnable if and only if the ring is Noetherian. In this paper, we are only concerned with the stronger requirement of EX-learnability. However, since we only consider uniformly recursive classes, BC-learnability implies EX-learnability.

a class of subsets of some universe U and $X \in 2^U$ (not necessarily in $\mathcal{L}$). A set $Y \in \mathcal{L}$ is an *upper-best approximation* of X if and only if $X \subseteq Y$ and for any $S \in \mathcal{L}$, $X \subseteq S$ implies $S \not\subset Y$. $\mathcal{L}$ has *upper best approximation property* (u.b.a.p.) if and only if for any $X \in 2^U$, an upper best approximation of X exists in $\mathcal{L}$. 2^U is *upper-best approximately identifiable in the limit from positive data by* $\mathcal{L}$ if and only if a learning algorithm M exists that when given any positive presentation of any $X \in 2^U$, the sequence outputted by M converges to some n such that $L_n \in \mathcal{L}$ and L_n is an upper best approximation of X.

Theorem 13 (Kobayashi and Yokomori [11]). *An indexed class of recursive sets $\mathcal{L}$ has u.b.a.p. if and only if $\mathcal{L}$ is closed under infinite intersections.*

Theorem 14 (Kobayashi and Yokomori [11]). *Let $\mathcal{L}$ be an indexed class of recursive sets with u.b.a.p. over some universe U. Then 2^U is upper-best approximately identifiable in the limit from positive data by $\mathcal{L}$ if and only if $\mathcal{L}$ contains no infinite strictly ascending chains of sets.*

Corollary 4. *An indexed class of recursive sets $\mathcal{L}$ has u.b.a.p. and upper-best approximately identifies 2^U in the limit from positive data if and only if $\mathcal{L}$ is a Noetherian algebraic closed set system of indexed recursive sets.*

Proof. Note a class of sets is a closed set system if and only if it is closed under arbitrary intersections of closed sets. □

Corollary 4 has the interesting consequence that for any class $\mathcal{L}$ with u.b.a.p that can upper-best approximately identify arbitrary sets, the sets of $\mathcal{L}$ are precisely the subalgebras of some algebra $\mathcal{A}$.

6 Discussion and Conclusion

In this paper we have given some theorems for the inferability of closed set systems from positive data. In particular, we have shown that an algebraic closed set system is inferable from positive data if and only if it is Noetherian. We have also shown that a class of sets with upper-best approximation property that can upper-best approximate arbitrary sets forms an algebraic closed set system. These results give a general framework for the inferability of various types of families of sets. The results on algebraic closed set systems are applicable to inferability of classes of subalgebras, ideals (of rings and partially ordered sets), and congruence relations, to name just a few examples.

One important aspect of learning that we did not discuss is mind change complexity. Stephan and Ventsov [16] show that the optimal mind change bound for learning the class of ideals of the integers is ω, and the optimal bound for ideals of the polynomial ring in n variables is ω^n. Further work by de Brecht and Yamamoto [7] gave upper and lower bounds for the mind change complexity of learning unbounded unions of a restricted class of pattern languages. The upper bound of $\omega^{\omega^\alpha}+\beta$, where $\alpha, \beta < \omega$ are constants determined by properties such as the size of the alphabet, was proven by embedding the class of pattern languages into a closed set system (see the discussion below). The corresponding closure operator and connections with order theory were then used to prove the upper bound. It is a subject for further research to see to what degree these results can be generalized to algebraic closed set systems. Such a generalization would give good insight into the relationship between mind change complexity and the structure of the class of languages.

Although closed set systems are useful for studying the inferability of algebraic structures, closure under arbitrary intersection is too strong of a requirement in most cases. However, some of the theorems in this paper still hold when embedding an arbitrary class of languages into a minimal closed set system. Given a class of recursive languages $\mathcal{L}$ over a universe U, we can define the closure of $X \subseteq U$ to be $C_{\mathcal{L}}(X) = \bigcap\{L \mid X \subseteq L, L \in \mathcal{L}\}$. In the closed set system defined by $C_{\mathcal{L}}(\cdot)$, every $L \in \mathcal{L}$ is a closed set, but not all closed sets are in $\mathcal{L}$ (and are not guaranteed to be recursive). The natural interpretation of $C_{\mathcal{L}}(\cdot)$ is given a sample subset X of an unknown language in $\mathcal{L}$, $C_{\mathcal{L}}(X)$ is the most information about the unknown language that we can unambiguously extract from X. It is easy to see that $L \in \mathcal{L}$ has a characteristic set if and only if it is a finitely generated closed set, and that $\mathcal{L}$ has finite elasticity if and only if the closed set system of $C_{\mathcal{L}}(\cdot)$ is Noetherian. However, Theorems 5 and 6 no longer hold since a language in $\mathcal{L}$ that has a finite tell-tale but no characteristic set will be equal to the union of an infinitely increasing chain of closed sets.

The closure operator defined above is reminiscent of finding the least Herbrand model of a logic program, and closure operators are also used in defining abstract logics. This suggests that closure operators may be useful in clarifying the relationship between deductive and inductive reasoning.

References

1. D. Angluin: Inductive Inference of Formal Languages from Positive Data. *Information and Control*, **45** (1980) 117–135.
2. D. Angluin: Inference of Reversible Languages. *Journal of the ACM*, **29** (1982) 741–765.
3. S. Arikawa, S. Miyano, A. Shinohara, T. Shinohara, A. Yamamoto: Algorithmic Learning Theory with Elementary Formal Systems. *IECE Trans. Inf. & Syst.* Vol. E75-D **4** (1992) 405–414.
4. G. Birkhoff: *Lattice Theory, Third Edition.* American Mathematical Society (1967).
5. S. Burris, H. P. Sankappanavar: *A Course in Universal Algebra.* Springer-Verlag (1981).
6. P. M. Cohn: *Universal Algebra.* Harper & Row (1965).
7. M. de Brecht and A. Yamamoto: Mind Change Complexity of Inferring Unbounded Unions of Pattern Languages From Positive Data. *Proceedings of the 17th International Conference on Algorithmic Learning Theory* (*Lecture Notes in Artificial Intelligence* **4264**) (2006) 124–138.
8. E. M. Gold: Language Identification in the Limit. *Information and Control*, **10** (1967) 447–474.
9. S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory, Second Edition.* MIT Press (1999).

10. M. Kobayashi, H.Tokunaga, and A. Yamamoto: Ideals of Polynomial Rings and Learning from Positive Data (in Japanese). *Proceedings of the 8th Workshop on Information-Based Induction Science* (2005) 129–134.
11. S. Kobayashi, T. Yokomori: On approximately identifying concept classes in the limit. *Proceedings of the 6th International Workshop on Algorithmic Learning Theory* (*Lecture Notes in Artificial Intelligence* **997**) (1995) 298–312.
12. S. Kobayashi: Approximate Identification, Finite Elasticity and Lattice Structure of Hypothesis Space. Technical Report, CSIM 96-04, Dept. of Compt. Sci. and Inform. Math., Univ. of Electro- Communications (1996).
13. T. Motoki, T. Shinohara, and K. Wright: The Correct Definition of Finite Elasticity: Corrigendum to Identification of Unions. *Proceedings of COLT '91* (1991) 375.
14. T. Shinohara: Inductive Inference of Monotonic Formal Systems From Positive Data. *New Generation Computing*, **8** (1991) 371–384.
15. T. Shinohara, H. Arimura: Inductive Inference of Unbounded Unions of Pattern Languages From Positive Data. *Theoretical Computer Science*, **241** (2000) 191–209.
16. F. Stephan and Y. Ventsov: Learning Algebraic Structures from Text. *Theoretical Computer Science*, **268** (2001) 221–273.
17. K. Wright: Identification of Unions of Languages Drawn from an Identifiable Class. *Proceedings of COLT '89*, (1989) 328–333.

An Extended Branch and Bound Search Algorithm for Finding Top-N Formal Concepts of Documents

Makoto Haraguchi and Yoshiaki Okubo

Division of Computer Science
Graduate School of Information Science and Technology
Hokkaido University
N-14 W-9, Sapporo 060-0814, Japan
{mh, yoshiaki}@ist.hokudai.ac.jp

Abstract. This paper presents a branch and bound search algorithm for finding only top N number of extents of formal concepts w.r.t. their evaluation, where the corresponding intents are under some quality control. The algorithm aims at finding potentially interesting documents of even lower evaluation values that belong to some highly evaluated formal concept. The experimental results show that it can effectively find such documents.

1 Introduction

One of the core tasks of *Information Retrieval* (IR) is to effectively find useful and important documents including Web pages. For this purpose, many retrieval engines compute *ranks* of documents and show them in the order of their ranks [3,2,9]. Highly ranked documents are easily checked by users, while documents ranked lower are rarely examined. Any retrieval system based on document ranking has its own ranking scheme. So, even potentially interesting documents are sometimes ranked low and are therefore actually hidden and invisible to users. In this sense, we might be missing many useful documents. If we can make such hidden significant documents visible, our chance to obtain valuable information and knowledge can be enhanced.

The standard approach to cope with the above problem is to use the techniques of *clustering* [1,4] by which we classify various documents into several clusters of similar documents. We pick up a few clusters that seem to be relevant, and then examine them in details to look for interesting documents. However, if the number of clusters is small, clusters tend to be larger ones involving even non-similar documents, and are hard to be examined. Conversely, if we have many number of clusters, it is also hard to check every cluster, although each cluster is smaller and involves only similar documents. Thus, it is not an easy task to have an adequate method for controlling the number of clusters.

For this reason, instead of dividing whole data set into several clusters by clustering, we have developed some strategy in [14,17,15,18] for finding only top N

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 276–288, 2007.

number of clusters of similar documents with respect to their evaluation values reflecting the ranks of documents in them. According to the strategy, the similarity is defined by the standard cosine measure for vector representation of documents and is used to draw edges among documents to form an undirected graph of documents. Then an algorithm has been designed as an extension of branch and bound *maximum clique search algorithms* [8,7] to find top N (pseudo-) cliques as clusters of documents. We have already verified that the algorithm found some clusters in which lowly ranked documents appear in them together with highly ranked documents contributing toward raising the whole evaluation of clusters.

However, as has already been pointed out in the area of conceptual clustering [10,11], as long as the similarity of documents is derived from the cosine measure for vector representation, it is generally difficult to understand the meaning of clusters (cliques in this case) by means of feature terms. In our case of finding interesting documents with lower ranks, the detected lower ranked documents together with highly ranked documents in one cluster are in fact similar vectors. However, it is always difficult to judge if the former and the latter share the same meaning or not. In other words, the conceptual classes they belong to may differ. In order to avoid such a conceptually indistinct argument, we make an informal constraint on the clusters to be obtained as follows:

> *The notion of relevance or interestingness depends only on a conceptual class of documents, not dependent on particular instance documents. Then the clusters we have to find must be concepts of documents that can be definable by means of feature terms.*

As the primary data for a document set is a document-term relationship, we adopt the notion of *Formal Concept Analysis* (FCA) [12,10,11]. As is well known, a *formal concept* consists of two closures, ψA called the *extent* of the concept, a set of documents, and $A = \varphi\psi A$ called the *intent* of the concept, a set of terms, where A is a term set and (φ, ψ) is a Galois connection defined from the document-term relationship. Then, the extent ψA has a definition that $d \in \psi A$ iff $A \subseteq \varphi d$, where RHS means that a document d has all terms in A. Thus, if some higher-ranked documents and lower-ranked ones share a set of terms, they could form an extent of a concept, that is, a conceptual cluster of documents.

It is also well known that formal concepts can be computed by finding maximal *bipartite cliques* of a bipartite graph or equivalently by finding *closures* of documents or terms. Therefore, keeping the evaluation scheme for extents as clusters of documents, it can be a strategy to find only Top-N extents by using some very fast enumeration algorithm, *LCM* [13] for instance, for finding all the closures.

The problem for such an approach is however that the number of possible extents is still large. Particularly, there exist a numerous number of extents of concepts whose corresponding intents are very smaller set of terms. For smaller intents we have, the extents tend to be larger sets of documents and to involve documents with less similarity. In other words, the quality of those extents becomes worse. For the reason, we revise in this paper our former algorithms [17,15,18] so that we try to find only Top-N extents w.r.t. the same evaluation schema for clusters, keeping the quality of extents by a given lower bound for the corresponding intents. Our method is summarized as follows:

FC1: Evaluation
Extents of formal concepts are evaluated by some *monotone function*. The evaluation becomes higher, as the extents grow as sets of documents, and as each document in them shows higher rank.

FC2: Graph Formation under Static Quality Control
Two documents are judged similar if they share at least a given number of common terms. We draw an edge between any similar two documents, and form a weighted undirected graph of documents, where each document is assigned a weight based on its rank. It should be noted here that any extent with enough quality of intent is always a clique in the graph.

FC3: Extent Search under Dynamic Quality Control
To enumerate only Top-N extents, closures of documents, our algorithm adopts again a branch and bound method, where

- **Candidate Closures of Documents:** a list of candidate top N closures is always kept,
- **Branch and Bound Pruning due to Monotone Evaluation:** for any search node, a closure of documents, whose evaluation value can never become larger than the minimum of those candidates, we cut off the nodes below, and
- **Dynamic Quality Control:** for any search node whose corresponding intent has less number of feature terms than a given lower bound, we also cut off the nodes below.

Clearly the two pruning rules are safe ones never missing any of Top-N extents satisfying the requirements.

In the graph formation process, we can exclude document pairs *in advance* which are never included in any extent with enough quality of intent. Furthermore, a theoretical property of cliques can provide us several *upper-bounds* of evaluation values for extents. For example, we can obtain a tight upper-bound with a *sequential approximate coloring* [7,8,17]. Based on the bounds, we can prune many useless extents which never be in Top-N. Thus, the clique search-based approach enables us to efficiently find Top-N extents.

In addition, we can utilize another pruning rule based on a simple theoretical property of formal concepts. It is effective for reducing generation of redundant extents, that is, duplications of already generated ones.

Top-N extents actually obtained are highly dependent on weighting functions for documents. If we always prefer an extent with higher-ranked documents, they might be assigned particularly larger weights and the other documents relatively small weights. In such a case, lower-ranked documents will always be clustered with some higher-ranked ones. On the other hand, if we assign a uniform weight to each document, an extent without higher-ranked documents could be obtained as one of the top N.

Now let us go on to the details of our algorithm in Section 3. Our experimental results are presented in Section 4. Particularly in Section 5, we compare our algorithm with a very fast enumerator of closures from a computational viewpoint. Finally in Section 6, we discuss some future works.

2 Preliminaries

In this section, we introduce some terminologies used throughout this paper.

Let $\mathcal{O}$ be a set of *objects* (individuals) and $\mathcal{F}$ a set of *features* (attributes). A pair of $\mathcal{O}$ and $\mathcal{F}$, $<\mathcal{O},\mathcal{F}>$ is called a *formal context*. Under a formal context $<\mathcal{O},\mathcal{F}>$, each object in $\mathcal{O}$ is represented as the set of all features in $\mathcal{F}$ which the object has.

Given a formal context $<\mathcal{O},\mathcal{F}>$, for a set of objects $O \subseteq \mathcal{O}$ and a set of features $F \subseteq \mathcal{F}$, we define two mappings $\varphi : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{F}}$ and $\psi : 2^{\mathcal{F}} \rightarrow 2^{\mathcal{O}}$ as follows.

$$\varphi(O) = \{f \in \mathcal{F} \mid \forall o \in O, f \in o\} = \bigcap_{o \in O} o \quad \text{and}$$

$$\psi(F) = \{o \in \mathcal{O} \mid F \subseteq o\}.$$

The former computes the feature set shared by every object in O. The latter, on the other hand, returns the set of objects with F.

Based on the mappings, a *formal concept* (FC) under the formal context is defined as a pair of object set and feature set, (O, F), where $O \subseteq \mathcal{O}$, $F \subseteq \mathcal{F}$, $\varphi(O) = F$ and $\psi(F) = O$. Especially, O and F are called the *extent* and *intent* of the concept, respectively. From the definition, it is obvious that $\psi(\varphi(O)) = O$ and $\varphi(\psi(F)) = F$. That is, a formal concept is defined as a pair of *closed* sets of objects and features under the mappings.

Let $\mathcal{T}$ be a set of *feature terms*. A document d can be represented as a set of feature terms in $\mathcal{T}$ appearing in the document, that is, $d \subseteq \mathcal{T}$. For a set of documents $\mathcal{D}$, therefore, we can consider a formal concept (D, T) under the formal context $<\mathcal{D},\mathcal{T}>$. The extent D is regarded as a cluster of documents and the intent is the feature terms which are shared by the documents in D.

Let $G = (V, E, w)$ be a weighted undirected graph, where V is a set of vertices, E ($\subseteq V \times V$) a set of edges and w a mapping which assigns a (positive) weight to each vertex. For a set of vertices $V' \subseteq V$, the graph G' defined as $G' = (V', E \cap (V' \times V'), w)$ is called a *subgraph of G induced by V'*. If a subgraph $G' = (V', E', w)$ of G is complete, then G' is called a *clique* in G, where the clique is often denoted by simply V'.

3 Finding Document Clusters Based on Formal Concept Analysis

In order to provide clear meanings of clusters, we try to improve our previous method of pinpoint clustering with *Formal Concept Analysis* (*FCA*) [12]. Formal Concept Analysis is a theory of data analysis which identifies *conceptual structures among objects* (individuals).

3.1 Document Clusters as Formal Concepts

Let $\mathcal{T}$ be a set of feature terms and $\mathcal{D}$ a set of documents. We consider a document cluster based on a formal concept under a formal context $\mathcal{C} =< \mathcal{D}, \mathcal{T} >$. More precisely speaking, for a formal concept (D, T) under $\mathcal{C}$, the extent D is regarded as a document cluster. It should be noted here that we can clearly explain why the documents in the cluster are grouped together. Each document in D shares the set of feature terms T and any other document never contains T. In this sense, D can form a meaningful grouping (cluster) of documents. Thus, with the help of formal concepts under $\mathcal{C}$, we can explicitly consider the meanings of clusters based on their intents. We call this kind of clusters *Formal Concept-Based clusters* (*FC*-clusters in short).

The meaningfulness of FC-cluster is affected by both of its intent and extent. For example, a cluster with smaller intent might be unconvincing because the evidence for the grouping seems to be weak, though its extent tends to be larger. Conversely, although a cluster with larger intent might have more convincing evidence, its extent tends to be smaller. Thus, it is required to control the quality of FC-clusters in order to obtain useful ones. From these observations, we formalize FC-clusters to be found as follows:

Quality Control : Constraint on Intents: Quality of FC-cluster to be found is controlled by imposing a constraint on intent. As such a constraint, we give a threshold δ for evaluation value of intent. For an FC-cluster, if the evaluation value of the intent is greater than or equal to δ, then the cluster is said to be *δ-valid.* As we will see later, the constraint can work statically and dynamically to prune useless search.

Preference in Extents: Among the δ-valid FC-clusters, we prefer ones with higher evaluation values. Especially, we try to extract only clusters which have Top-N evaluation values.

In order to evaluate intents and extents, we assume that each document d and feature term t have their (positive) weights which are referred to as $w(d)$ and $w(t)$, respectively. For example, each document is given a weight based on its rank assigned by an information retrieval system. Furthermore, a weight of feature term might be defined as the *inverted document frequency*(IDF) of the term. Then, an evaluation function might be defined as the *sum of weights* of documents (terms) in each extent (intent).

Thus, intents and extents of formal concepts can be evaluated from several viewpoints. We can actually define various evaluation functions for them. From the computational point of view, however, a function which behaves *monotonically* according to expansion of extents (intents) is strongly preferred. More precisely speaking, we prefer an evaluation function f such that for any set S and its superset S', $f(S) \leq f(S')$ holds. It is obvious that the above evaluation function based on the sum of weights behaves monotonically. The reason why such a function is preferable will become clear shortly.

We can now define our problem of finding Top-N δ-valid FC-clusters as follows:

[Given] $\mathcal{T}$: a set of feature terms
$\mathcal{D}$: a set of documents each of which is represented as a subset of $\mathcal{T}$
w_d : an evaluation function for sets of documents
w_t : an evaluation function for sets of feature terms
δ : a threshold for the minimum evaluation value for intent
[Find] the set of formal concepts $\{(D, T)\}$ such that the evaluation value of D, $w_d(D)$, is in the top N among δ-valid formal concepts under the formal context $< \mathcal{D}, \mathcal{T} >$

We present in the next subsection our *clique search-based* algorithm for the problem.

3.2 Algorithm for Finding Top-N δ-Valid FC-Clusters by Clique Search

Top-N δ-valid FC-clusters can be extracted by finding certain *cliques* in a weighted undirected graph.

Graph Construction:
Given a formal context $\mathcal{C} =< \mathcal{D}, \mathcal{T} >$ and a threshold δ for the quality of intents, we first construct a weighted undirected graph $G = (\mathcal{D}, V, w)$, where V is defined as

$$V = \{(d_i, d_j) \mid d_i, d_j \in \mathcal{D}(i \neq j) \wedge w_t(d_i \cap d_j) \geq \delta\}.$$

That is, if a pair of documents share a set of feature terms whose evaluation value is greater than or equal to δ, then they are connected by an edge. For example, consider a set of documents

$$\begin{aligned} \mathcal{D} = \{\ & d_1 = \{a, b, c, d\}, \\ & d_2 = \{a, b, f\}, \\ & d_3 = \{b, c, f\}, \\ & d_4 = \{b, d, e\}, \\ & d_5 = \{a, b, e, g\}\ \}. \end{aligned}$$

Let w_t be an evaluation function defined as $w_t(T) = |T|$ for a term set T. Under the setting of $\delta = 2$, we have a weighted undirected graph

$$G = (\mathcal{D}, \{(d_1, d_2), (d_1, d_3), (d_1, d_5), (d_2, d_3), (d_2, d_5), (d_4, d_5)\}, w).$$

From the definition of the graph G, any extent of *delta*-valid FC can be found as a clique in G. In other words, finding only cliques in G is sufficient for our task. In the example, although $(\{d_1, d_2, d_3, d_5\}, \{b\})$ is an FC, it is not δ-valid. Since the extent $\{d_1, d_2, d_3, d_5\}$ is not a clique in G, we do not have to examine it in our search. In our graph construction process, thus, many useless extents can be eliminated *statically* based on the threshold δ.

It should be noted here that any clique does not always correspond to an extent of formal concept. In the above example, $\{d_1, d_2, d_3\}$ is a maximal clique, but it is not an extent of formal concept. Each d_i in the clique shares the term b, that is, $\varphi(\{d_1, d_2, d_3\}) = \{b\}$. However, since $\psi(\{b\}) = \{d_1, d_2, d_3, d_5\}$ holds, $\{d_1, d_2, d_3\}$ is not a closed set. That is, there exists no FC whose extent is the clique $\{d_1, d_2, d_3\}$.

Search Strategy:
From the graph G, we try to extract δ-valid FCs whose extents have Top-N evaluation values. Our algorithm finds Top-N FCs with *depth-first branch-and-bound search strategy.* During our search, we maintain a list which stores Top-N FCs among ones already found. That is, the list keeps *tentative* Top-N FCs.

Basically speaking, for each clique (a set of documents) Q, $\varphi(Q)$ is computed and then evaluated. If $w_t(\varphi(Q)) \geq \delta$, the set of feature terms $\varphi(Q)$ can become the intent of a δ-valid FC with its extent $\psi(\varphi(Q))$. Then, the tentative Top-N list is adequately updated for the obtained FC, $(\psi(\varphi(Q)), \varphi(Q))$. The procedure is iterated until no Q remains to be examined.

Each clique Q in G is explored in *depth-first manner*. For a clique Q, a new clique can be obtained by adding a vertex $\alpha \in cand(Q)$ to Q, where $cand(Q)$ is the set of vertices adjacent to all vertices in Q. Starting with the initial Q of the empty set, Q is expanded step by step in depth-first manner.

As has been mentioned above, in our algorithm, we assume that as an intent becomes larger under set inclusion, our evaluation function for intents behaves monotonically. Especially, we prefer a monotonically increasing function. As long as we evaluate intents with such a monotone function, we can utilize a simple pruning rule. That is, for a clique Q, if $w_t(\varphi(Q)) < \delta$ holds, then we can never obtain δ-valid formal concepts from any extension of Q, because its corresponding intent becomes smaller. Therefore, we can immediately stop expanding Q and backtrack. Thus, the quality of intents is *dynamically* controlled in our search.

Moreover, since we also assume an evaluation function for extents is monotonically increasing according to expansion of extents, we can enjoy a pruning based on tentative Top-N clusters. From a theoretical property of cliques, if $w_d(Q) + w_d(cand(Q))$ is less than the minimum evaluation value of extents in the tentative Top-N list, we do not have to examine any expansion of Q. $w_d(Q) + w_d(cand(Q))$ gives an upper bound of extent values obtained by expanding Q. If the upper bound is less than the tentative minimum value, any extension of Q is no longer useful and can be pruned safely.

In addition to the prunings, our current algorithm also utilizes another pruning rule based on a theoretical property of formal concepts. It is quite effective for reducing generation of useless FCs, that is, duplications of already generated ones.

The pruning rule is based on the following simple property:

Observation 1. *Let Q be a set of documents. For any documents α and β, if $\alpha \in \psi(\varphi(Q \cup \{\beta\}))$ and $\beta \in \psi(\varphi(Q \cup \{\alpha\}))$, then $\psi(\varphi(Q \cup \{\alpha\})) = \psi(\varphi(Q \cup \{\beta\}))$.*

For a clique Q, let W_Q be the set of vertices already used to expand Q. Assume we try to expand Q with a vertex α. From the above property, if there exists a vertex $w \in W_Q$ such that $\alpha \in \psi(\varphi(Q \cup \{w\}))$ and $w \in \psi(\varphi(Q \cup \{\alpha\}))$, then we do not have to expand Q with α. Expanding Q with α will generate a duplication of an extent already generated. Therefore, the search branch can be pruned safely.

A pseudo-code of our algorithm is shown in Figure 1.

```
[Input]  δ : a threshold for the minimum evaluation value for intent.
         (D, E, w) : an weighted undirected graph constructed under δ,
                     where D is an ordered set of documents each of which
                     is represented as a set of feature terms in T.
         w_d : an evaluation function for document sets.
         w_t : an evaluation function for term sets.
[Output]  FC : the set of Top-N δ-valid formal concepts under < D, T >
--------------------------------------------------------------------------
procedure main() :
  FC ← φ ;
  for each d ∈ D in predefined order do
    begin
      TopNFCSearch({d}, N(d), d, φ, FC, 0.0) ;
    end
  return FC ;
--------------------------------------------------------------------------
procedure TopNFCSearch(Q, C, I, P, FC, min) :
  TopNListUpdate(FC, (ψ(φ(Q)), I), min) ;
  for each d ∈ C such that tail(Q) ≺ d in predefined order do
    begin
      if w_t(I ∩ d) < δ or
         |Q ∪ {d}| + |C ∩ N(d)| < min then
          continue ;
      else
        if ∄ α ∈ C s.t.
                α ≺ d, α ∈ ψ(φ(Q ∪ {d})) and d ∈ ψ(φ(Q ∪ {α})) then
            TopNFCSearch(Q ∪ {d}, C ∩ N(d), I ∩ d, ψ(φ(Q)), FC, min) ;
      endif
    end
--------------------------------------------------------------------------
procedure TopNListUpdate(FC, FC, min) :
  FC ← FC ∪ {FC} ;
  if FC tentatively contains N-th ones then
    min ← N-th extent value ;
    Remove M-th ones from FC such that N < M ;
  else
    min ← 0.0 ;
  endif
```

Fig. 1. Algorithm for Finding Top-N δ-Valid Formal Concepts

4 Experimental Result

In this section, we present our experimental result.

We have conducted an experimentation to observe characteristics of *FC*-clusters.

A set of web pages $\mathcal{P}$ to be clustered has been retrieved by using *Google SOAP Search API* [5] with keywords "Presidential" and "Election". The number of retrieved pages is 968. For each page, its *summary* and *snippet* extracted by Google API are gathered as a document. After the stemming process, we have obtained 3600-terms in the documents and extracted 947 of them as feature terms[1]. Therefore, each page (document) is represented as a 947-dimensional binary vector.

Each web page p is assigned a linear weight, $w(p) = |\mathcal{P}| - rank(p) + 1$, where $rank(p)$ is the rank assigned by Google. Each extent (a set of pages) is evaluated by the sum of the page weights.

The weight of feature term t is given as *inverted document frequency* of t, that is, $w(t) = \log(|\mathcal{P}|/df(t))$, where $df(t)$ is the number of pages in $\mathcal{P}$ containing the term t. Each intent is evaluated by the sum of term weights in the intent.

Under the setting of $\delta = 33.0$, we constructed a weighted undirected graph. Roughly speaking, since the average of term weights is approximately 5.0, a pair of web pages sharing about $6 - 7$ feature terms are connected by an edge.

We tried to extract Top-10 δ-valid *FC*-clusters from the graph. The computation time was just 0.52 second. Some of the obtained clusters are shown in Table 1. In the table, page IDs are equivalent to their ranks.

For comparison, we extracted several clusters based on the standard cosine measure for vector representation of documents. As one of them, we can obtain a cluster D consisting of the pages with the ranks 176, 191, 193, 194, 203, 204, 205, 210 and 465. Note here that FC_1 in Table 1 is a subset of D. Needless to say, it is difficult to understand the meaning of D by means of feature terms. On the other hand, we can clearly understand FC_1 based on the intent. Thus, we can conceptually understand each cluster by referring to the intent. It would be easy to judge whether the cluster is interesting for us or not.

The cluster FC_2 shows that we can obtain 709-th page with (relatively) higher-ranked pages with the ranks 20, 21 and 66. Such a lower-ranked page would not be browsed in many cases. However, our cluster tells us the page might be significant if the concept (intent) would be interesting for us. Thus, our chance to find significant lower-ranked pages can be enhanced.

The cluster FC_3 shows a remarkable characteristic of our method. It consists of web pages concerned with the presidential election by *Tamil people*. The pages would not be so popular and their ranks are really lower. Therefore, almost people will miss them. However, our method can make such clusters visible. More precisely speaking, such an effect can be controlled by our definition of document weights.

[1] All terms with the frequencies above 100 and below 3 have been removed from 3600-terms.

Table 1. *FC*-Based Clusters

Cluster ID.	Extent (Page IDs)	Intent
FC_1	194 203 205 210	Adam, Archive, Back, Carr39, National, Psephos, Top, middot, summary
FC_2	20 21 66 709	Administration, Bush, COVERAGE, Coverage, Electoral, FULL, Full, News, RELATED, Reform, Yahoo, amp
FC_3	246 280 405 600 608	2005, Sri Lanka, Tamil, TamilNet, accurate, concerning, feature, focus, information, issue, news, people, providing, reliable, service
FC_4	176 205 444	2001, Adam, Archive, Carr39, National, Psephos, province, summary
FC_5	70 326 479	Ukrainian, allegation, controversy, example, fraud, include, irregularity, massive

In the experimentation, each document has been assigned a linear weight based on its rank. In other words, such a linearly weighting gives some degree of importance to documents with not higher ranks. Therefore, a cluster consisting of only middle-ranked pages might be extracted as in Top-N, if its size is relatively large. On the other hand, if we assign a weight to each document d such as $w(d) = 1/rank(d)^2$, only clusters with higher-ranked pages will be extracted as in Top-N.

5 Discussion

Formal Concept Analysis is closely related to the problem of *closed itemset mining* (e.g. [6,13]). There is an exact correspondence between a formal concept and a closed itemset. For a context $\mathcal{C} =< \mathcal{O}, \mathcal{F} >$ in FCA, $\mathcal{O}$ and $\mathcal{F}$ can be viewed as a set of transactions and a set of items, respectively, in a problem of itemset mining. In this case, the intent of a formal concept is equivalent to a closed itemset. Thus, finding formal concepts is equivalent to mining closed itemsets.

Uno *et al.* have designed an efficient algorithm, named *LCM*, for *enumerating* all frequent closed itemsets [13]. From the above correspondence, *LCM* can efficiently enumerate all formal concepts. Therefore, one might claim that we would be able to efficiently extract Top-N FC-clusters by simply enumerating all FCs and then sorting δ-valid ones in descending order of extent values. Our preliminary experimentations have shown that the claim is really true under several problem settings. However, the authors emphasize the following remarkable characteristics and advantages of our algorithm:

- *LCM* finds all closed itemsets whose frequencies are greater than or equal to a given threshold. From a theoretical property of itemsets, an itemset with larger size tends to have a lower frequency. Therefore, if we require a higher value of δ to retain a certain quality of formal concepts, extents of preferable formal concepts will become smaller. This means that when we try to extract such formal concepts with the help of *LCM*, we have to provide a lower threshold of frequency. However, as the threshold becomes lower, the number of frequent closed itemsets becomes larger. In such a case, therefore, the computation necessarily takes longer time, because *LCM* is an enumeration algorithm for such formal concepts. As the result, we need much computation time for finding Top-N FC-clusters with *LCM*.

– On the other hand, our algorithm takes both quality and preference of formal concepts into account and can prune many useless formal concepts during search. Especially, in case of higher δ, the computation of Top-N FC-clusters is much faster than one with *LCM*.

The above has been really observed in our preliminary experimentation. In the experimentation, the set of documents to be clustered consists of 5619 articles on politics in newspapers, where the number of feature terms is 2793. The documents and the feature terms are simply assigned uniform weights. That is, intents and extents are evaluated by their sizes. In case of $\delta = 30$, our algorithm takes just 1.21 second for finding Top-5 FC-clusters. On the other hand, in order to obtain the same clusters with *LCM*, we have to give a frequency threshold 5. In that case, *LCM* enumerates about 30 million formal concepts taking 198 second. Furthermore, we have to extract Top-5 30-valid clusters from them. Thus, our algorithm is quite efficient in case of higher δ.

6 Concluding Remarks

In this paper, we discussed a method for pinpoint clustering of documents based on Formal Concept Analysis. Our cluster can consist of similar higher-ranked and lower-ranked pages. Although we are usually careless of pages with lower ranks, they can be *explicitly* extracted together with significant higher-ranked pages. As the result, our clusters can provide new valuable information for users.

Our clusters can be explicitly provided with more convincing meanings, with the help of FCA. By restricting our clusters to extents of formal concepts, we can consider their clear conceptual meanings as their intents (the set of shared feature terms). We designed an algorithm for finding Top-N FC-clusters. It can be viewed as an extended algorithm of our previous one. The extended algorithm utilizes new pruning rules based on a theoretical property of formal concepts. In our experimentation, we confirmed that meaningful clusters can be really extracted according to our method. Furthermore, we verified that our algorithm can efficiently finds Top-N clusters compared with *LCM* in case of higher δ. From the observations, we expect that our method based on Formal Concept Analysis would be a promising approach to finding meaningful clusters of documents.

In our evaluation of intents, each feature term is dealt with *independently*. In general, however, we can observe some correlation among feature terms. The authors expect that the current evaluation mechanism for intents can be improved by taking such a correlation into account. It will provide intents more clear and convincing conceptual interpretations.

The method of *Latent Semantic Indexing* (LSI) [9] is useful for reducing the number of feature terms. Based on the method, we can capture semantically similar terms which can define a compound term corresponding to a new feature term. Improving our current method from this viewpoint would be an interesting future work.

It is known that formal concepts are quite sensitive to *noise* or *exceptions* in a formal context (data set) we are concerned with. In general, existence of such noise or exceptions will increase the number of possible FCs. Especially, we often obtain many FCs which are *slightly different*. Therefore, one might reasonably claim that some *data cleaning* process is indispensable.

As another approach, *approximating* FCs would be also effective. For example, if several FCs are almost the same, an approximate FC can be obtained by grouping them together. The notion of *pseudo-cliques* [18] will become a basis of this kind of approximation. Furthermore, an effective method for approximating closed itemsets has been investigated in the literature [16]. As has been mentioned, since closed itemsets exactly correspond to intents of FCs, an approximation of our valid FCs can be proposed by extending the method in [16].

Needless to say, our method is not only for document clustering. Since it is a general framework, we can easily apply the method for another kind of data in which each object to be clustered can be represented as a set of attributes, like *relational data*. Applying the method to other real data will be also an interesting work.

References

1. M.W. Berry (Ed.), Survey of Text Mining : Clustering, Classification, and Retrieval, Springer, 2004.
2. L. Page, S. Brin, R. Motwani and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", http://dbpubs.stanford.edu/pub/1999-66, 1999.
3. G. Salton and M. J. McGill, Introduction to modern information retrieval, Mcgraw-Hill, 1983.
4. A. Vakali, J. Pokorný and T. Dalamagas, "An Overview of Web Data Clustering Practices", Proceedings of the 9th International Conference on Extending Database Technology - EDBT'04, Springer-LNCS 3268, pp. 597 - 606, 2004.
5. Google SOAP Search API, `http://www.google.com/apis/`.
6. N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal, Efficient Mining of Association Rules Using Closed Itemset Lattices, Information Systems, 24(1), pp. 25 - 46, 1999.
7. E. Tomita and T. Seki, "An Efficient Branch and Bound Algorithm for Finding a Maximum Clique", Proceedings of the 4th International Conference on Discrete Mathematics and Theoretical Computer Science - DMTCS'03, Springer-LNCS 2731, pp. 278 - 289, 2003.
8. T. Fahle, "Simple and Fast: Improving a Branch and Bound Algorithm for Maximum Clique", Proceedings of the 10th European Symposium on Algorithms - ESA'02, Springer-LNCS 2461, pp. 485 - 498, 2002.
9. K. Kita, K. Tsuda and M. Shishibori, "Information Retrieval Algorithms", Kyoritsu Shuppan, 2002 (in Japanese).
10. A.Hotho, G. Stumme, Conceptual Clustering of Text Clusters, In Proc. of the Machine Learning Workshop (FGML'02), 37–45, 2002.
11. A.Hotho, S. Staab, G. Stumme, "Explaining Text Clustering Results Using Semantic Structures", Proc. of the 7th European Conference on Principles of Data Mining and Knowledge Discovery - PKDD'03, pp. 22 - 26, 2003

12. B. Ganter and R. Wille, "Formal Concept Analysis: Mathematical Foundations", Springer, 1999.
13. T. Uno, M. Kiyomi and H. Arimura, "LCM ver. 2: Efficient Mining Algorithm for Frequent/Closed/Maximal Itemsets", IEEE ICDM'04 Workshop FIMI'04, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-126/, 2004.
14. M. Haraguchi, Concept Learning Based on Optimal Clique Searches, JSAI SIG-FAI-A202-11, pp. 63 - 66, 2002 (in Japanese).
15. M. Haraguchi and Y. Okubo, "A Method for Pinpoint Clustering of Web Pages with Pseudo-Clique Search", Federation over the Web, International Workshop, Dagstuhl Castle, Germany, May 1 - 6, 2005, Revised Selected Papers, Springer-LNAI 3847, pp. 59 - 78, 2006.
16. K. Kanda, M. Haraguchi and Y. Okubo: "Constructing Approximate Informative Basis of Asscoiation Rules", Proceedings of the 4th International Conference on Discovery Science - DS'01, Springer-LNAI 2226, pp. 141 - 154, 2001.
17. Y. Okubo and M. Haraguchi, "Creating Abstract Concepts for Classification by Finding Top-N Maximal Weighted Cliques", Proceedings of the 6th International Conference on Discovery Science - DS'03, Springer-LNAI 2843, pp. 418 - 425, 2003.
18. Y. Okubo, M. Haraguchi and B. Shi, "Finding Significant Web Pages with Lower Ranks by Pseudo-Clique Search", Proceedings of the 8th International Conference on Discovery Science - DS'05, Springer-LNAI 3735, pp. 346 - 353, 2005.

N-Gram Analysis Based on Zero-Suppressed BDDs

Ryutaro Kurai, Shin-ichi Minato, and Thomas Zeugmann

Division of Computer Science
Hokkaido University, N-14, W-9, Sapporo 060-0814, Japan
{ryu,minato,thomas}@mx-alg.ist.hokudai.ac.jp

Abstract. In the present paper, we propose a new method of n-gram analysis using ZBDDs (Zero-suppressed BDDs). ZBDDs are known as a compact representation of combinatorial item sets. Here, we newly apply the ZBDD-based techniques for efficiently handling sets of sequences. Using the algebraic operations defined over ZBDDs, such as union, intersection, difference, etc., we can execute various processings and/or analyses for large-scale sequence data. We conducted experiments for generating n-gram statistical data for given real document files. The obtained results show the potentiality of the ZBDD-based method for the sequence database analysis.

1 Introduction

One of the important data model that has been used for text analysis are n-grams (cf., e.g., [2, 7, 8, 9]). Recently, n-grams have been used not only for text analysis but also for text indexing in some search engines [1]. If we can compactly represent n-gram data and efficiently manipulate them, it will greatly facilitate text database analysis and machine learning applications.

In the present paper, we propose a method of n-gram computation with a new sequence data structure based on Zero-suppressed BDDs. BDDs (Binary Decision Diagrams) are graph-based representations of Boolean functions, now widely used in system design and verification. Zero-suppressed BDDs (ZBDDs) are a special type of BDDs that are suitable for handling large-scale sets of combinations. Using ZBDDs, we can implicitly enumerate combinatorial item set data and efficiently compute set operations over the ZBDDs. For n-gram computations, we need to manipulate sets of "sequences," which is a more complicated data model than sets of "combinations." We present a method of manipulating sets of sequences using ZBDDs and generate efficiently n-gram data for given sequence databases. We have implemented a prototype system and show experimental results to evaluate our new n-gram analysis method.

2 ZBDDs and Item Set Manipulation

Within this section, we briefly describe the basic techniques of ZBDDs for representing sets of combinations efficiently.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 289–300, 2007.

2.1 Sets of Combinations and Their Representation

A *set of combinations* consists of the elements each of which is a combination of a number of items. There are 2^n combinations that can be chosen from n items, so we have 2^{2^n} variations of sets of combinations. For example, for a domain of five items a, b, c, d, e, we can express examples of sets of combinations as:

$\{ab, e\}$, $\{abc, cde, bd, acde, e\}$, $\{1, cd\}$, $\emptyset$. Here "1" denotes a combination of null items, and "$\emptyset$" means the empty set. Sets of combinations are one of the basic data structures for handling combinatorial problems. They often appear in real-life problems, such as combinations of switching devices, sets of faults, paths in the networks, etc., and of course, they can be used for representing frequent item set data.

a	b	c	F	
0	0	0	0	
0	0	1	1	$\rightarrow c$
0	1	0	0	
0	1	1	1	$\rightarrow bc$
1	0	0	0	
1	0	1	0	
1	1	0	1	$\rightarrow ab$
1	1	1	1	$\rightarrow abc$

As a Boolean function:
$F = ab + \overline{a}c$

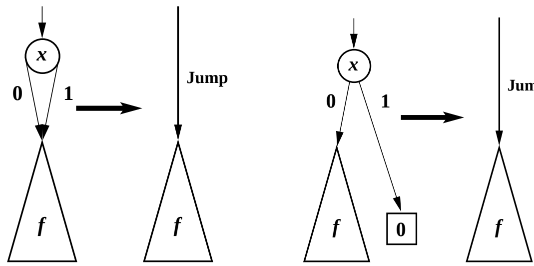
As a set of combinations:
$F = \{ab,\ abc,\ bc,\ c\}$

Fig. 1. Correspondence of Boolean functions and sets of combinations

A set of combinations can be mapped into Boolean space of n input variables. For example, Fig. 1 shows the truth table of the Boolean function $(ab + \overline{a}c)$, but it also represents the set of combinations $\{ab, abc, bc, c\}$. Such Boolean functions are called *characteristic functions* for the sets of combinations. Using BDD manipulation for characteristic functions, we can implicitly represent and manipulate large-scale sets of combinations. In addition, we can enjoy more efficient manipulation using "Zero-suppressed BDDs" (ZBDD) (cf. [4]), which are a special type of BDDs optimized for handling sets of combinations.

ZBDDs are based on the reduction rule different from the one used in ordinary BDDs. As illustrated in Fig. 2(a), the ordinary reduction rule deletes the nodes whose two edges point to the same node. However, in ZBDDs, we do not delete such nodes but delete another type of nodes whose 1-edge directly points to the 0-terminal node, as shown in Fig. 2(b).

In ZBDDs, a 0-edge points to the subset (cofactor) of combinations not including the decision variable x, and a 1-edge points to the subset (cofactor) of combinations including x. If the 1-edge directly points to the 0-terminal node, it means that the item x never appears in the set of combinations. The Zero-suppressed reduction rule automatically deletes such a node with respect to the



(a) Ordinary BDD. (b) ZBDD.

Fig. 2. Different reduction rules for BDD and ZBDDs

irrelevant item x, and thus ZBDDs more compactly represent sets of combinations than ordinary BDDs do.

The detailed techniques of ZBDD manipulation are described in the articles [4, 5]. A typical ZBDD package supports cofactoring operations to traverse 0-edge or 1-edge, and binary operations between two sets of combinations, such as union, intersection, and difference. The computation time for each operation is almost linear to the number of ZBDD nodes related to the operation.

2.2 Item Set Histograms and ZBDD Vectors

An *item set histogram* is the table for counting the number of appearances of each item combination in the given database. An example of an item set histogram is shown in Fig. 3. This is just a compressed table of the database to combine the same tuples appearing more than once into one line with the frequency.

Record ID	Comb.
1	*abc*
2	*ab*
3	*abc*
4	*bc*
5	*ab*
6	*abc*
7	*c*
8	*abc*
9	*abc*
10	*ab*
11	*bc*

Original Database

Comb.	Freq.
abc	5
ab	3
bc	2
c	1

Item set histogram

Fig. 3. Database example and item set histogram

Comb.	frequency	F_2	F_1	F_0
abc	5 (101)	1	0	1
ab	3 (011)	0	1	1
bc	2 (010)	0	1	0
c	1 (001)	0	0	1

$F_0 = \{abc,\ ab,\ c\}$
$F_1 = \{ab,\ bc\}$, $F_2 = \{abc\}$

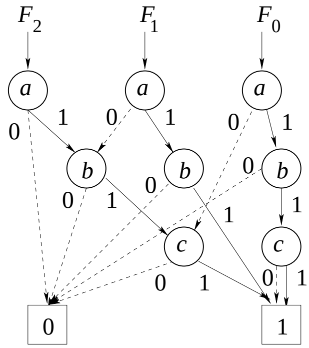


Fig. 4. ZBDD vector for item set histogram

Our n-gram data structure is based on the item set histogram representation [6] using ZBDDs. Since ZBDDs are representation of sets of combinations, a simple ZBDD distinguishes only the existence of each combination in the database. In order to represent the numbers of combination's appearances, we decompose the number into m-digits of ZBDD vector $\{F_0, F_1, \ldots, F_{m-1}\}$ to represent integers up to $(2^m - 1)$, as shown in Fig. 4. Namely, we encode the appearance numbers into binary digital code, where F_0 represents a set of tuples appearing odd times (LSB = 1), F_1 represents a set of combinations whose appearance number has a 1 in the second lowest bit, and similarly we define the set of each digit up to F_{m-1}.

In the example of Fig. 4, the item set frequencies are decomposed as: $F_0 = \{abc, ab, c\}$, $F_1 = \{ab, bc\}$, $F_2 = \{abc\}$, and then each digit can be represented by a simple ZBDD. The three ZBDDs share their sub-graphs to one another.

When we construct a ZBDD vector of an item set histogram, the number of ZBDD nodes in each digit is bounded by the total appearance of items in all combinations. If there are many partially similar combinations in the database, the sub-graphs of ZBDDs are shared very well, and a compact representation is obtained. The bit-width of ZBDD vector is bounded by $\log S_{max}$, where S_{max} is the appearance of most frequent items.

Once we have generated a ZBDD vector for the item set histogram, various operations can be executed efficiently. Here are the instances of operations used in our pattern mining algorithm.

- H.factor0(v): Extracts sub-histogram of combinations not including item v.
- H.factor1(v): Extracts sub-histogram of combinations including item v and then delete v from the combinations. (also considered as the quotient of H/v)
- $v \cdot H$: Attaches an item v on each combination in the histogram F.
- $H_1 + H_2$: Generates a new item set histogram with sum of the frequencies of corresponding tuples.
- H.tuplecount: The number of tuples appearing at least once.

These operations can be composed as a sequence of ZBDD operations and the result is also compactly represented by a ZBDD vector. The computation time is roughly linear in the total ZBDD sizes. For a more detailed description of the techniques of ZBDD vector manipulation, we refer the reader to [6].

3 ZBDD-Based Representation for Sets of Sequences

Next we explain, in some more detail, our method of generating n-gram data representations using ZBDDs.

3.1 Sets of Sequences and Sets of Combinations

First, we describe the way to extend the representation of sets of combinations to sets of sequences. In our method, we define an item for each symbol (or character) used in each position of sequences. Namely, the item represents not only a symbol but also its position in the sequence. An example is given below.

Sequences	Combinations
AABC	$a_1a_2b_3c_4$
CABC	$c_1a_2b_3c_4$

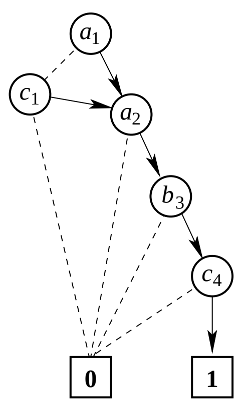


Fig. 5. ZBDDs example which expresses sets of sequences

In the combination for the sequence "AABC," the item a_1 represents the symbol "A" at the first (leftmost) position. Similarly, c_4 express the symbol "C" at the 4th position. In our method, if the same symbol appears at different positions, we treat them as different items. For example, the sequence "AABC" includes the same symbol "A" at the first and the second position, but we treat them as a_1 and a_2, i.e., as different items. After such an assignment of items, a set of sequences can be represented by a set of combinations. Then, the set of combinations can be represented by a ZBDD. An example is shown in Fig. 5. Here, the ZBDD represents the set of sequences {AABC, CABC}. In this way, the sets of sequences are decomposed to sets of combinations, and they can be manipulated by ZBDD-techniques.

3.2 *n*-Gram Representation

An n-gram data is a histogram of all possible subsequences of length n included in a given sequence. As described in Subsection 2.2, we can compute item set histograms by using ZBDD vectors. So, the ZBDD-based representation for sets of sequences can be extended to sequence histograms by using ZBDD vectors.

3.3 Binary Coding of Symbols

In our method, we use a number of items for the respective positions in the sequences. The number is the product of the number of symbols $|\Sigma|$ and the sequence length m. If we apply this method for a language using not so many symbols such as gene sequence data, the number of items would be feasible. However, for a language that has many symbols such as Japanese Kanji, too many items would be required. To address this problem, we propose a method of using binary coding to save the number of items.

When we assume the three symbols A, B and C for each position, we may use the 2-bit binary coding shown in the following table.

Symbols	Items	Binary code ($x_k^1 x_k^0$)	Encoded combinations
A	a_k	01	x_k^0
B	b_k	10	x_k^1
C	c_k	11	$x_k^1 x_k^0$

Using this encoding, we can express the sequences "AABC" and "CABC" by the following table, and the resulting ZBDD is shown in Fig. 6.

Sequences	Encoded combinations
AABC	$x_1^0 x_2^0 x_3^1 x_4^1 x_4^0$
CABC	$x_1^1 x_1^0 x_2^0 x_3^1 x_4^1 x_4^0$

We may use ASCII code for this purpose. ASCII code expresses conventional western characters with 8bit vectors. We treat one bit as one item. One can distinguish at most 255 symbols with a combination of eight items for each position. Examples are shown below.

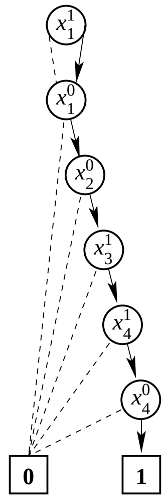


Fig. 6. ZBDDs example which expresses sets of sequences

Symbols	ASCII (decimal)	Binary code($x^7x^6x^5x^4x^3x^2x^1x^0$)	Encoded combinations
a	97	01100001	$x^6x^5x^0$
b	98	01100010	$x^6x^5x^1$
c	99	01100011	$x^6x^5x^1x^0$
z	122	01111010	$x^6x^5x^4x^3x^1$
A	65	01000001	x^6x^0

If the probability distribution of the symbols is given, we shall be able to design a more efficient binary coding assignment.

4 Experimental Results

4.1 Generating n-Gram Data Based on ZBDDs

For evaluating the feasibility of our method, we performed experiments for generating ZBDDs of the n-grams for a given text data. We used the English plain text data "Alice in Wonderland." This text consists of over 3,000 lines and 26,000 words, having a total of 138KB of plane text. In our experiments, we used a 3.00GHz Pentium 4 Linux PC with 512MB main memory.

The results of generating n-grams for $n = 1$ through 10 are shown in Table 1. In this table, the column "ZBDD nodes" shows the total number of nodes in the ZBDD vector representing the n-gram. "#Sequences" stands for the number of

Table 1. Experimental results

	ZBDD				ruby-Hash
n	ZBDD nodes	#Sequences	Max frequency	Time(sec)	Time(sec)
1	121	26	13,681	3.17	0.28
2	1,428	518	3,813	6.18	0.29
3	9,487	4,750	2,402	9.45	0.41
4	33,199	19,004	500	12.96	1.01
5	73,940	38,913	417	16.73	2.12
6	129,738	57,510	221	22.4	2.83
7	198,311	72,285	211	30.33	3.90
8	283,544	82,891	116	44.02	5.01
9	388,863	90,247	116	63.49	6.05
10	522,175	95,366	58	89.51	7.45

different n-grams actually found in the text. "Max frequency" shows the number of appearances of the most popular subsequences of length n in the text. Finally, "Time" displays the time needed to generate the data structure ZBDD. The ZBDD structure is generated by VSOP (see [6]), the script language which can manipulate large scale ZBDDs. In this experiment, we used the symbol encoding method with ASCII code, as described in the previous section.

From the result, we can observe that the size of ZBDDs is almost proportional to the product of the length n and the number of sequences. The CPU time is also almost linear in size of the ZBDD.

Then we compare the performance of generating speed between ZBDD and other data structure. As an existing method, we use the Ruby [10] "Hash" class implementation. Ruby is the object oriented script language which is popularly used in web applications. The "Hash" class provides the function of indexing tables, and is included in the ruby standard library. The result of this experiment is the rightmost column "Time" in the Table 1. The result shows that our method is almost ten times slower than the ruby-Hash implementation, so our method is not so effective if we simply generate n-grams. However, ZBDD-based method has an advantage of flexible post processing after generating n-gram data. If we use ruby-Hash data structure we can only operate simple matching of sequences. So we would have to write several ruby codes to calculate complicated operations.

After generating the ZBDD vector of the n-gram, we can easily retrieve the frequent sequences by using algebraic operations over ZBDDs. The additional computation cost is relatively smaller than the cost for generating n-grams. Table 2 shows the top 10 frequent sequences in the 5-gram of "Alice in Wonderland." The most frequent sequence is "alice," and it appears 417 times. The 2nd, 3rd, 4th, and 10th sequences seem a part of same sequence "he/she said the." In this experiment, we removed spaces, special symbols such as question mark, and periods from the text. We only used case-independent alphabets and numbers to create n-grams.

Table 2. Top 10 subsequences in the 5-gram of "Alice in Wonderland"

Rank	Sequence	Appearance
1	"alice"	417
2	"saidt"	265
3	"aidth"	223
4	"idthe"	220
5	"thing"	170
6	"andth"	169
7	"dalic"	163
8	"ofthe"	156
8	"ndthe"	156
10	"esaid"	133

4.2 Post Processing for *n*-Gram Data

An advantageous feature of our ZBDD-based method is that it allows flexible post processing after the n-gram data have been generated. The post processing is done by using algebraic operations such as intersection, union, difference, and, additionally, some numeric operations.

For example, the operation that calculates the rank of frequency of the sequences is represented by the following short command scripts. These command scripts are written in VSOP. The uppercase letters are register variables which show sequences of n-grams, while the lowercase letters are symbols which represent each character in n-grams.

```
T1 = {a1 a2 b3 + a1 b2 c3 + ...}
M = T1.MaxVal
print M
print T1 / M
```

Using only four commands we can get the ranking and the frequency of appearance. The first command generates the ZBDD of item sets of n-gram. The three symbols "a1 a2 b3" represent the 3-gram "aab". The meaning of symbol "a1" is as follows: "a" means that symbol "a1" expresses the character "a", and "1" specifies that character "a" exists as 1st character in the 3-gram. Similarly, the three symbols "a1 b2 c3" express another 3-gram, i.e., "abc". Therefore, the variable "T1" contains some 3-grams including "aab", "abc" as item sets. `MaxVal` is the calculation to get the number of the largest coefficient. So, the 2nd command returns the frequency of the most popular item sets as M. And the last line returns the most popular item sets.

For having another example, suppose that we want to obtain the subset of the n-gram data such that the last letter is either "a," "e," "i," "o," or "u." In this case, we first generate the five subsets of the n-gram data such that the first subset includes all n-grams having the last letter "a," the second subset includes all n-grams having the last letter "e," . . . , and the fifth subset includes all

n-grams having the last letter "u." Then we apply union operations to combine all the subsets.

These instructions can easily be described in a command script of ZBDD operations as shown below.

```
T2 = {a1 a2 b3 + a1 b2 c3 + ...}
T3 = (T2 / a3) * a3
T4 = (T2 / e3) * e3
T5 = (T2 / i3) * i3
T6 = (T2 / o3) * o3
T7 = (T2 / u3) * u3
T8 = T3 + T4 + T5 + T6 + T7
```

In the first line command, we store all n-grams as T2. T3 to T7 are the subsets including the corresponding letter at the last position. Finally, we get the combination of all subsets T3 to T7 as T8.

Next, we show a good example of applying ZBDD operations after generating n-gram data. We prepared a command script to compare two sets of n-grams. That is, we compare the first half of the text in "Alice in Wonderland" to the second half of it. The corresponding VSOP commands to compare these two sets are displayed below.

```
S1 = {a1 b2 c3 + b2 c3 d4 + ...}
S2 = {c1 d2 e3 + d2 e3 f4 + ...}
C = S1 - S2
M = C.MaxVal
print M
print C / M
```

In the first and the second command we store the first and second half of n-grams as S1 and S2, respectively. Then we store the difference of S1 and S2 as C. Finally, `C.MaxVal` and `C / M` return the frequency of the most popular sequence and the sequence itself, respectively.

The obtained results are displayed in Table 3, 4, and 5 below.

Table 3 and 4 do not differ much, since the same sequences appear many times in the first and the second half of the text. Therefore, we also check the histogram of differences of the histogram. As shown in Table 5 below, the sequence "the

Table 3. Top 5 subsequences in the 5-gram of "The first half of Alice in Wonderland"

Rank	Sequence	Appearance
1	"alice"	201
2	"thing"	93
2	"saidt"	93
4	"littl"	85
4	"ittle"	85

Table 4. Top 5 subsequences in the 5-gram of "The Second half of Alice in Wonderland"

Rank	Sequence	Appearance
1	"alice"	216
2	"saidt"	172
3	"idthe"	145
4	"aidth"	143
5	"andth"	103

Table 5. Top 5 subsequences in the 5-gram of "Difference between the First half and the Second half of Alice in Wonderland"

Rank	Sequence	Appearance
1	"saidt"	79
2	"idthe"	70
3	"queen"	68
4	"thequ"	65
5	"heque"	65
5	"equee"	65
5	"urtle"	65

queen" appears only in this table. This is caused by the fact that "queen" appears in the story only in the second half of the text (frequently).

5 Conclusions

In the present paper, we have proposed a new method to represent sequences in sets of item sets, i.e., by using ZBDDs. In order to demonstrate the efficiency of this method, we have performed a preliminary experiment of generating an n-gram set for a given text file. The experimental result shows that we can construct the ZBDD-based n-gram data structure in a feasible computation time and memory space. Regrettably, our data structure does not perform faster than previous data structure implemented by Ruby. So we improve our method to utilize ZBDD's effect that share the ZBDD nodes and compress size of data.

Though we restricted ourselves within this paper to generate n-gram statistical data for a given real document file, i.e., "Alice in Wonderland," ZBDDs can also be used to perform more advanced calculations. For example, we may also apply our method to represent a text dictionary and to perform quickly dictionary search operations.

It should be noted that our new method can also be applied to analyze and to manipulate DNA data or some semi-structured data such as linguistic data annotated by NLP tools (e.g., [3]). This will be done in the future.

References

[1] B. Hoffmeister and T. Zeugmann: Text Mining Using Markov Chains of Variable Length, In (K.P. Jantke, A. Lunzer, N. Spyratos, Y. Tanaka, eds), *Federation over the Web, International Workshop, Dagstuhl Castle, Germany, May 2005, Revised Selected Papers*, Lecture Notes in Artificial Intelligence Vol. 3847, pp. 1–24, Springer, Feb. 2006.

[2] P. Jokinen, E. Ukkonen: Two algorithms for approximate string matching in static texts, In *Mathematical Foundations of Computer Science 1991, 16th International Symposium, MFCS'91, Kazimierz Dolny, Poland, September 9-13, 1991, Proceedings*, Lecture Notes in Computer Science Vol. 520, 240-248, Springer-Verlag 1991.

[3] T. Kudo, K. Yamamoto, Y. Tsuboi, Y. Matsumoto: Text mining using linguistic information, IPSJ SIG-NLP NL-148, (in Japanese), pp. 65–72, 2002.

[4] S. Minato: Zero-suppressed BDDs for set manipulation in combinatorial problems, In Proc. 30th Design Automation Conference (DAC-93), pp. 272–277, ACM Press, June 1993.

[5] S. Minato: Zero-suppressed BDDs and their applications, International Journal on Software Tools for Technology Transfer (STTT), 3(2):156–170, Springer, May 2001.

[6] S. Minato: VSOP (Valued-Sum-of-Products) Calculator for Knowledge Processing Based on Zero-Suppressed BDDs, In (K.P. Jantke, A. Lunzer, N. Spyratos, Y. Tanaka, eds), *Federation over the Web, International Workshop, Dagstuhl Castle, Germany, May 2005, Revised Selected Papers*, Lecture Notes in Artificial Intelligence Vol. 3847, pp. 40-58, Springer, Feb. 2006.

[7] M. Nagano, and S. Mori: A new method of N-gram statistics for large number of n and automatic extraction of words and phrases from large text data of Japanese, In *Proc. 15th Conference on Computational Linguistics*, Vol. 1, pp. 611–615, Association for Computational Linguistics, Morristown, NJ, USA, 1994.

[8] Y. Tsuboi: Mining frequent substrings, Technical Report of IEICE, NLC2003-47, 2003 (in Japanese).

[9] E. Ukkonen: Approximate string-matching with q-grams and maximal matches, *Theoretical Computer Science*, 92(1):191–211, 1992.

[10] Ruby Home Page http://www.ruby-lang.org/en/

Risk Mining - Overview

Shusaku Tsumoto[1] and Takashi Washio[2]

[1] Department of Medical Informatics, Shimane University, School of Medicine
89-1 Enya-Cho, Izumo 693-8501 Japan
tsumoto@computer.org
http://www.med.shimane-u.ac.jp/med_info/tsumoto/index.htm
[2] The Institute of Scientific and Industrial Research (I.S.I.R), Osaka University
8-1, Mihogaoka, Ibarakishi, Osaka 567-0047, Japan
washio@ar.sanken.osaka-u.ac.jp
http://www.ar.sanken.osaka-u.ac.jp/washio/washpreg.html

Abstract. International workshop on Risk Mining (RM2006) was held in conjunction with the 20th Annual Conference of the Japanese Society for Artificial Intelligence(JSAI2005), Tokyo Japan, June 2005. The workshop aimed at sharing and comparing experiences on risk mining techniques applied to risk detection, risk clarification and risk utilization. In summary, the workshop gave a discussion forum for researchers working on both data mining and risk management where the attendees discussed various aspects on data mining based risk management.

1 Introduction

Organizations in our modern society grow larger and more complex to provide advanced services due to the varieties of social demands. Such organizations are highly efficient for routine work processes but known to be not robust to unexpected situations. According to this observation, the importance of the organizational risk management has been noticed in recent years. On the other hand, a large amount of data on the work processes has been automatically stored since information technology was introduced to the organizations. Thus, it has been expected that reuse of collected data should contribute to risk management for large-scale organizations.

This workshop focused on risk mining, where data mining techniques were applied to detection and analysis of risks potentially existing in the organizations and to usage of risk information for better organizational management. The workshop brought together researchers who are interested in both of the areas of data mining and risk management, where the attendees discussed intensively various aspects on the data mining based risk management.

The topics of the workshop in call for papers included:

- Data mining for risk management
- Chance discovery for risk management
- Active Mining for risk management

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 303–304, 2007.

- Machine learning for risk management
- Other techniques for risk detection, analysis and utilization
- Applications in the fields of medicine, marketing, security, decision support in business, social activities, human relationships, chemistry and sensor data

2 Workshop Overview

All submitted papers were carefully peer reviewed by program committee members. In the workshop, we had total 13 excellent presentations in the following four sessions: “Social Risk Mining (4 papers)”, “Power Plant Risk Mining (2 papers)”, “Network Risk Mining (4 papers) ” and “Medical Risk Mining (3 papers).”

From those presentations, PC members selected the following 7 excellent papers for this proceedings:

(Social risk mining) We selected Washio’s paper, where the authors present risk clarification on relations between enterprise profit and financial Sate by using data mining techniques.
(Power plant risk mining) The following two papers are selected: Onoda et al. show unusual condition detection of bearing vibration in hydroelectric power plants. Shu presents an application of data mining to a nonlinear structural health assessment in nuclear power plants.
(Network risk mining) The following three papers are selected: first, Hu et al. report grid-based e-finance portals for intelligent risk management. Secondly, Yada and Ohno show knowledge discovery from click stream data and effective site management Finally, Yoshida discussed sampling-based stream mining for network risk management.
(Medical risk mining) We selected Abe’s paper, where the authors presents relations between abductive and inductive types of nursing risk management.

We would like to thank all the authors who submitted papers to the workshop and participated in the interesting discussions at the workshop. We would also like to thank the all active program committee members for their efforts in careful reviewing of papers and supporting the success of the workshop.

3 Program Committee

Workshop Chair
Shusaku Tsumoto (Shimane University)
Takashi Washio (Osaka University)

Program Committee (Alphabetical)
Yukio Ohsawa (The University of Tokyo)
Einoshin Suzuki (Kyushu University)
Shusaku Tsumoto (Shimane University)
Takashi Washio (Osaka University)
Ning Zhong (Maebashi Institute of Technology)

Analysis on a Relation Between Enterprise Profit and Financial State by Using Data Mining Techniques

Takashi Washio[1], Yasuo Shinnou[2], Katsutoshi Yada[3], Hiroshi Motoda[1], and Takashi Okada[2]

[1] I.S.I.R., Osaka University
washio@sanken.osaka-u.ac.jp
[2] Kwansei Gakuin University
okada-office@ksc.kwansei.ac.jp
[3] Kansai University
yada@ipcku.kansai-u.ac.jp

Abstract. The knowledge on the relation between a financial state of an enterprise and its future profit will efficiently and securely reduce the negative risk and increase the positive risk on the decision making needed in the management of the enterprise and the investment in stock markets. Generally speaking, the relation is considered to have a highly complicated structure containing the influences from various financial factors characterizing the enterprise. Recent development of data mining techniques has significantly extended the power to model such a complicated relation in accurate and tractable manners. In this study, we assessed the feasibility to model the relation in the framework of data mining, and analyzed the characteristics of the model.

1 Introduction

The relation between a financial state of an enterprise and its future profit is very crucial for the management of the enterprise. Moreover the model of the relation, if it is available, is expected to provide an important measure for the investment in stock markets. The knowledge on the relation will efficiently and securely reduce the negative risk and increase the positive risk on the decision making needed in the management and the investment. However, a financial settlement of an enterprise represents the state of the enterprise management during the financial term immediately before, but does not reflect the enterprise management and its surrounding economic environment in the next financial term. In addition, the rules of enterprise accounting is highly complicated, and contains various ambiguities due to the variety of measures to tune important financial indices for the optimality of the management and tax saving. Accordingly, the relation between a financial state of an enterprise and its future profit has a highly ambiguous and complicated structure containing the influences from various financial factors characterizing the enterprise. The theoretical development of a precise model of the relation is known to be highly difficult [1].

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 305–316, 2007.

On the other hand, along the development of communication network and the trend of information disclosure, the detailed data on the financial states of listed enterprises are now available. Moreover, data mining technology including machine learning and statistical mathematics is rapidly growing [2], and its efficient tools easy to use are now widely available [3]. The significance of the data mining is not limited to the modeling based on massive data but the ability to capture complicated and ambiguous characteristics of the data. Under these circumstances, the data mining technology has high possibility to develop an appropriate model of the relation between a financial state of an enterprise and its future profit. In this study, we attempt to develop some accurate models to predict the profit of an enterprise in a next financial term from the financial data in the past terms by using the data mining techniques. Moreover, the relation between a financial state of an enterprise and its future profit is analyzed through the assessment on the characteristics of the model.

More concretely speaking, we searched the best accurate modeling approach among the learning techniques provided in a data mining tool Weka [4] for the data consisting of the profit of enterprises in a financial term and the financial data in their past terms. The data contains the records of enterprises listed on the first section of Tokyo stock exchange market. Some financial indices have been preprocessed by an accounting expert to give the indices independent and clear meaning in terms of accounting. In the result, a modeling approach named Random Tree based Committee Learning provided the best accurate prediction model. Furthermore, through the analysis of the model and the data, the final result indicates that every financial index in the data does not have any strong correlation with the future profit, but most of the indices synthetically determine the future profit.

2 Objective Data

Financial settlement data of the enterprises listed on the first section of Tokyo stock exchange market from 2000 year to 2004 is recorded in the CD-ROM provided by Nikkei NEEDS service [5]. The original data in the CD-ROM contains missing values of some financial indices, and these missing values are complemented by an accounting expert based on the other disclosed information of the financial settlement of the enterprises. Some financial indices are converted from the originals to the other forms having independent and clear meaning in terms of accounting, and finally the data of 60 financial indices are defined and prepared for 1326 enterprises among 1480 listed in the market [1]. A binary class $Pr_{i,t+1}$ representing the profit of an enterprise i in a financial term $t+1$ is defined as follows.

$$Pr_{i,t+1} = \begin{cases} 0, \, for \;\; e.s.p._{i,t+1} - e.s.p._{i,t} - drift_{i,t+1} < 0, \\ 1, \, for \;\; e.s.p._{i,t+1} - e.s.p._{i,t} - drift_{i,t+1} \geq 0, \end{cases} \tag{1}$$

where $e.s.p._{i,t}$ is profit per share of an enterprise i at a financial term t, $drift_{i,t+1}$ average annual change of $e.s.p._{i,t}$ over the terms $[t-3, t]$. Hence, the class value is 1 if the increase of the profit in the term $t+1$ from that in t is greater than or equal to the average annual change of the profit over the past four years. Otherwise, the class value is 0.

We define a financial term as an annual period and $t = 2003$. Accordingly, the class value is defined for the profit in the term 2004, and thus the models are developed to predict the class value of each enterprise in 2004 from its financial settlement data in the past four terms from 2000 to 2003. Among 1326 enterprises in the data, 62.92% of them have the class value 1. Given a blind model always predicting the class of an enterprise as 1, this model achieves the accuracy more than 60%. To reduce this biased class distribution, the instances having the class value 0 have been randomly chosen and duplicated to equalize the class distribution. Through this preprocessing, total number of the instances in the data has been increased to 1670.

3 Model Selection for Profit Prediction

We used Weka [4] to search the best classifier for the aforementioned data. Weka is a generic and open source data mining tool supporting many machine learning approaches. We adopted all classification methods under their default parameter settings in Weka except ones not applicable to the data containing numeric attributes, because all attributes of the objective data are numeric. In total, 49 classification methods were applied to the data, and their results sorted by the classification accuracies of 10 hold cross validations are listed in Table 1. The type “meta” indicates that the type of a classifier is the meta learning to combine multiple classifiers and learning strategies. The type “tree” is a classifier to use tree structured classification schemes and the type “rules” to use rule-based classification. The top 7 classifiers except NNge showing high accuracies are known to use learning strategies of tree based ensemble learning where the predictions of multiple decision trees are combined to derive a classification result. On the other hand, the standard approaches to use a single classifier such as C4.5 do not provide very good results. “Random Tree based Committee Learning”, “Random Committee” in short, provides the best accuracy to predict the future profit of enterprises from the financial data in their past terms [6,7].

“Committee Learning” is an approach to use multiple base classifiers and iteratively train the base classifiers by using critical instances which resulted large discrepancy among the class predictions of the base classifiers. “Random Tree based Committee Learning” is a committee learning to use “Random Trees” as the base classifiers. The details of these approaches will be explained later. Besides the Random Trees, we assessed to apply the other 13 classifiers as the base classifiers of the Committee Learning. Table 2 shows their results for the objective data. In total, the Random Tree based Committee Learning we have already seen is known to give the best result.

We further assessed the optimum number of base classifiers, *i.e.*, Random Trees, used in the Random Tree based Committee Learning. Table 3 shows the relation between the number of Random Trees and the classification accuracy

Table 1. Accuracies of 49 classification methods under default settings

Type	Method	Acc.[%]	Type	Method	Acc.[%]
meta	Random Committee	81.3	rules	PART	59.9
meta	Decorate	76.8	rules	Decision Table	59.3
tree	Random Forest	76.7	rules	Ridor	59.0
rules	NNge	75.1	functions	SMO	57.9
tree	Random Tree	74.1	rules	OneR	56.7
tree	LMT	72.6	meta	Raced Incremental	55.8
meta	Bagging	71.3		Logit Boost	
meta	Classification	69.9	tree	Decision Stamp	55.4
	Via Regression		lazy	LWL	55.2
lazy	Kstar	68.3	rules	Conjunctive Rule	54.5
tree	NB Tree	67.6	bayes	Naive Bayes Simple	54.3
meta	Ordinal Class	64.3	bayes	Naive Bayes	54.1
	Classifier		bayes	Naive Bayes	54.1
tree	C4.5(J48)	64.3		Update Table	
meta	Ada Boost M1	62.9	functions	RBF Network	53.3
meta	Multi Boost AB	62.9	meta	Threshold Selector	50.8
tree	REP Tree	62.8	functions	Voted Perceptron	49.7
tree	Ad Tree	62.5	meta	CV Parameter	49.7
meta	Attribute Selected	62.2		Selection	
	Classifier		meta	Grading	49.7
meta	Logit Boost	61.9	meta	Multi Scheme	49.7
rules	Jrip	61.5	meta	Stacking	49.7
functions	Simple Logistic	61.3	meta	Stacking C	49.7
bayes	Bayes Net	60.5	meta	Vote	49.7
meta	Filtered Classifier	60.5	rules	Zero R	49.7
functions	Logistic	60.0	lazy	IB1	49.6
functions	Multilayer Perceptron	60.0	lazy	IBk	49.6
meta	Multi Class Classifier	60.0			

Table 2. Accuracies of 14 classification methods with committee learning

Type	Method	Acc.[%]	Type	Method	Acc.[%]
meta	Random Trees	81.3	meta	Raced Incremental	57.0
tree	Random Forest	81.0		Logit Boost	
meta	Decorate	78.4	meta	Threshold Selector	50.5
meta	Bagging	75.4	meta	Grading	49.7
meta	Ada Boost M1	62.9	meta	Multi Scheme	49.7
meta	Multi Boost AB	62.9	meta	Stacking	49.7
meta	Logit Boost	61.9	meta	Stacking C	49.0
meta	Multi Class Classifier	60.0			

Table 3. Accuracies for the number of Random Trees

Number	10	20	25	30	50	100	150	200
Acc.[%]	81.3	82.5	82.6	83.8	84.5	84.3	84.3	84.2

obtained through 10 fold cross validations. The optimum number is shown to be 50 for the objective data. To increase the classification accuracy by the ensemble learning, the set of instances miss-classified by each base classifier must be statistically independent from the sets miss-classified by the other base classifiers. Because the statistically independent miss-classified sets are finite in many cases, the increase of the classification accuracy stops at a certain number of base classifiers. The optimum number of Random Trees is only for the current objective data, and thus more precise optimum number may be slightly different from 50.

According to an accounting specialist who is an author of this paper, the prediction accuracy 84.5% provided by the Random Tree based Committee Learning is sufficient for the management of enterprise profit and the decision making needed in the stock investment. The practical and commercial use of the profit prediction results is out of scope of this paper.

4 Outline of Random Tree Based Committee Learning

In the latter section, we are going to analyze the mechanism determining the future profit of an enterprise from its past financial conditions through the characteristics of the Random Tree based Committee Learning which provided the best prediction accuracy. Here, the Random Tree based Committee Learning is outlined, and its characteristics are reviewed. First, the Random Tree, the base classifier, of this learning approach is briefly reviewed. Second, the framework of Committee Learning to combine the base classifiers is described.

4.1 Random Tree

A decision tree is generated by recursively adding decision nodes and its successive braches, while the attribute at each node to distribute instances to the branches is randomly chosen from the objective data for classification in a Random Tree [6]. The threshold value to distribute instances to the braches is also randomly determined in case of a numeric attribute. This recursive extension of the tree is repeated until the most of the instances distributed to each branch become to have a unique class label by chance. Because the tree is randomly generated, the computational amount needed to construct a tree is so small that the required time for the generation is almost 1/100 of that of C4.5 in the average. On the other hand, the size of the generated tree is usually 4 to 5 times larger than the ordinary decision trees such as C4.5, because the tree extension continues until the class label of the instances in each branch becomes almost unique.

In the framework of the standard Random Tree classification, multiple Random Trees are generated, and a Bagging approach to take their ensemble is applied to derive an entire classification result for each instance [8]. In the Bagging, the class label of an instance is predicted by each Random Tree, and the total prediction of the label is determined by the voting of all Random Trees.

The prediction by a Random Tree is highly independent from the others due to the randomness of the tree generation process. In addition, the prediction accuracy of a Random tree for unseen data is more than 50% in many cases, since the tree has been generated to over fit to the training data which are likely to be similar to the unseen data. Accordingly, the majority decision based on the independent predictions having more than 50% accuracy produces highly accurate and robust classification.

4.2 Committee Learning

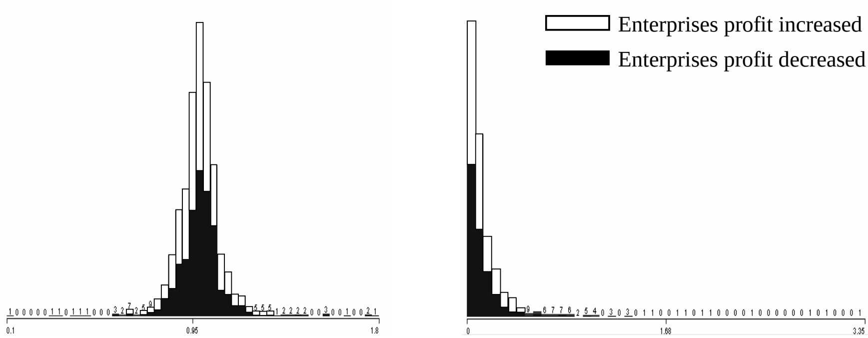
Committee Learning is a framework of iterative ensemble learning to choose instances and base classifiers for learning based on the classification result in the previous learning step. The Committee Learning of Weka applies a learning principle named Query by Committee [7]. We focus its explanation to a binary classification problem such as our profit prediction without loss of generality. Given an objective data set, we split it to a small training data set and the rest evaluation data set. Then a set of base classifiers is produced by using the small training data set. In case of the Random Tree based Committee Learning, each base classifier is a Random Tree generated from the training data set by the aforementioned procedure. Subsequently, all base classifiers are applied to an instance removed from the evaluation data set. If the instance is misclassified by almost a half of the base classifiers, the instance is added to the training data set. Then, the wrong base classifiers are retrained by the revised training data set. Otherwise, the instance is discarded and not used for learning. This learning procedure is repeated until the evaluation data set exhausts. The class prediction of the committee for an unseen instance is given by the Bagging of the base classifiers as described in the previous section.

If the discrepancy among predictions on an instance by the base classifiers is the maximum, *i.e.*, a half of the base classifiers misclassifies the instance, the instance provides the maximum information amount to the committee of the base classifiers through its learning. Accordingly, the above procedure to select instances and base classifiers for learning is considered to be the best in terms of learning efficiency while reducing the over learning possibility.

5 Characteristics of Financial Settlement Data

The prediction accuracy of the enterprise profit obtained by the Random Tree based Committee Learning was 84.5%, while that of C4.5(J48) using a decision tree constructed by information gain was 64.3%. We consider two possibilities to explain this difference, and check the validities of the assumptions.

The first possibility is the strong heterogeneity among the 1326 enterprises in the objective data in terms of the relations between the profits and the past financial settlements. The accounting rules of the enterprises in practice is heterogeneous, since industrial sectors, management styles and size of the enterprises listed in the market are diverse. Due to this diversity of the data generation



(a) LT debt to equity. (b) Net profit margin.

Fig. 1. Two types of attribute histograms

processes, the relation between the profits and the past financial settlements represented in the data is potentially heterogeneous, and thus the heterogeneous relations may not be sufficiently captured by a unique classifier such as C4.5. Under this condition, the classification approaches to use multiple base classifiers in conjunction with a strategic ensemble learning such as Random Tree based Committee Learning may well capture the relations rather than the approaches to use a unique classifier.

The evaluation of the varieties of the relations among the attributes is difficult in generic sense. Here, we check the varieties from some view points of the data distributions for each attribute. If the data contains various relations among the attributes, the data distributions are supposed to be complex or widely spread. Fig. 1 (a) is the histograms of an attribute named "LT debt to equity" over the profit increased/decreased enterprises in the original data where they are unimodal distributions. (b) is the histograms of "Net profit margin" where they have one side unimodal distributions due to the limitations of their value ranges. The average values of their distributions are almost identical over the two data sets of increased and decreased profits. Each attribute in the objective data has one of these two shapes of distributions, and hence the heterogeneity of the data is not observed from the distribution type of each attribute. Next, we evaluated the volume of the tail parts of each distribution to check if the data contains significant outliers. If the number of the outliers is significant, the relations in the data can be diverse. Table 4 shows representative 60 attributes and their fractions of the enterprises within $\pm 2\sigma$ (std) from the average for both enterprises of increased and decreased profits. When the distribution follows the normal distribution, the fraction should be within $95.44\% \pm 2.62\%$ (std). Accordingly, almost all attributes in the both data sets of increased and decreased profits do not contain more outliers than expected from the normal distribution. Based on these observations, the objective data forms a unique cluster over the both class data sets, and thus the heterogeneity of the data is not concluded.

The second possibility to explain the high accuracy of Random Tree based Committee Learning is that many attributes contribute to the enterprise profit. Random Tree based Committee Learning is an ensemble classifier to use many Random Trees in each of which many attributes tend to be used due to the random attribute selection. Whereas, the unique decision tree based classifiers such as C4.5(J48) does not use very many attributes since the attributes are selected by some information measures. This observation introduces the aforementioned possibility.

Table 4. Fractions of enterprises within $\pm 2\sigma$

Attribute	Inc.[%]	Dec.[%]	Attribute	Inc.[%]	Dec.[%]
Current ratio	95.7	96.1	Return on total assets	99.2	95.9
Attribute 2	96.4	96.1	Return on total equity	98.6	95.5
Quick ratio	95.5	95.7	Gross margin ratio	93.8	94.3
Attribute 4	95.5	96.5	% gross margin ratio	98.9	95.9
Days sales in accs.	99.5	99.2	Attribute 35	95.5	94.3
% days sales in accs.	99.8	97.3	Attribute 36	92.9	97.0
Inventory turnover	98.6	98.1	Pretax income to sales	96.9	99.4
% inventory turnover	98.9	98.8	% pretax income to sales	97.1	99.0
Inventory / t-assets	95.7	95.3	Net profit margin	97.7	97.6
% inventory / t-assets	95.3	95.1	% net profit margin	99.4	97.0
% inventory	96.7	98.6	Sales to total cash	96.9	96.1
% sales	95.6	95.7	Sales to accs.receivable	99.5	99.2
% depreciation	98.0	99.2	Sales to inventory	98.3	95.7
% dividend per share	98.6	99.0	Attribute 44	99.9	95.8
Depreciation / plant assets	97.8	95.9	Sales to working cap.	97.6	96.7
% depreciation / plant assets	97.0	95.1	% Sales to working cap.	98.6	99.4
Return on opening equity	98.7	99.6	Sales to fixed assets	99.9	98.8
% return on opening equity	97.8	99.6	Attribute 49	98.6	94.5
% (cap. exp. / t-assets)	96.8	95.5	Attribute 50	97.3	94.5
% (cap. exp. / t-assets) 1y lag	95.2	93.5	% advertising expense	98.9	98.3
Debt - equity ratio	99.5	96.7	% (advertising / sales)	97.0	93.7
Attribute 22	99.4	99.4	% total assets	95.0	96.3
LT debt to equity	96.8	96.9	Cash flow to t-assets	96.4	95.3
% LT debt to equity	98.7	96.7	Working cap. / t-assets	99.9	96.1
Equity to fixed assets	99.3	96.7	% working cap. / t-assets	99.9	94.9
Attribute 26	99.6	99.4	Op. income / t-assets	99.8	97.2
Time interest earned	97.7	99.2	% op. income / t-assets	99.9	95.3
% time interest earned	99.3	97.8	% LT debt	98.2	97.1
Sales / t-assets	95.0	95.7	% working cap.	99.0	99.2

Table 5 shows the numbers of the representative 60 attributes used in the classifiers of the Random Tree based Committee Learning and the C4.5(J48). The Random Tree based Committee Learning consists of 50 Random Trees where the best classification accuracy was obtained. Because an attribute sometimes appears at multiple nodes in a decision tree, the numbers of the attributes become more than one for the C4.5. In case of the Random Tree based Committee Learning using 50 Random Trees, the numbers are almost more than 100 which imply that many attributes appear more than twice in each Random Tree. This fact supports that the class of the objective data can be accurately predicted by a classifier such as the Random Tree based Committee Learning which uses many attributes in the data.

Table 5. Comparison of number of attributes between two classifiers

Attribute	RC	C4.5	Attribute	RC	C4.5
Current ratio	105	3	Return on t-assets	114	1
Attribute 2	110	2	Return on t-equity	94	0
Quick ratio	102	0	Gross margin ratio	120	1
Attribute 4	113	0	% gross margin ratio	102	3
Days sales in accs. Receivale	102	0	Attribute 35	20	0
Inventory turnover	122	0	Pretax income to sales	110	2
Inventory / t-assets	111	0	Net profit margin	118	1
% inventory / t-assets	124	1	% net profit margin	110	0
% inventory	109	0	Sales to t-cash	126	1
% sales	137	2	Sales to accs.receivable	103	0
% depreciation	108	0	Sales to inventory	121	2
% dividend per share	130	2	Attribute 44	125	1
Depreciation / plant assets	124	2	Sales to working cap.	96	0
% depreciation / plant assets	121	2	% Sales to working cap.	95	2
Return on opening equity	102	1	Sales to fixed assets	111	3
% return on opening equity	107	1	Attribute 49	82	5
% (cap. exp. / t-assets)	119	2	Attribute 50	98	4
% (cap. exp. / t-assets) 1y lag	134	0	% advertising expense	115	0
Debt - equity ratio	125	0	% (advertising / sales)	94	0
Attribute 22	115	3	% t-assets	108	4
LT debt to equity	126	4	Cash flow to t-assets	120	0
% LT debt to equity	121	0	Working cap. / t-assets	115	0
Equity to fixed assets	112	1	% working cap. / t-assets	107	0
Attribute 26	105	0	Operating income / t-assets	114	2
Time interest earned	110	1	% operating income / t-assets	117	2
% time interest earned	112	0	% LT debt	112	1
Sales / t-assets	123	0	% working cap.	101	2
% sales / t-assets	131	0	Net income over cash flows	102	1

RC stands for Random Tree based Committee Learning using 50 trees.

Table 6 indicates the correlation coefficients of the representative 60 attributes with the class in the objective data. If some attributes are highly correlated with the class, a classifier such as C4.5 can well predict the class based on the limited number of the attributes. However, all attributes in the table show very low correlations. This observation also supports that the accurate class prediction is not achieved by the limited number of the attributes in the objective data but by the use of the various attributes in the data.

Table 6. Correlation coefficients of attributes and class

Attribute	Cor. Coef.	Attribute	Cor. Coef.
Current ratio	0.068	Return on t-assets	-0.075
Attribute 2	-0.049	Return on t-equity	-0.033
Quick ratio	0.055	Gross margin ratio	-0.065
Attribute 4	-0.018	% gross margin ratio	0.038
Days sales in accs. Receivale	-0.025	Attribute 35	-0.014
% days sales in accs. Receivable	0.006	Attribute 36	-0.011
Inventory turnover	-0.033	Pretax income to sales	-0.101
% inventory turnover	0.019	% pretax income to sales	-0.063
Inventory / t-assets	0.019	Net profit margin	-0.011
% inventory / t-assets	-0.027	% net profit margin	0.013
% inventory	-0.065	Sales to t-cash	0.053
% sales	-0.099	Sales to accs.receivable	-0.015
% depreciation	-0.016	Sales to inventory	0.044
% dividend per share	-0.028	Attribute 44	0.010
Depreciation / plant assets	0.045	Sales to working t-cap.	0.010
% depreciation / plant assets	-0.060	% Sales to working t-cap.	-0.047
Return on opening equity	-0.055	Sales to fixed assets	0.022
% return on opening equity	-0.050	Attribute 49	0.051
% (cap. exp. / t-assets)	-0.015	Attribute 50	-0.003
% (cap. exp. / t-assets) 1y lag	0.036	% advertising expense	-0.034
Debt - equity ratio	0.032	% (advertising / sales)	0.020
Attribute 22	-0.033	% t-assets	-0.109
LT debt to equity	0.008	Cash flow to t-assets	-0.069
% LT debt to equity	0.010	Working t-cap. / t-assets	-0.003
Equity to fixed assets	0.034	% working t-cap. / t-assets	-0.026
Attribute 26	-0.023	Operating income / t-assets	-0.075
Time interest earned	0.023	% operating income / t-assets	-0.052
% time interest earned	0.032	% LT debt	0.008
Sales / t-assets	-0.041	% working t-cap.	0.036
% sales / t-assets	-0.019	Net income over cash flows	-0.043

To further investigate the dependency of the prediction accuracy on the number of attributes in a decision tree and its size, we evaluated the accuracy of C4.5(J48) for the data under some tree pruning options as indicated in Table 7. “Conf.” is a threshold parameter to estimate pessimistic accuracies before and after pruning a subtree. In general, if “Conf”. is higher, then the pruning becomes less. “Min. Obj.” is another threshold parameter to prune a leaf if the number of the instances in the leaf is less than “Min. Obj.” If it is smaller, then the pruning becomes less. “Attr.#” is the number of attributes appeared in a decision tree, and “Tree Size” is the number of the decision nodes in the tree. As easily understood, the less pruning, *i.e.*, using more attributes, provides better prediction accuracy. However, the prediction accuracy obtained even by the unpruned decision tree is far less than those provided by the ensemble learning classifiers in Table 1. This again indicates that many attributes contribute to the enterprise profit.

Table 7. Predication accuracies of variously pruned C4.5

Conf.	Min. Obj.	Attr.#	Tree Size	Acc.	Conf.	Min. Obj.	Attr.#	Tree Size	Acc.
1.00	25	19	49	60.6%	0.25	2	36	149	64.3%
0.01	2	29	119	63.4%	1.00	2	40	167	64.4%
1.00	10	34	97	64.2%	1.00	1	38	159	65.7%

In a short summary, the relations between the profits and the past financial settlements are complex, associated with various attributes and implicitly embedded in the objective data. For the accurate prediction of the future profit of an enterprise, a classifier able to include the effects from wide varieties of financial attributes such as Random Tree based Committee Learning is needed.

6 Discussion and Conclusion

In this report, the characteristics of the profits and financial settlements data obtained from the enterprises listed on the first section of Tokyo stock exchange market have been analyzed through the classification modeling based on machine learning and statistical techniques. This analysis clarified that the profit of an enterprise is determined by various financial factors in the past rather than some limited number of dominant factors. Consequently, a classifier which can capture influences from various attributes to a class attribute provides significant accuracy such as Random Tree based Committee Learning. Moreover, the homogeneity of the relations between the profits and the financial settlements became clear. The relations are not significantly different among the listed enterprises. Most of the listed enterprises adjust the settlements by using multiple cost accounting methods for the optimality of the management and the tax saving. However, as these cost accounting methods are regulated and disclosed by financial authorities, most of the enterprises follow a uniform scheme of the accounting. This situation may result the homogeneity of the data.

The accurate profit prediction model given in this report is expected to provide an efficient measure to reduce the negative risk and enhance the positive risk in the enterprise management, since the financial situation required to increase the future profit of the enterprise can be found by the model. The model will also provide an excellent measure to decrease the negative risk and increase the positive risk for the investment in the stock of the enterprise. If the model indicates the future increase of the enterprise's profit, the investor will buy the stock otherwise sell. Though the model developed in this report is empirical, and does not provide any comprehensive insight on the profit mechanism, it enables data oriented and scientific management and investment in economy.

Acknowledgment

This research was partially supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research, 16300045 and 17650042, 2006.

References

1. Suda, K., Yada, K., Shiba, K.: Making an earnings prediction model to gain excess returns: The test of the efficient market hypothesis, RCSS Discussion Paper Series, Vol.26 (2005) 1–24
2. Han, J., Kamber, M.: Data Mining: Concepts and Techniques; Second Edition, The Morgan Kaufmann Series in Data Management Systems (2006)
3. Witten, I.H, Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Series in Data Management Systems (2005)
4. Weka 3 - Data Mining with Open Source Machine Learning Software in Java, http://www.cs.waikato.ac.nz/ml/weka/
5. Nikkei Needs, http://www.nikkeimm.co.jp/service/macro/needs/index.html
6. Cutler, A, Zhao, G.: PERT - Perfect Random Tree Ensembles, Computing Science and Statistics, (Frontiers of Data Mining and Bioinformatics, Procom the 33rd Symposium on the Interface), **33** (2001) 490–497
7. Seung, H.S., Opper, M., Sompolinsky, H.: Query by Committee, Proc. of the Fifth Annual Workshop on Computational Learning Theory, (1992) 287–294
8. Breiman, L.: Bagging Predictors, Machine Learning, **26** (2) (1996) 123–140

Unusual Condition Detection of Bearing Vibration in Hydroelectric Power Plants for Risk Management

Takashi Onoda[1], Norihiko Ito[1], and Kenji Shimizu[2]

[1] Central Research Institute of Electric Power Industry, Communication & Information Laboratory, 2-11-1 Iwado Kita, Komae-shi, Tokyo 201-8511 Japan
{onoda, norihiko}@criepi.denken.or.jp

[2] Kyushu Electric Power Co.,Inc., 2-1-82 Watanabe-Dori, Chuo-ku, Fukuoka 810-8720 Japan
shimizu@kyuden.co.jp

Abstract. Kyushu Electric Power Co.,Inc. collects different sensor data and weather information to maintain the safety of hydroelectric power plants while the plants are running. In order to maintain the safety of hydroelectric power plants, it is very important to measure and collect the sensor data of abnormal condition and trouble condition. However, it is very hard to measure and collect them. Because it is very rare to occur abnormal condition and trouble condition in the hydroelectric power equipment. In this situation, we have to find abnormal condition sign as a risk management from the many sensor data of normal condition. In this paper, we consider that the abnormal condition sign may be unusual condition. This paper shows results of unusual condition patterns detection of bearing vibration. The unusual condition patterns are detected from the collected different sensor data and weather information by using one class support vector machine. The result shows that our approach may be useful for unusual condition patterns detection in bearing vibration and maintaining hydroelectric power plants. Therefore, the proposed method is one method of risk management for hydroelectric power plants.

1 Introduction

Recently, electric power companies have begun to try to shift a Time Based Maintenance (TBM) to a Condition Based Maintenance (CBM) for electric equipment management to realize an efficient maintenance and reduce the cost of maintenance[1]. TBM is to check and change the equipment based on the guaranteed term recommended by makers. And CBM is to check, repair and change the equipment based on the state of equipment. The state consists of the present state of equipment, the operation term of equipment, the load in an operation, and etc. Therefore, this CBM is a kind of risk management for electric power companies' management.

In order to realize the CBM, it is important for electric power companies to collect the data of equipment. The data consist of normal operation information

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 317–331, 2007.

of equipment, abnormal operation information of equipment, and etc. Especially, to reduce the maintenance cost by the accuracy improvement of the equipment management and maintenance based on CBM, it is the most important to collect much data utilized to make management and maintenance.

It is necessary to collect and analyze the past abnormal condition data and the past trouble condition data, in order to discover an abnormal condition signs of power equipment. For instance, there are the discovery of an abnormal condition signs from sensor information of the hydroelectric power plant bearing vibration and the discovery of an abnormal condition sign form the operation data of the power generation plant. However, it is very rare to occur abnormal condition and trouble condition in the power equipment. And in order to collect the abnormal and trouble condition data, it is hard to construct experimental power generation plant and hydroelectric power plant. Because its cost is very high.

In the above situation, Kyushu Electric Power Co.,Inc. has been analyzing the causal relation between the bearing vibration and various sensor information to establish the detection method of an abnormal condition sign for the bearing vibration in the hydroelectric power plant. However, it is hard to discover the causal relation between the bearing vibration and various sensor information. Because the relation between the bearing vibration and various sensor information is too complex, and it is impossible to acquire the abnormal and trouble condition data and we can measure the normal condition data only.

In order to discover abnormal condition signs of hydroelectric power plants, Kyushu Electric Power Co.,Inc. and Central Research Institute of Power Industry are researching the detection method of the unusual condition for the bearing vibration using various sensor information, now. In our research, we consider that the unusual condition pattern coincide with the abnormal condition pattern nearly, because we can measure the normal condition patterns only from a regular operation hydroelectric power plant. And we developed the detection method of the unusual condition for bearing vibration from regular condition data in the hydroelectric power plant.

In this paper, we describe the measured sensor data briefly in the next section. In the third section, we briefly explain our proposed approach to detect an abnormal condition sign for bearing vibration and our proposed detection method of the unusual condition for bearing vibration from regular operation condition data. The experimental results are shown in the forth section. Finally, we conclude the concept of the abnormal sign discovery of the hydroelectric power plant bearing vibration based on the detection method of the unusual patterns.

2 Measurement Data

Table 1 shows the outline of a hydroelectric power plant. The hydroelectric power plant has various sensors to measure data related to bearing vibration. In this paper, the measurement data is collected from the hydroelectric power plant and

Table 1. Outline of a targeted Hydroelectric Power Plant

<table>
<tr><td colspan="2">Generated Output</td><td>18,000kW</td></tr>
<tr><td colspan="2">Working Water</td><td>$45m^3/s$</td></tr>
<tr><td colspan="2">Effective Head</td><td>46.3m</td></tr>
<tr><td colspan="2">Turbine Type</td><td>Vertical Shaft Francis Turbine</td></tr>
<tr><td colspan="2">Rated Revolutions Per Minute</td><td>240rpm</td></tr>
<tr><td rowspan="4">Bearing Type</td><td>Upper Bearing</td><td>Oil Self Contained Type Segment Bearing (Natural Cooling)</td></tr>
<tr><td>Bottom Bearing</td><td>Oil Self Contained Type Segment Bearing (Natural Cooling)</td></tr>
<tr><td>Turbine Bearing</td><td>Oil Self Contained Type Cylindrical Bearing (Natural Cooling)</td></tr>
<tr><td>Thrust Bearing</td><td>Oil Self Contained Type Pivot Spring Bearing (Natural Cooling)</td></tr>
<tr><td colspan="2">Operation Pattern</td><td>Process Control Operation
(An operation pattern is generated at everyday.)</td></tr>
</table>

analyzed by our proposed method. The measurement data, which is related to bearing vibration, is collected from March 16, 2004 to November 23, 2004.

One data has been composed of the sensor and weather information on 38 measurement items shown in the table 2 for five seconds the measurement interval. All measurement data is normal condition data and does not include an abnormal condition such as accidental condition, trouble condition, and etc.

Table 2 shows one example of the measured data value.

3 Abnormal Condition Sign Detection

In this section, we describe the outline of the approach of detecting the abnormal condition sign using the unusual condition patterns.

Generally, the discovery of an abnormal sign is to detect a peculiar case that appears only before an existing abnormal condition by comparing between normal conditions and abnormal conditions. However, it is a fact that there is little data of an abnormal conditions in the electric power equipment, because the electric power plants is designed with the high safety factor and maintained appropriately. Currently, our bearing vibration data of the hydroelectric power plant also does not have abnormal conditions and accidental conditions. Therefore, it is impossible to detect a peculiar case before abnormal conditions and accidental conditions happen, because it is hard to take the abnormal or accidental conditions and it is impossible to compare normal conditions with abnormal or accidental conditions. Then, we think the relation between a peculiar condition before an abnormal condition happen (hereafter, we call it the abnormal condition sign) and unusual conditions as the following relation.

$$\text{The abnormal condition sign} \approx \text{The unusual condition.}$$

It is possible to change the discovery of the abnormal sign condition to the detection of the unusual condition in the normal conditions. In other words, it

Table 2. An Example of Online Collected Sensor Information

Measurement Item	Value
Time	04:20:45
Generated Output(MW)	16.26
Generated Reactive Power(Mvar)	-0.708
Generator Voltage(V-W)(kV)	10.901
Generator Current(R)(A)	893
Guide Vane Opening(%)	85
Revolutions Per Minute(rpm)	242
Generator Stator Winding Temperature(U Phase)(℃)	74.4
Generator Stator Winding Temperature(V Phase)(℃)	75.0
Generator Stator Winding Temperature(W Phase)(℃)	74.8
Dam Water Level(m)	-2.49
Surge Tank Water Level(m)	128.73
Tail Water Outlet(m)	85.08
Outside Temperature(℃)	25.9
Room Temperature(℃)	28.6
Water Temperature()	27.4
Oil Cooler Inlet Air Temperature(℃)	31.2
Oil Cooler Outlet Air Temperature(℃)	40.6
Upper Bearing Temperature(℃)	50.7
Bottom Bearing Temperature(℃)	51.0
Upper Bearing Oil Temperature(℃)	47.6
Turbine Bearing Temperature(℃)	44.1
Turbine Bearing Oil Temperature(℃)	37.4
Thrust Bearing Temperature(℃)	59.8
Bottom Oil Tank Oil Temperature(℃)	47.8
Air Duct Outlet Air Temperature(℃)	50.7
Bottom Bearing Inlet Air Temperature(℃)	29.0
Pressured Oil Tank Oil Temperature(℃)	35.6
Generator Shaft Vibration(X axis)(μm)	22
Turbine Shaft Vibration(X axis)(μm)	51
Air Tank Pressure(MPa)	3.167
Pressured Oil Tube Oil Pressure(MPa)	2.732
Generator Bottom Bearing Oil Tank Oil Surface(mm)	-4
Generator Upper Thrust Bearing Oil(mm)	3
Turbine Bearing Oil Surface(mm)	-6
Pressured Oil Tank Oil Surface(mm)	-58
Pressured Oil Tube Oil Surface(mm)	17
Main Shaft Sealed Water Volume(ℓ/min)	-0.44

is assumed that the unusual condition with low probability of existing in the normal condition data has high probability of abnormal condition sign.

The figure 1 shows a concept of detection of unusual condition pattern in the normal condition data. In this figure, the gray oblique line area denotes the normal condition pattern area. In this research, the unusual condition patterns are detected from this normal condition patterns. From the figure 1, if we can find a hyper-sphere, which can cover the 99% of the normal condition patterns, we can think that the other 1% patterns unusual condition patterns. This 99% normal condition patterns are called "general condition patterns". In the figure 1, the inside of a circle shown a black solid line is general condition pattern area, and black stars denote unusual condition patterns. Therefore, if we can find a boundary of α% area in the normal condition area correctly, it is possible to detect unusual condition patterns which do not belong the α% area of normal condition patterns[2]. We adopt One Class Support Vector Machine (hereafter One Class SVM) to find the boundary correctly[3][4].

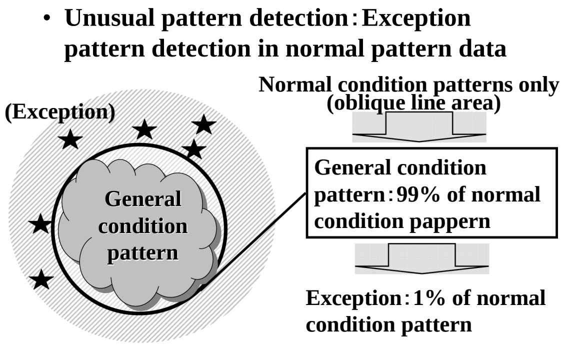


Fig. 1. Image of Unusual Condition Detection

4 One Class SVM

Schölkopf et al. suggested a method of adapting the SVM methodology to one class classification problem[4]. Essentially, after transforming the feature via a kernel, they treat the origin as the only member of the second class. The using "relaxation parameters" they separate the image of the one class from the origin. Then the standard two class SVM techniques are employed.

One Class SVM [4] returns a function f that takes the value +1 in a "small" region capturing most of the training data points, and -1 elsewhere.

The algorithm can be summarized as mapping the data into a feature space H using an appropriate kernel function, and then trying to separate the mapped vectors from the origin with maximum margin(see Figure 2).

Let the training data be

$$\mathbf{x}_1, \ldots, \mathbf{x}_\ell \tag{1}$$

belonging to one class X, where X is a compact subset of R^N and ℓ is the number of observations. Let $\Phi : X \to H$ be a kernel map which transforms the training examples to feature space. The dot product in the image of Φ can be computed by evaluating some simple kernel

$$k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) \tag{2}$$

such as the Gaussian kernel

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{\|\mathbf{x} - \mathbf{y}\|^2}{c}\right). \tag{3}$$

The strategy is to map the data into the feature space corresponding to the kernel, and to separate them from the origin with maximum margin. Then, To separate the data set from the origin, one needs to solve the following quadratic program:

$$\begin{aligned} &\min_{\mathbf{w}\in H, \xi\in R^\ell \rho\in R^N} \quad \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu\ell}\sum_i \xi_i - \rho \\ &\text{subject to} \;\; (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0. \end{aligned} \tag{4}$$

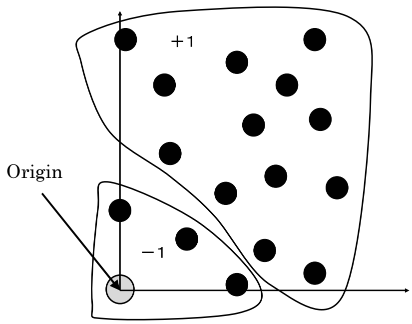


Fig. 2. One Class SVM Classifier: the origin is the only original member of the second class

Here, $\nu \in (0, 1)$ is an upper bound on the fraction of outliers, and a lower bound on the fraction of Support Vectors.

Since nonzero slack variables ξ_i are penalized in the objective function, we can expect that if $\mathbf{w}$ and ρ solve this problem, then the decision function

$$f(\mathbf{x}) = \operatorname{sgn}\left((\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho\right) \tag{5}$$

will be positive for most examples $\mathbf{x}_i$ contained in the training set, while the SV type regularization term $\|w\|$ will still be small. The actual trade-off between these two is controlled by ν. For a new point $\mathbf{x}$, the value $f(\mathbf{x})$ is determined by evaluating which side of the hyperplane it falls on, in feature space.

Using multipliers $\alpha_i, \beta_i \geq 0$, we introduce a Lagrangian

$$\begin{aligned} L(\mathbf{w}, \boldsymbol{\xi}, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu\ell}\sum_i \xi_i - \rho \\ & - \sum_i \alpha_i((\mathbf{w} \cdot \mathbf{x}_i) - \rho + \xi_i) - \sum_i \beta_i \xi_i \end{aligned} \tag{6}$$

and set the derivatives with respect to the primal variables $\mathbf{w}, \xi_i, \rho$ equal to zero, yielding

$$\mathbf{w} = \sum_i \alpha_i \mathbf{x}_i, \tag{7}$$

$$\alpha_i = \frac{1}{\nu\ell} - \beta_i \leq \frac{1}{\nu\ell}, \quad \sum_i \alpha_i = 1. \tag{8}$$

In Eqn. (7), all patterns $\{\mathbf{x}_i : i \in [\ell], \alpha_i > 0\}$ are called Support Vectors. Using Eqn. (2), the SV expansion transforms the decision function Eqn. (5)

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho\right). \tag{9}$$

Substituting Eqn. (7) and Eqn. (8) into Eqn. (6), we obtain the dual problem:

$$\min_{\boldsymbol{\alpha}} \ \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{10}$$

$$\text{subject to } \ 0 \leq \alpha_i \leq \frac{1}{\nu\ell}, \ \ \sum_i \alpha_i = 1. \tag{11}$$

One can show that at the optimum, the two inequality constraints Eqn. (4) become equalities if α_i and β_i are nonzero, i.e. if $0 < \alpha \leq 1/(\nu\ell)$. Therefore, we can recover ρ by exploiting that for any such α_i, the corresponding pattern $\mathbf{x}_i$ satisfies

$$\rho = (\mathbf{w} \cdot \mathbf{x}_i) = \sum_j \alpha_j \mathbf{x}_j \cdot \mathbf{x}_i. \tag{12}$$

Note that if ν approaches 0, the upper boundaries on the Lagrange multipliers tend to infinity, i.e. the second inequality constraint in Eqn. (11) become void. The problem then resembles the corresponding *hard margin* algorithm, since the penalization of errors becomes infinite, as can be seen from the primal objective function Eqn. (4). It is still a feasible problem, since we have placed no restriction on ρ, so ρ can become a large negative number in order to satisfy Eqn. (4). If we had required $\rho \geq 0$ from the start, we would have ended up with the constraint $\sum_i \alpha_i \geq 1$ instead of the corresponding equality constraint in Eqn. (11), and the multipliers α_i could have diverged.

In our research we used the LIBSVM. This is an integrated tool for support vector classification and regression which can handle One Class SVM using the Schölkopf etc algorithms. The LIBSVM is available at *http://www.csie.ntu.edu.tw/*˜cjlin/libsvm.

5 Unusual Pattern Detection Experiment

In this section, we describe our experiment by using the measurement data, which is explained in section 2. Especially, we briefly introduce our experimental setup, how to verify our experimental results, experimental results, and the evaluation.

5.1 Experimental Setup

Our experiment analyzed the measurement data, which is explained in section 2. The measurement data is composed of 38 measurement items. However, in order to detect the unusual condition patterns, we extracted the related measurement items to the bearing vibration from all measurement items. So, 16 measurement items were selected by the expertise of the bearing vibration of the experts to analyze the unusual condition patterns. Table 3 shows these selected 16 measurement items.

In this section, the hydroelectric power plant operations were defined by the expertise of the experts. These operations are the starting condition, the parallel operation condition, the parallel off condition and the stopping condition.

Table 3. Measurement Items for Detection Analysis

Measurement Items for Detection Analysis	
A. Generated Output(MW)	B. Revolutions Per Minute
C. Room Tempe.(℃)	D. Water Tempe.(℃)
E. Oil Cooler Inlet Air Tempe.(℃)	F. Oil Cooler Outlet Air Tempe.(℃)
G. Upper Bearing Tempe.(℃)	H. Bottom Bearing Tempe.(℃)
I. Upper Bearing Oil Tempe.e(℃)	J. Turbine Bearing Tempe.(℃)
K. Turbine Bearing Oil Tempe.(℃)	L. Thrust Bearing Tempe.(℃)
M. Bottom Oil Tank Oil Tempe.(℃)	N. Bottom Bearing Inlet Air Tempe.(℃)
O. Generator Shaft Vibration(X axis)(μm)	P. Turbine Shaft Vibration(X axis)(μm)

Table 4. The Number of Patterns in Each Condition

Group	The number of patterns
Stopping condition	433,935
Starting condition	120
Parallel operation condition	2,368,756
Parallel off condition	132
Total	2,804,113

In practice, these operations should be defined by the operation record of the hydroelectric power plant. However, in order to verify our proposed approach, we defined the four groups by using the expertise of the experts intentionally.

The starting condition patterns and the parallel off condition patterns are very few in our dataset relatively. The parallel operation condition patterns are very large. If we analyze the all measurement data to detect the unusual condition patterns, the detected condition patterns are the starting condition patterns or the parallel off condition patterns. This is not good situation for our analysis. So, the all measurement data is divided into the following four groups by expertise of experts.

Starting condition group. Generator Voltage(V-W) $<$ 10kV and Guide Vane Opening $\geq$ 10% and Revolutions Per Minute $\geq$ 200 rpm.

Parallel operation condition group. Generator Voltage(V-W) $\geq$ 10kV and Revolutions Per Minute $\geq$ 200 rpm.

Parallel off condition group. Generator Voltage(V-W) ¡ 10kV and Guide Vane Opening $<$ 10% and Revolutions Per Minute $\geq$ 200 rpm.

Stopping condition group. Otherwise.

These groups defined by the expertise of the experts. Table 4 shows the number of patterns in each group.

In the stopping condition group, the bearing does not rotate. This group patterns were omitted from the analyzed patterns. In other words, the unusual condition patters were detected in each group, which is starting condition or parallel operation condition or parallel off condition. In order to ignore the different measurement units, the measurement data is normalized into the average 0 and the variance 1 at each measurement item.

Table 5. Unusual condition patterns in the starting condition patterns

Measurement Item	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
A.Generated Output(MW)	-0.192	-0.192	-0.192	*-0.192*	*-0.192*	-0.192	*-0.162*	-0.192	-0.180
B.Revolutions Per Minute	239	224	238	*243*	*243*	238	*330*	233	224
C.Room Tempe.(℃)	15.0	16.1	15.2	*23.1*	*23.1*	29.8	*28.8*	27.4	26.8
D.Water Tempe.(℃)	10.1	10.9	10.4	*21.1*	*21.1*	27.9	*27.4*	25.8	24.5
E.Oil Cooler Inlet Air Tempe.(℃)	20.3	22.9	18.0	*25.9*	*25.9*	36.6	*30.9*	28.6	27.9
F.Oil Cooler Outlet Air Tempe.(℃)	25.4	27.6	21.8	*35.0*	*34.8*	40.9	*40.5*	29.9	28.5
G.Upper Bearing Tempe.(℃)	40.0	41.0	36.2	*46.4*	*46.5*	50.3	*50.8*	32.2	30.6
H.Bttm Bearing Tempe.(℃)	36.3	35.1	32.9	*44.3*	*44.3*	44.5	*50.9*	31.2	29.6
I.Upper Bearing Oil Tempe. (℃)	35.9	36.8	32.1	*42.9*	*42.8*	47.7	*47.5*	31.6	29.5
J.Turbine Bearing Tempe.(℃)	34.1	31.7	30.8	*41.1*	*41.1*	40.8	*43.8*	24.6	22.5
K.Turbine Bearing Oil Tempe.(℃)	32.3	30.0	29.2	*33.1*	*33.1*	38.2	*37.0*	24.8	22.5
L.Thrust Bearing Tempe.(℃)	40.4	41.3	36.5	*54.6*	*54.8*	51.3	*59.7*	32.1	30.4
M.Bottom Oil Tank Oil Tempe.(℃)	33.3	32.5	29.7	*40.5*	*40.4*	43.3	*47.8*	30.1	28.6
N.Bottom Bearing Inlet Air Tempe.(℃)	18.8	21.0	18.1	*24.3*	*24.5*	33.5	*29.2*	29.1	27.8
O.Generator Shaft Vibration (X axis)(μm)	11	22	16	*12*	*8*	15	*23*	11	14
P.Turbine Shaft Vibration (X axis)(μm)	21	33	31	*34*	*27*	35	*33*	20	22

5.2 Detection Results and Evaluation

The unusual condition patterns were detected in each group data of the starting condition, the parallel operation condition and the parallel off condition by applying One Class SVM, which is introduced in section 3. Our experiments made the trial that changes the detected number of unusual condition patterns. In this section, we reports the analysis result to each group data in the number of unusual condition patterns. The experts judged that the number was appropriate.

Detection of the unusual condition patterns in the starting condition patterns. The starting condition patterns were applied One Class SVM to detect unusual condition patterns in these patterns. The parameter ν, which can determine the rate of the unusual condition patterns in the analyzed patterns, was set 0.05. In this case, $\nu = 0.05$ denotes that the number of the unusual condition patterns is about six.

We could detect 9 unusual condition patterns by the analysis. Table 5 shows the typical unusual condition patterns, which are detected by our experiment.

Figure 3 shows an existence probability of the starting condition patterns. In this figure, x axis denotes values of discriminate function and y axis denotes a cumulative probability. The discriminate function values of the unusual condition patterns, which are shown in the table 5, are smaller than 0. In the figure 3, the existence probabilities of the unusual condition patterns, whose discriminate function values are smaller than 0, are smaller than 0.05 and very rare condition patterns.

Table 6. Data Before and After Unusual Condition Case 4 and 5

Measurement Items	-15 sec.	-10 sec.	-5sec.	0 sec. Case 4	+5 sec. Case 5	+10 sec.	+15 sec.	+20 sec.
A.Generated Output(MW)	4.272	4.32	4.452	*-0.192*	*-0.192*	4.26	2.352	4.23
B.Revolutions Par Minute	243	243	243	*243*	*243*	243	243	243
C.Room Tempe.(℃)	23.2	23.1	23.1	*23.1*	*23.1*	23.1	23.2	23.1
D.Water Tempe.(℃)	21.2	21.0	21.1	*21.1*	*21.1*	21.2	21.2	20.9
E.Oil Cooler Inlet Air Tempe.(℃)	25.9	26.1	25.9	*25.9*	*25.9*	26.1	26.1	26.1
F.Oil Cooler Outlet Air Tempe.(℃)	34.8	35.0	34.9	*35.0*	*34.8*	34.9	34.8	34.9
G.Upper Bearing Tempe.(℃)	46.4	46.6	46.5	*46.4*	*46.5*	46.3	46.6	46.6
H.Bottom Bearing Tempe.(℃)	44.3	44.3	44.3	*44.3*	*44.3*	44.5	44.2	44.2
I.Upper Bearing Oil Tempe. (℃)	42.9	42.8	42.8	*42.9*	*42.8*	43.0	43.0	42.9
J.Turbine Bearing Tempe.(℃)	41.1	41.1	41.1	*41.1*	*41.1*	41.1	41.1	41.1
K.Turbine Bearing Oil Tempe.(℃)	33.1	32.9	33.2	*33.1*	*33.1*	33.0	33.2	33.0
L.Thrust Bearing Tempe.(℃)	54.8	54.8	54.8	*54.6*	*54.8*	54.6	54.8	54.6
M.Bottom Oil Tank Oil Tempe.(℃)	40.4	40.6	40.4	*40.5*	*40.4*	40.6	40.3	40.6
N.Bottom Bearing Inlet Air Tempe(℃)	24.5	24.4	24.5	*24.3*	*24.5*	24.6	24.4	24.5
O.Generator Shaft Vibration (X axis)(μm)	14	11	9	*12*	*8*	11	9	10
P.Turbine Shaft Vibration (X axis)(μm)	40	25	26	*34*	*27*	31	34	31

In the table 5, the case 4 and 5 should belong to the parallel operation condition actually. Table 6 shows the data before and after 15 seconds of the case 4 and 5. In the table 6, we can recognize that the case 4 and 5 should belong to the parallel operation condition. This situation shows that the expertise of experts could not describe the starting condition completely. However, our proposed approach can detect the case 4 and 5, which should belong to the parallel operation condition, as the unusual condition patterns. This fact shows that our approach can find the appropriate unusual condition patterns, which should be detected in the starting condition. In the table 5, the case 3 is an accidental interruption. This fact was researched by using an operation record, an daily report, and etc.

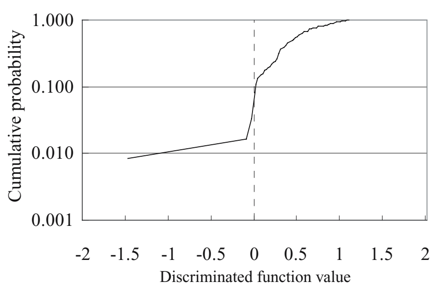


Fig. 3. The Relation between Classification Value and Existence Probability for Starting condition patterns

Table 7. Unusual condition patterns in the parallel operation patterns (1)

Measurement Item	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
A.Generated Output(MW)	1.602	2.112	1.542	2.622	1.122	16.68	16.26
B.Revolutions Per Minute	242	242	242	242	242	243	242
C.Room Tempe.(℃)	15.0	14.8	15.0	15.0	12.1	23.3	28.6
D.Water Tempe.(℃)	10.2	10.3	10.7	10.3	12.2	20.8	27.4
E.Oil Cooler Inlet Air Tempe.(℃)	18.4	18.3	16.9	16.7	11.1	25.9	31.2
F.Oil Cooler Outlet Air Tempe.(℃)	24.2	24.4	22.9	23.2	22.0	34.3	40.6
G.Upper Bearing Tempe.(℃)	37.9	38.1	37.0	37.3	39.6	45.8	50.7
H.Bottom Bearing Tempe.(℃)	35.0	35.2	34.6	35.6	44.3	42.6	51.0
I.Upper Bearing Oil Tempe.(℃)	34.1	34.2	33.0	33.2	35.1	42.2	47.6
J.Turbine Bearing Tempe.(℃)	32.5	32.9	32.9	33.5	43.3	39.7	44.1
K.Turbine Bearing Oil Tempe.(℃)	29.4	29.3	29.1	29.0	35.5	31.9	37.4
L.Thrust Bearing Tempe.(℃)	39.6	40.3	39.1	40.6	46.9	53.9	59.8
M.Bottom Oil Tank Oil Tempe.(℃)	32.1	32.4	31.6	32.0	39.2	38.7	47.8
N.Bottom Bearing Inlet Air Tempe.(℃)	16.3	16.3	15.9	15.7	12.1	24.5	29.0
O.Generator Shaft Vibration(X axis)(μm)	18	21	21	19	24	17	22
P.Turbine Shaft Vibration(X axis)(μm)	40	35	42	44	87	26	51

The accidental interruption denotes a power plant parallel off when the power system has an accident, and is very rare case.

Detection of the unusual condition patterns in the parallel operation condition patterns. The parallel operation condition patterns were applied One Class SVM to detect unusual condition patterns in the parallel operation patterns. The parameter ν, which can determine the rate of the unusual condition patterns in the analyzed patterns, was set 5×10^{-6}. In this case, $\nu = 5 \times 10^{-6}$ denotes that the number of the unusual condition patterns is about twelve. This analysis can detect the unusual condition patterns, whose existence probabilities are smaller than 5×10^{-6}.

We could detect 14 unusual condition patterns by the analysis. Table 7 and 8 show the unusual condition patterns, which are detected by our experiment. In the table 8, the three cases (case 10-12) of these unusual condition patterns were test operation. This fact was researched by using an operation record, an daily report, and etc. The test operation denotes a short term operation to check the electric power plant, and is very rare case. This fact shows that our approach can find the appropriate unusual condition patterns, which should be detected in the parallel operation condition.

Detection of the unusual condition patterns in the parallel off condition patterns. The parallel off condition patterns were applied One Class SVM to detect unusual condition patterns in these patterns. The parameter ν, which can determine the rate of the unusual condition patterns in the analyzed patterns, was set 0.05. In this case, $\nu = 0.05$ denotes that the number of the unusual condition patterns is about twelve. This analysis can detect the unusual condition patterns, whose existence probabilities are smaller than 0.05.

We could detect 11 unusual condition patterns by the analysis. Table 9 and 10 show the unusual condition patterns, which are detected by our experiment.

Table 8. Unusual condition patterns in the parallel operation condition patterns (2)

Measurement Item	Case 8	Case 9	Case 10	Case 11	Case 12	Case 13	Case 14
A.Generated Output(MW)	15.792	14.64	*-0.192*	*-0.192*	*-0.192*	-0.18	1.632
B.Revolutions Per Minute	242	243	*232*	*231*	*229*	233	242
C.Room Tempe.(℃)	28.6	28.7	*26.8*	*26.8*	*26.7*	27.6	12.4
D.Water Tempe.(℃)	27.4	27.6	*25.6*	*25.6*	*25.6*	25.7	13.1
E.Oil Cooler Inlet Air Tempe.(℃)	30.8	30.6	*28.1*	*28.0*	*28.0*	28.6	12.8
F.Oil Cooler Outlet Air Tempe.(℃)	40.3	40.0	*35.3*	*35.3*	*35.3*	30.6	22.5
G.Upper Bearing Tempe.(℃)	50.5	50.7	*45.6*	*45.6*	*45.8*	36.8	39.2
H.Bottom Bearing Tempe.(℃)	50.7	50.9	*43.5*	*43.7*	*43.8*	32.8	44.2
I.Upper Bearing Oil Tempe.(℃)	47.3	47.4	*42.4*	*42.6*	*42.4*	34.7	35.1
J.Turbine Bearing Tempe.(℃)	43.8	43.7	*39.9*	*40.0*	*39.9*	27.4	43.8
K.Turbine Bearing Oil Tempe.(℃)	36.5	36.0	*31.1*	*31.1*	*31.1*	25.2	36.3
L.Thrust Bearing Tempe.(℃)	59.7	59.4	*52.9*	*52.9*	*53.1*	38.1	46.8
M.Bottom Oil Tank Oil Tempe.(℃)	47.8	48.1	*40.3*	*40.2*	*40.2*	30.8	39.0
N.Bottom Bearing Inlet Air Tempe.(℃)	29.2	29.1	*26.8*	*26.7*	*26.8*	28.5	11.7
O.Generator Shaft Vibration(X axis)(μm)	20	22	*23*	*18*	*12*	10	23
P.Turbine Shaft Vibration(X axis)(μm)	54	63	*70*	*54*	*47*	24	87

Table 9. Unusual condition patterns in the parallel off operation patterns (1)

Measurement Item	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
A.Generated Output(MW)	-0.21	-0.192	-0.192	-0.192	-0.192	-0.192
B.Revolutions Per Minute	227	208	239	233	240	208
C.Room Tempe.(℃)	15.6	16.4	13.8	16.9	22.8	22.7
D.Water Tempe.(℃)	9.4	9.6	13.2	14.7	20.6	20.4
E.Oil Cooler Inlet Air Tempe.(℃)	17.6	18.6	13.9	20.1	26.3	26.3
F.Oil Cooler Outlet Air Tempe.(℃)	27.4	28.0	24.1	29.1	32.0	31.9
G.Upper Bearing Tempe.(℃)	42.4	42.3	40.3	43.0	42.6	42.8
H.Bottom Bearing Tempe.(℃)	45.6	45.6	44.7	44.7	38.1	38.1
I.Upper Bearing Oil Tempe.(℃)	38.3	38.3	36.2	39.0	39.0	39.3
J.Turbine Bearing Tempe.(℃)	42.4	42.6	42.6	41.5	35.5	35.6
K.Turbine Bearing Oil Tempe.(℃)	35.5	35.0	34.8	34.0	31.9	31.9
L.Thrust Bearing Tempe.(℃)	50.2	49.9	47.9	50.9	45.6	45.8
M.Bottom Oil Tank Oil Tempe.(℃)	41.0	41.1	39.7	40.2	35.4	35.4
N.Bottom Bearing Inlet Air Tempe.(℃)	16.4	17.1	14.1	17.9	24.6	24.6
O.Generator Shaft Vibration(X axis)(μm)	11	11	12	12	13	13
P.Turbine Shaft Vibration(X axis)(μm)	35	33	34	33	28	23

In the table 10, the case 8 and 9 of these unusual condition patterns were an accidental interruption. The other two cases (the case 10 and 11) of the unusual condition patterns were a test operation. These facts were researched by using an operation record, an daily report, and etc. The test operation and the accidental interruption are very rare case. These facts show that our approach can find the appropriate unusual condition patterns, which should be detected in the parallel off condition.

5.3 Inexplainable Unusual Condition Patterns

In our detection experiment, the trial that changes the extracted number of unusual condition patterns was carried out, and the detected unusual condition patterns were shown to experts. It is easy for experts to explain that the detected

Table 10. Unusual condition patterns in the parallel off operation patterns (2)

Measurement Item	Case 7	Case 8	Case 9	Case 10	Case 11
A.Generated Output(MW)	-0.192	*-0.18*	*-0.192*	*-0.192*	*-0.192*
B.Revolutions Per Minute	242	*261*	*208*	*225*	*236*
C.Room Tempe.(℃)	29.9	*28.8*	*28.8*	*26.7*	*27.6*
D.Water Tempe.(℃)	28.7	*27.6*	*27.6*	*25.7*	*25.6*
E.Oil Cooler Inlet Air Tempe.(℃)	32.3	*31.1*	*31.1*	*28.0*	*28.6*
F.Oil Cooler Outlet Air Tempe.(℃)	41.1	*40.6*	*40.5*	*35.3*	*30.8*
G.Upper Bearing Tempe.(℃)	50.8	*50.6*	*50.7*	*45.6*	*36.6*
H.Bottom Bearing Tempe.(℃)	51.6	*50.7*	*50.9*	*43.7*	*32.5*
I.Upper Bearing Oil Tempe.(℃)	48.1	*47.6*	*47.7*	*42.6*	*34.6*
J.Turbine Bearing Tempe.(℃)	48.1	*44.0*	*43.9*	*40.0*	*26.8*
K.Turbine Bearing Oil Tempe.(℃)	43.0	*37.0*	*37.0*	*31.1*	*25.3*
L.Thrust Bearing Tempe.(℃)	59.4	*59.7*	*59.7*	*53.1*	*37.2*
M.Bottom Oil Tank Oil Tempe.(℃)	48.8	*47.7*	*47.9*	*40.3*	*30.9*
N.Bottom Bearing Inlet Air Tempe.(℃)	31.0	*29.2*	*29.4*	*26.6*	*28.2*
O.Generator Shaft Vibration(X axis)(μm)	12	*17*	*10*	*12*	*9*
P.Turbine Shaft Vibration(X axis)(μm)	46	*37*	*37*	*34*	*23*

unusual condition patterns, which are larger than the above described number of the unusual condition patterns, are normal condition patterns.

However, there are some inexplainable unusual condition patterns in the above described unusual condition patterns. In our experiment, our approach could detect nine unusual condition patterns in the starting condition. The two patterns should belong to the parallel operation condition. The other one pattern is the accidental interruption. However, the rest six patterns are inexplainable unusual condition patterns. For the parallel operation condition patterns, our approach could detect fourteen unusual condition patterns. The three patterns could be explain as the test operation, but the rest eleven unusual patterns are inexplainable unusual condition patterns. For the parallel off condition patterns, our approach could detect eleven unusual condition patterns. The four patterns could be explain as the test operation or the accidental interruption, but the rest seven unusual patterns are inexplainable unusual condition patterns. These inexplainable unusual condition patterns are very important for the discovery of the abnormal condition sign.

It is easy for experts to manage the explainable unusual condition patterns, because the experts understand the evidence of generation of the unusual condition patterns. However, it is difficult to manage the inexplainable unusual condition patterns, because the experts can not understand the evidence of generation of the unusual condition patterns without the experience of the different abnormal conditions. Therefore, the following two parts are important for the discovery of abnormal condition signs and the risk management for hydroelectric power plants.

- The first part is to discover the unusual condition patterns by a system, to evaluate the discovered unusual condition patterns and select the inexplainable unusual condition patterns by human experts.

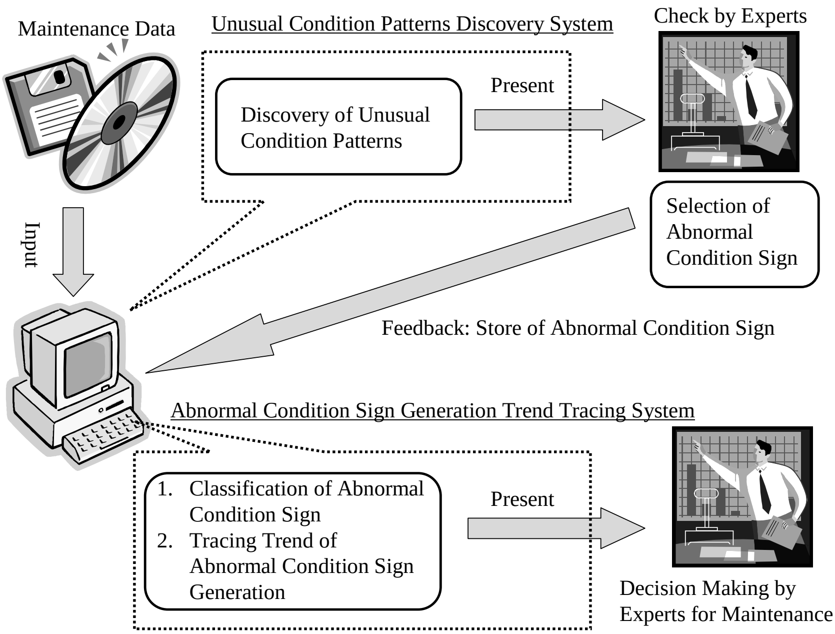


Fig. 4. The Concept of Discovery of Abnormal Condition Sign

– The second part is to classify the inexplainable unusual condition patterns as abnormal condition signs in unseen data and trace the trend of generation of the inexplainable unusual condition patterns.

Figure 4 shows our concept of discovery of abnormal condition signs for hydroelectric power plants. In the other words, figure 4 shows our concept of a risk management system for hydroelectric power plants. In the upper part of this figure, the system discovers the unusual condition patterns and presents the patterns to human experts. Then the human experts evaluate the patterns and select the inexplainable unusual condition patterns as abnormal condition signs. And the human experts give the selected unusual condition patterns to the system. In the lower part of the figure 4, the system classify the abnormal condition signs in unseen data by using the selected unusual condition patterns. And the system traces the trend of the generation of abnormal condition signs. Then the human experts get the trend of the generation of abnormal condition signs from the system and determine how to maintain hydroelectric power plants.

6 Conclusion

In this paper, we described an unusual condition detection of bearing vibration in hydroelectric power plants for the discovery of abnormal condition signs. This approach is a kind of the risk management for hydroelectric power plants. We proposed an unusual condition detection of bearing vibration in hydroelectric power plants base on the One Class SVM. And we show that our proposed approach could find the unusual condition patterns in hydroelectric power plants by using expertise of experts. Moreover, we proposed a concept of the discovery of abnormal condition signs by using the detected unusual condition patterns, and introduced a system image for the abnormal condition signs discovery. It is hard for our problem to evaluate effectiveness of our approach, because we do not have an abnormal condition and a trouble condition in hydroelectric power plants in Japan. However, our proposed approach can analyze an abnormal condition sign without an abnormal condition.

In our future work, we plan to develop our proposed system and evaluate the effectiveness of our system in hydroelectric power plants from risk management point of view.

References

1. Yamana, M., Murata, H., Onoda, T., Oohashi, T., Kato, S.: Comparison of pattern classification methods in system for crossarm reuse judgement on the basis of rust images. In: Proceedings of Artificial Intelligence and Applications 2005. (2003) 439–444
2. Onoda, T., Ito, N., Shimizu, K.: Unusual condition detection of bearing vibration in hydroelectric power plants. In Ras, Z., Tsumoto, S., Zighed, D., eds.: Mining Complex Data Proceedings of a Workshop held in Conference with 2005 IEEE International Conference on Data Mining. (2005) 73–80
3. Schölkopf, B., Platt, J., Shawe-Taylor, J., A, J.S., Williamson, R.: Estimation support of high-dimensional distribution. Technical Report MSR-TR-99-87, Microsoft Research (1999)
4. Schölkopf, B., Smola, A., Williamson, R., Bartlett, P.: New support vector algorithms. Neural Computaion **12** (2000) 1083 – 1121

Structural Health Assessing by Interactive Data Mining Approach in Nuclear Power Plant

Yufei Shu

CCSE, Japan Atomic Energy Agency
6-9-3 Higashiueno, Taitou-ku, Tokyo 110-0015, Japan
shu.yufei@jaea.go.jp

Abstract. This paper presents a nonlinear structural health assessing technique, based on an interactive data mining approach. A data mining control agent emulating cognitive process of human analyst is integrated in the data mining loop, analyzing and verifying the output of the data miner and controlling the data mining process to improve the interaction between human user and computer system. Additionally, an artificial neural network method, which is adopted as a core component of the proposed interactive data mining method, is evolved by adding a novelty detecting and retraining function for handling complicated nuclear power plant quake-proof data. To demonstrate how the proposed technique can be used as a powerful tool for assessment of structural status in nuclear power plant, quake-proof testing data has been applied.

Keywords: Cognitive model, Interactive data mining, Structural health assessing, Information inference.

1 Introduction

There are 53 Nuclear Power Plant (NPP) units in operation in Japan, with designed around 40-years lifespan. Among these NPP units, 7 units have been in operation more than 30 years, and 20 units have been in operation more than 20 years. Some of these power plants are entering the final phases of their design life and are becoming structurally deficient. To maintain a higher level reliability and safety of the NPP and to effectively use limited maintenance resource, it is essential to understand the true structure condition and rate of degradation of each important part or subsystem of the NPP infrastructure. The field of structural health assessing has explosive growth in the past two or three decades. However, despite exponential technological progress, use of these technologies to deal with nuclear power plants is still practically limited. For example, signal acquisition and process techniques provide real time assessment, but for the large scale infrastructure, i.e. NPP, we only can afford to deploy limited sensor. Also model-based finite element method (FEM) is widely employed to analysis structural behavior. For the objective of determining if damage is present, focusing on individual parts in the NPP, a model-based FEM analysis can work very well. However, if an infrastructure with thousands of parts, like a real case in NPP, the answer is hard to determine immediately. For example, within large-scale

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 332 – 345, 2007.

3-dimensional virtual earthquake test-bed (3D-VET), which is constructing by Japan Atomic Energy Agency (JAEA) on a grid computational environment to numerically simulate the behavior of the entire NPP, under static load excitation, the computing time took 7633,232 seconds by using 64PE CPUs, and generated massive data set of 3.5GB [1]. In addition, there are still some unsolved technical issues for model-based numerical simulation, such as model is limited by the capability of their mathematical description. Most of physical models are simplified or just stripped out such factors [2]. To increase the accuracy, the large mesh is used. However, the mesh is larger, the time of computation is longer and size of data is bigger. At another hand, in the nuclear industry, there are a huge amount of plant quake-proof data, which were created by model-based simulation, or collected from experiment data, and signal data. Clearly, the main issue facing for a structural health assessment is not the lack of data, but rather how to process and analyze those collected massive amount of raw data. Therefore, the challenging tasks for assessing NPP infrastructure health assessment are 1) if it is possible to develop model-free statistical data analysis method to represent behavior of plant infrastructure, and furthermore, 2) if it is possible to employ such model-free data analysis method to screen real plant structure condition and integrate it into model-based analysis platform to provide reliable information regarding the integrity of the plant structure, rapidly and accurately.

Among the various data analysis and interpreting approaches, the data mining technique is considered as an advanced means [3]. The core components of data mining technology have been under development for decades and various models have been proposed, for example classical statistics regression analysis, decision tree [4], rough sets [5], the machine learning based method (generic programming, SVM, artificial neural network, etc.), and so on. These technologies have already proven successful in a variety of application domains, such as stock market assessment and prediction [6], medical diagnosis and protein structure discovery [7, 8], aircraft component failure detection [3], analysis of insurance packages [9], and structural health monitoring [10, 11]. However, the application of the data mining technique to handle massive nuclear power plant quake-proof information is a new attempt. Compared to other application domains, the nuclear industry has more complex databases with more indigestible physical implication related to plant parameters. In addition, for power plant infrastructure health assessment, it is impossible to either acquire ground truth data for all possible damage scenarios that are likely to occur and also to acquire in advance for some critical situations. These characteristics of plant quake-proof data require that human experts collaborate more closely to the computer-based data mining system to guide data analysis process and to test hypotheses.

To meet the requirements of using data mining technology to interpret nuclear power plant quake-proof information, the first objective of this research program is to establish an interactive data mining approach. To improve the interaction between users and computers, the data mining agent based on human cognitive model will be integrated into an automated data mining loop. The second objective is to develop an intelligent information inference system based on the proposed interactive data mining method. This inter-disciplinary computer science and structural engineering information inference system will synergize the resolution of the basic technical challenge and speed up the discovery of knowledge related to two types of infrastructure deterioration in NPP, 1) progressive deterioration in time due to environmental effects, and 2) sudden deterioration due to severe loading events (caused by natural hazards or man-made events). It aims to contribute to the vitality of the NPP, as rehabilitation, renewal, replacement, and maintenance of infrastructure are required to mitigate the effects of human operating errors, natural disasters (i.e. extra-large earthquake, tsunami), or terrorism.

Following in the section 2, the framework of the proposed method is explained. In section 3, based on the method, an information inference system is developed. To demonstrate the validation of proposed method, typical NPP structure deterioration data are applied to the developed system in section 4. At last, the conclusion is given.

2 Conceptual Framework of Interactive Data Mining Method

Various machine learning techniques such as genetic programs (GPs), neural networks (NN), case-based reasoning (CBR), etc., have been employed to handle wide spectrum data mining tasks. Among these techniques, neural networks (NN) offer the advantage of dealing with unseen but similar data. NN have powerful capability as a universal approximation of nonlinear mapping functions. It has been applied to different fields of science and engineering, for example, detecting the defect of a chemical processing plant [12], predicting typhoon damage to electric power system [13]. There has also been increasing interest in using it as a statistical pattern classifier to identify and predict behavior of structural systems [14, 15, 16]. In our study, the neural networks approach is adopted as a core component of the interactive data mining method to represent and predict structure damage typically encountered in the NPP.

2.1 Adverseness of Neural Network in Interpreting Plant Quake-Proof Data

However, neural network approach has an inherent disadvantage because of its black-box nature. For example, information or knowledge embedded in trained neural networks is hard to be verified or interpreted by human beings. Lack of an explanation facility becomes the major barrier of using neural network to interpreting NPP quake-proof data, which is desirable to enable the extraction of rules from trained neural network for analysts to gain a greater understanding of the problem at hand.

Furthermore, after training, the neural network structure is fixed. The networks recognize new type data as if it is “unknown”. The point of controversy among researchers is that in the realistic NPP, it is impossible to get ground true data of all possible damage cases or some critical situations. Therefore, when the new situations occur, the trained NN cannot recognize those data and classify it. It becomes another disadvantage of using neural networks to identify structural damage under real NPP operational situation. Obviously, the ability to detect new type data and to re-train the NN with these data is required. It is therefore of considerable practical benefit to develop a data interpreting method, which is inherently able to decide if input data sets are genuine members of the known classes, and if they simply correspond to new data classes.

Moreover, data mining is a computer-based discovery process, which does not require human to create the hypotheses about relationship of data. It does however require human's guidance to some degree. Data mining should be an interactive process, which allows the user to be front and center in the mining process. Analysts must provide specific guidance to the data mining system, for example selecting the portion of data to explore, or by specifying potential target concepts. Also, whenever the user chooses to, the neural network should be able to progressively refine its structure to give better solutions. In other words, the human analyst should be incorporated into the data mining process and play an important role in analyzing NPP quake-proof data.

2.2 Interactive Data Mining Approach

To overcome above or part of above adverse aspects of NN for interpreting plant quake-proof data, neural networks approach is evolved by adding three functions: novelty detection, re-training, and explanation. Data mining control agent based on the cognitive model has been introduced to help to carry out these functions. Figure 1 shows the conceptual framework of the proposed interactive data mining approach.

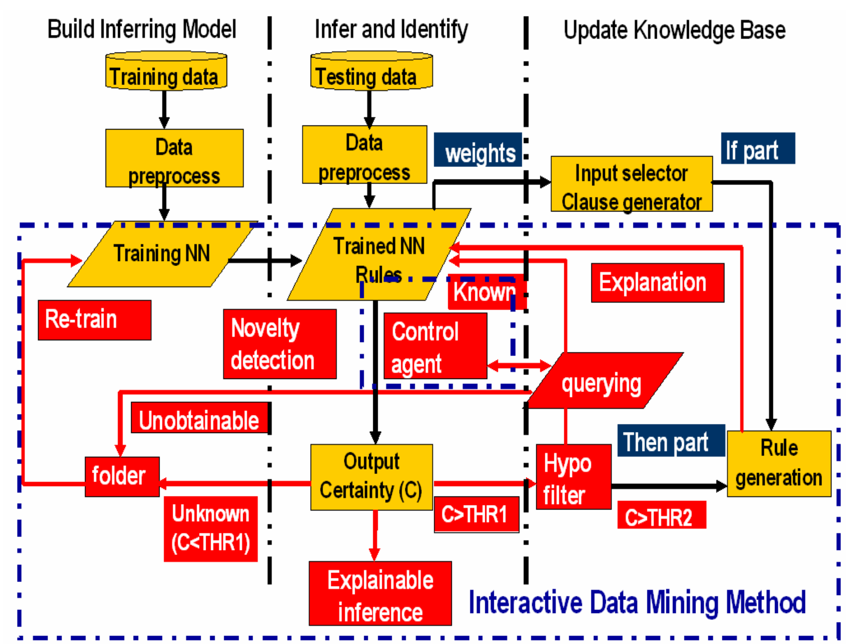


Fig. 1. Block diagram of interactive data mining method

Novelty detection and retraining functions are added to NN. Here, novelty is implied the quake-proof data corresponding to the new status of the NPP, which will be the input of NN. This function will help to recognize the complicated potential relation amongst the quake-proof data and new status of the system, which is not known initially. By incorporating the extracted rules into the system, an explanation function is introduced to NN. Rules will be in an “IF-THEN” style to explain the inferred result. In this way we not only predict what happened in the system, but also why it happened. Data mining control agent is developed based on the human cognitive model. It embedded in the data mining process providing guidance to recognize or to detect the new class and verifying the extracted rules by using domain expertise to improve interaction between human user and computer system.

There are three phases included in the interactive data mining approach. The first phase is to build the neural network inferring model. A neural network will be created with randomly weighted links and network structure, then perform the training process. The feed forward network mapping from inputs x_i to output y_k is (1)

$$y_k = f(\sum_m w_{kj} g(\sum_n w_{ji} x_i + w_{j0}) + w_{k0}) \tag{1}$$

here, w_{ji} is the weight of the connection from the i^{th} input neuron to the j^{th} hidden neuron, and w_{kj} denotes the weight from the j^{th} hidden neuron to the k^{th} output neuron, w_{j0} and w_{k0} represent the bias for the hidden and output unit and *g( ), f( )* denote the activation function from input to hidden layer and hidden to output layer, respectively. As a supervised, error correcting and gradient-based technique, back-propagation algorithm [17, 18] is adopted to train the network under a given set of ground truth data. After training, the network size and link weight of the inferring model are decided for a particular application domain.

The second phase is to infer and identify the health condition of the infrastructure with test data sets. In this process, the mining agent is introduced to progressively refine the trained inferring model performance. This agent embedded in the data mining process to provide any obtainable feature to validate the hypotheses, or to detect the novelty of input data, and to decide whether to re-train NN. In our approach, the novelty detection process is modeled as a hypothesis generation by the neural network learning algorithm. The certainty factor of the output neuron *Ci* is calculated over the neuron network inferring model constructed in first phase with activation function *f( )*. The mining control agent will verify the inference result by comparing *Ci* with a particular threshold *THRi*, which is stored in the knowledge base. The value of the threshold *THRi* corresponds to the concrete situations decided by domain dependent knowledge.

(1) If the certainty factor *Ci* of hypothesis exceeds a particular threshold of the certainty factor, THR1, (> THR1) the output is stored into the hypothesis filter.

(2) If the certainty factor of the output neuron *Ci* is high enough, (> THR2), it will be changed to belief and stored into the knowledge base as a new extracted rule to re-use for further prediction.

(3) If the certainty factor of the output neuron is between (THR1< *Ci* <THR2), and when mining agent judged that data accumulated in the hypothesis filter are enough, the system will query if there is any other important obtainable input feature and if

yes, the system will go back to the phase 1 to retrain the NN and update the network size and link weight of inferring model. If there is not new obtainable input, the system will treat these input data as new type data corresponding to the new state and move them into the new data sets folder.

(4) If the certainty factor of the output neuron is low (Ci < THR1), i.e. all outputs show that there is no single clear winning node (class), the input data can be deemed as novel and will be collected. When the mining agent judged that accumulated novel input data are enough in the new data sets folder, the system will go back to the phase 1 to retrain the NN and update the network size and link weight of inferring model.

After the inferring model had been refined, the test data will be applied to the updated network again to identify the structure's health state.

The third phase is to update the knowledge base. After verifying the inference results by the mining control agent, extracted rules are incorporated into the network to update the knowledge base. It is to be noted that these rules are not represented explicitly in the knowledge base at the beginning; they are generated by the inferring system as purpose for explanations, so that not only prediction can be made, but also the reason for the prediction.

Data mining agent guides and controls the mining process, which emulates how a human analyst interprets data sets based on domain knowledge. By introducing the mining control agent, novelty detection, re-training and explain function can be carried out in the data mining loop. It helps to improve interaction between human users and computer system. The proposed interactive data mining method then can accomplish the mission of representing and analyzing complicated quake-proof data, further, of discovering knowledge related to infrastructure deterioration in the nuclear power plant.

3 Architecture of Information Inference System

To realize full potential as a data manager, the inference system should be constructed with different functional capabilities. The domain independent proposal is developed in order to manage different types of data and complete different data manage tasks [14]. The architecture of information inference system consists of three different technical modules: Data miner, Mining control agent, and Knowledge base as shown in Figure 2. The system is developed in a programming language C++ and Java.

The datasets are retrieved from database where plant quake-proof data are stored, such as simulation data, signal data and experience data. With various kind of knowledge prescribed in the knowledge base, the data miner will generate its inference results guided by mining control agent. This distributed and paralleled architecture makes the information inference system have very flexible structure. We can add or remove certain sub-module of Data miner and correspondingly modify knowledge base. So that mining control agent can accomplish wide spectrum data mining task in light to the different data contents and requirements. The primary function of each technical module is as following:

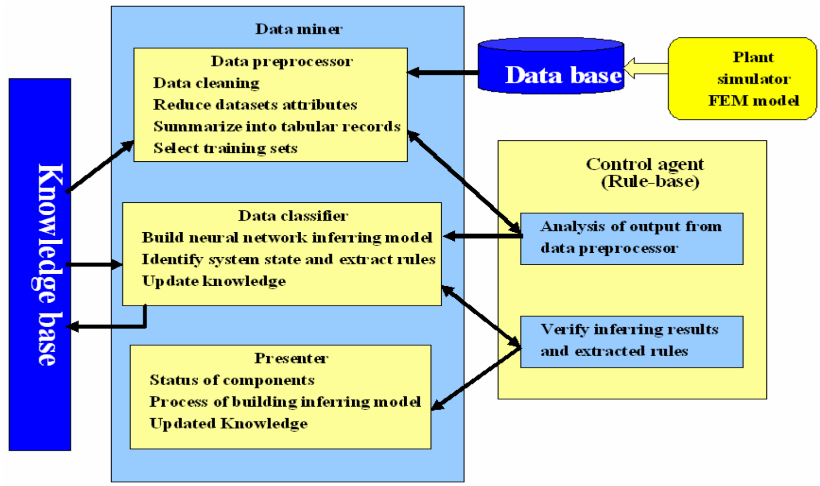


Fig. 2. Architecture of information inference system

Mining control agent emulates human analyst to judge and verify output of data miner. At this stage, we employed rule-based human cognitive model as a basis to develop mining control agent. All cognitive rules are derived from domain dependent expertise.

Knowledge Base stores domain dependent knowledge and case specified knowledge extracted from datasets. On this stage of our study, automatically update knowledge base is not accomplished yet.

Data miner is core of the system, which is domain independent. It includes three distinct but tightly coupled sub-modules: data preprocessor module, data classifier module, and presenter module to carry out the different data mining functions. At this stage, we develop the data miner with main function to classify data based on the proposed interactive data mining approach, as explained in the section 2.2.

1. Data Preprocessor Module The initial work is to retrieve data from database. Assembling structures, such as equipment, pipe, valve, support, and bending pipe, are selected as our study object due to their critical role in power plan safety. At this step, we will analyze data generated from power plant numerical simulator. This sub-module provides the information inference system with FEM simulation results, such as displacement, strain, stress or acceleration of component of infrastructure under static or dynamic loads and boundary conditions to cover real power plant operating environment.

After selecting certain data sets, the Principle Component Analysis (PCA) [19] method is employed to reduce feature of dataset. Then, these data sets are summarized into discrete tabular records containing numbers of reduced essential attributes and save as data files.

2. Data Classifier Module Classifier will chose pre-processed data through call-level interface and use them as training and test data sets. Its primary focuses are to infer if the infrastructure has degraded or in an unsafe status, and if yes, to locate it, and assess the level of the structural damage by interpreting plant quake-proof data. Meanwhile, classifier will extract re-useable rules, which explain the causality of component invisible structure status and measurable parameters. Mining control agent judges and verifies the output of data classifier. Extracted case-specified rules are stored in the knowledge based for reuse.

3. Presenter Module This GUI module is developed in a programming language Java. It provides platform for users to interact with computer-based data mining system. User can select the learning parameters, network initial structure and training, test data through the interface. It also presents the process of training and result from information inference system. Furthermore, the training time, the need to re-train, or if novelty is detected are shown in the interface. In this module, visualization techniques are employed to present the training process and how the inferring model is build up. At this step, we have developed Japanese version GUI, shown in Figure 3.

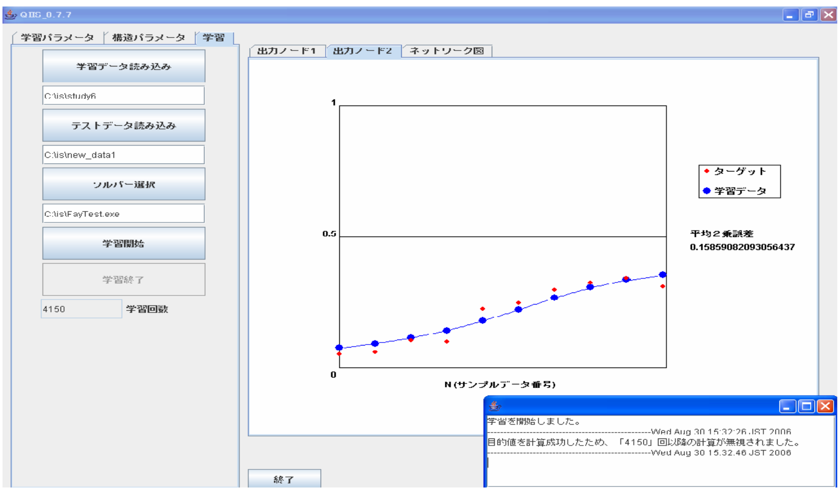


Fig. 3. GUI of information inference system

4 Application: Detection of Structural Damage

There are four levels of damage assessment as the standard across the filed, which are adopted in our study [20].

- Is damage present?(detection)
- Where is the damage located?(localization)
- How sever is the damage?(diagnosis)
- How much longer can the structure perform?(prognosis)

A typical scenario related to infrastructure deterioration in NPP was selected and applied to the developed information inference system to check if the proposed interactive data analysis method can answer above or parts of above questions well.

4.1 Problem Identification and Data Generation

It is known that when structural damage occurs, the structural property and structural dynamic response will be changed accordingly. There is underlying relation between a damage pattern and corresponding structural dynamic response. Therefore, the damage detection problem becomes fundamentally one of statistical pattern recognition.

As a pilot study, a computational model of pipe system with supports was developed as our virtual test-bed, which simulated the pipe system structural configuration and characteristics in the HTTR of JAEA [21]. The pipe is 30 meter long and supported by 5 supports, which are distributed averagely. Left and right end of the pipe was fixed. The structure is built on a hard solid foundation, which is assumed to be rigid. In order to simplify the problem, the following assumptions are made,

1) The pipe segment is approximated as uniform beam of constant mass

2) The displacements of each central point between two supports are considered at vertical directions of the pipe and

3) Deflections of pipe at horizontal direction are neglected

4) The structure damping is neglected

The finite element method using concentrated masses is adopted to derive the pipe system FEM model. In this study, the static analysis of FEM generated deterioration patterns by reducing the stiffness of the support, which was represented by Young's modulus. We assumed that there is only one support damaged at a time. That is, the damage pattern can be classified into 5 classes in accordance with locations as labeled from 1 to 5, shown in Figure 4. Thus, the problem can be defined as task of identifying the pipe system in the presence of deterioration excitations, such as the location of support damaged from the changes of the system dynamic response, and the level of deterioration at certain location corresponding to the initial damage point.

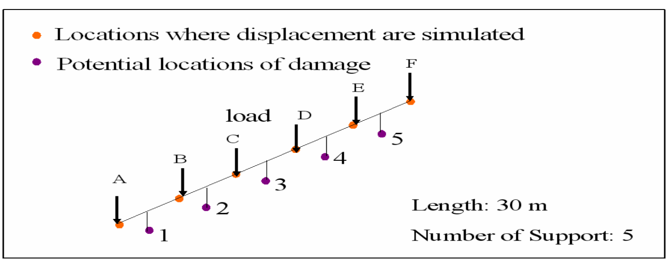


Fig. 4. FEM model for pipe and support system

To generate ground-truth training data, how the system response to the external load is simulated using FEM within the virtual test-bed. We simulated the situations where the stiffness of each supports was reduced by 10, 20, 30, 40 and 50%. There are 25 damage patterns for the supports as training data sets. In addition, the test-bed generated other damage patterns. Firstly, we randomly damage one of those 5 supports with reducing stiffness between 10% and 50%. These 25 damage patterns are served as validation data sets to test how well the network generalizes to new data set. On this basis, the optimal network configuration is chosen according to the minimum error of this validation data set. Since it is not the always case that we know all damage situations at the beginning. Secondly, we generate another two groups of testing data sets, D1 and D2, including some unknown damage situations. D1 contains all known and two unknown damage classes, total 35 data sets and D2 contains three known and one unknown damage classes, 20 data sets. With these data sets, we try to demonstrate how the proposed interactive data mining method can recognize plant structural healthy statues.

4.2 Detail of Data Classifier

The virtual test-bed generated various numerical simulated results, such as stress, strain, displacement and so on. To identification of pipe system deterioration, the essential features are selected from these multi-dimension simulated parameters. We build different type data classifier to carry out different level of damage assessment [20].

For detecting damage and identifying damage location, we use a 2-layer neural network integrated with cognitive model as a statistical data classifier. The structure behavior is characterized by vertical displacements at central point between 2 supports, A-F nodal point. The input patterns consist of these six nodal displacements, and the output units correspond to each support location, labeled 1 to 5, where structural deterioration occurs likely. Sigmoid activation function (2) was used for hidden units.

$$h_i = g(x) = \frac{1}{1+e^{-\beta x}} \tag{2}$$

where *hi* is the value of corresponding out put of hidden unit.

The "softmax" activation function (3) was used for output unit. By using "softmax" function, the posterior probability can be predicted, which is deemed as the output confidence factor. It will be used as a threshold to detect novelty.

$$y_i = e^{a_i} \Big/ \sum_j e^{a_j} \tag{3}$$

Since there are five output units in the network train phase, we set the target to be a vector made up of "1" and "0", where "1" indicates that the damage pattern belongs to that class, i.e., damage is located at the corresponding support, and "0" indicates that the damage pattern does not belong to the class. The data classifier output is a vector of positive values whose sum is 1.0, and each component of the vector represents the posterior probabilities of the corresponding damage class given the input. If the data classifier can predict the damaged foundation with the highest posterior probability, we can conclude that the information inference system successfully detected the damage and the location.

For diagnosis or predicting the damage level, we also employ a 2-layer neural network with cognitive model as a data classifier. In this network, the nodal displacements at central point between two supports, labeled A to F, are used as inputs of network, and the percentage of stiffness reductions of each support are served as output patterns. Sigmoid activation function (2) is selected as activation function for hidden and output layer.

4.3 Training and Test Result Analysis

The data classifier for identification of damage location will be trained in two phases. At first phase, the data classifier is trained with the training data set and test data sets D1 is applied to it. The aim of this phase is to investigate the capability of novelty detection function of data classifier with mining control agent. At second phase, we generate a new training data sets, which contain the original training data and newly added data detected from phase 1 and use it to re-training data classifier. Retrained data classifier is tested by test data sets D2. Re-training is performed incrementally rather than completely new retraining with same number of input and hidden units but with different output units, which corresponds to the damage classes. Before retraining, another attempt is to duplicate new detected data set several times to ensure balanced training, since it has less number of novel samples in training than other known classes.

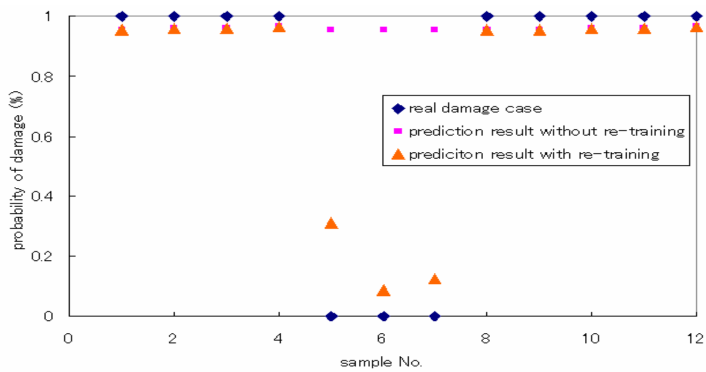


Fig. 5. Real damage location vs. predicted result

We test a situation with and without retraining function. The predicted damage location is compared with the target location, shown in Figure 5, where x axis is sample number and y axis represents probability of damage at one support location. “1” indicates that the damage pattern belongs to this class, i.e., damage is located at the corresponding support, and “0” indicates that the damage pattern does not belong to the class. The prediction accuracy of the inference system is about 75% without retraining function, however, after retraining function is added, the prediction accuracy of the inference system increased to about 83.3%. There are only 2 cases out of 12 data sets were not identified precisely. It shows that the proposed interactive data mining approach can carry out more accurate damage identification even unknown classes including in the data sets. In addition, for most damage patterns, the posterior probability assigned to the damaged support was higher than 97.6%, which implied that the developed inference system can answer the questions, such as is damage present?(detection); where is the damage located?(localization), appropriately based on proposed interactive data mining approach.

Next, we train and test data classifier for predicting of the damage level located at damaged support using training data sets and test data sets D1, and D2. When the test data set D1 is applied to the information inference system, the predicted damage level at support point 4 and 5 was compared to the real damage level, shown in Figure 6, where x axis is sample No. and y axis corresponds to damaged level at support location. We can see that the prediction results of damage level do not dovetailed with real damage level. Since the test data set D1 includes two classes that data classifier treated it as unknown. However, after retraining the data classifier with those newly added classes, the D2 data set is applied to the inference system, the prediction result dovetailed with real damage level very well at point 4 and 5, as shown in Figure 7, where x axis is sample No. and y axis represents damaged level at support location. It demonstrated that with novelty detection and retrain function, the ability to identify damage level of the information inference system is improved considerably.

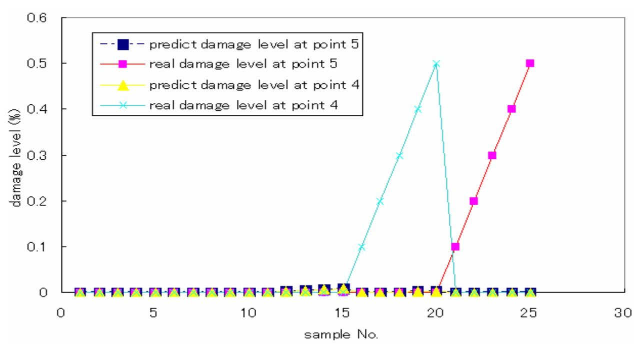


Fig. 6. Real damage level vs. predicted results without novelty detection function

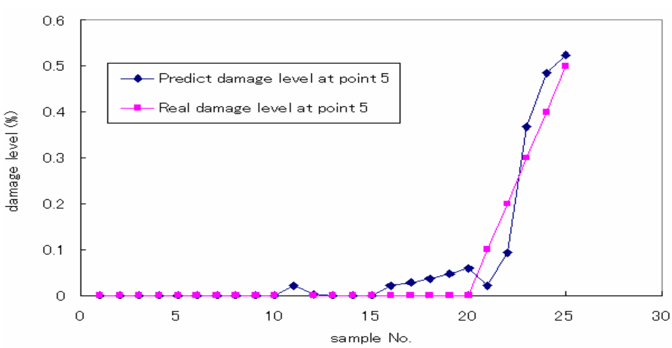


Fig. 7. Real damage level vs. predicted results with novelty detection function

These test results show that the proposed method can identify structural health status accurately and rapidly by adding novelty detection and retraining functions.

5 Concluding Remarks

Using of data mining technology to interpret nuclear power plant quake-proof information for discovering knowledge related to plant infrastructure health status is a new attempt. However, compared to other industries, the nuclear industry has more complex and complete databases with more indigested physical meanings related to health status of thousands parts in NPP, which gives it the competitive advantage and challenge of being able to make good use of the data mining techniques and human domain knowledge in this field.

To deal with complex NPP quake-proof data, an interactive data mining method has been proposed, which integrates human cognitive model in the data mining loop and overcome the adverseness of neural network by adding novelty detection function and retrain ability. An information inference system has been developed based on the proposed interactive data mining approach. This IT-based data analysis framework is expected to develop a practicable structural health assessment methodology, as well as the discovery of new knowledge related to the NPP infrastructure. Through application of plant simulation data to the information inference system, it demonstrated that the developed system can be used as powerful tool to identify the structural health state from massive data, and further, such tool will be of practical benefit for rapid health condition screening and providing reliable information regarding the integrity of the structure under the real operating environment of NPP.

Nevertheless, the cognitive capability of proposed data mining approach current is however limited, only simple rule-based human model is employed at the stage. To prove it will be one of the next step works. Also, at this stage, the system is not integrated with single process technologies. In the future, we will apply the system to process the real time signal data and combine it with numerical simulation analysis method to strengthen conventional approaches. For long-term, the output of the system is expected to be periodically updated information regarding the ability of the structure to perform its intended function in light of the inevitable ageing and degradation resulting from operational environments. For short term, for example after earthquake, the system is expected for rapid condition assessing and aims to provide reliable information regarding the integrity of the infrastructure. It will be a challenge of the next step.

Acknowledgements. Work described in this paper was sponsored by Japan Atomic Energy Agency.

References

1. Matsubara, T., 2006. The large-scale numerical analysis of nuclear power plant in distributed computational environment, Proc. Work on Large-Scale Simulation, Tokyo, pp. 34-40 (in Japanese).
2. Furukawa, T., Hoffman, M., 2005. Implicit material modeling: a challenge to reliable inelastic finite element analysis, IACM expressions, Vol. 17, pp. 20-24.
3. Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P., 1996. From Data Mining to Knowledge Discovery: An Overview, Advances in Knowledge Discovery and Data mining, AAAI/MIT press, Cambridge, MA.

4. J.R.Quinlan., 1993. C4.5: Programs for Machine learning. San Mateo, California: Morgan Kaufmann.
5. Pawlak, Z., 1991. Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publisher, Dordrecht, Boston, London.
6. Mizuno, H., Kosaka, M., and Yajima, H., 1998. Application of Neural network to technical analysis of stock market prediction, Retrieved Sep., 8, 2004, from http://wwww.oco.ro/ici/revosta/soc98_2/art03.html.
7. Langley, P., and Simon, H., 1995. Application of Machine learning and rule induction, Communications on ACM, Vol.38, pp. 55-64.
8. Veropoulos, K., 2002. Machine learning approaches to medical decision making, Phd thesis, Bristol University.
9. Chuang, Y., Hsu, M., and Wang, Y., 2004. Analysis of insurance packages by using data mining technology. Proc. Int. Conf. on Systemics, Cybernetics and Informatics, Florida, CDROM.
10. Chang, C.C., Chang, T.Y.P., Xu, Y.G, and Wang, M.L., 2000. Structural Damage diction using an iterative neural network, Journal of intelligent material systems and structures, Vol.11, pp. 32-42.
11. Worden, K., and Lane, A.J., 2001. Damage Identification Using Support vector machines, Smart Materials and Structures, Vol.10, pp. 540-547.
12. Stites, Robert, L., Ward, Bryan, Walters, Robert, V., 1991. Defect Prediction with Neural networks. Proc. Int. Conf on Analysis of Neural Network Applications, ACM press, pp. 196-206.
13. Taketa, H., Yanase, M., Waki, T., and Hachino, T., 2002. A prediction method of electric power damage by typhoons in Kagoshima via GMDH and NN, Proc. Conf. on Society of Instrument and Control Engineers, Osaka, pp. 2423-2425.
14. Yufei, S., Norihiro, N., 2005. Development of a quake-proof information inference system by using data mining technology, Proc. Int. Conf on Human-Computer Interaction, Nevada, CDROM.
15. Masri, S.F., Smyth, A.W., Chassiakos, A.G., Caughey, T.K., and Hunter, N.F., 2005. Application of neural network for detection of changes in nonlinear systems, Journal of Eng Mechanics, July, pp. 666-676.
16. Choi, M.Y., and Kwon, I.B., 2000. Damage detection system of a real steel truss bridge by neural networks, Smart structures and materials 2000: Smart systems for bridges, structures, and highways, Proc. SPIE, California Vol.3, pp. 295-306.
17. Duda, O.R., Hart, E.P., Stork, G. D., 2001. Pattern Classification, John Wiley & Sons, Inc. New York, pp. 289-319
18. Rumelhart, D., and McClelland, J., 1986. Parallel distributed processing: exploration in the microstructure of cognition, Vol.1, and 2, MIT Press, Cambridge, MA..
19. Neumann, D., 2002. An Enhanced Neural Network Technique for Software Risk Analysis, IEEE Trans Soft Eng, pp. 904-912.
20. Sohn, H., Farrar, R.C., Hemez, M.F., Shunk, D.D., Stinemates, W.D., Nadler, R.B., 2003. A Review of structural health monitoring literature: 1990-2001, The Los Alamos National Laboratory Report LA 13976-MS, pp.1-307.
21. http://www2.tokai.jaeri.go.jp/httr/eng/index_top_eng.html

Developing Mining-Grid Centric e-Finance Portals for Risk Management

Jia Hu, Muneaki Ohshima, and Ning Zhong

Maebashi Institute of Technology
460-1 Kamisadori-Cho, Maebashi 371-0816, Japan
hujia@kis-lab.com
ohshima@wi-lab.com
zhong@maebashi-it.ac.jp

Abstract. E-finance industry is rapidly transforming and evolving toward more dynamic, flexible and intelligent solutions. This paper describes a model with dynamic multi-level workflows corresponding to a multi-layer Grid architecture, for multi-aspect analysis in building e-finance portals on the Wisdom Web. The application and research demonstrate that mining-grid centric three-layer Grid architecture is effective for developing intelligent risk management and decision making financial systems.

This paper concentrates on how to develop an mining-grid centric e-finance portal (MGCFP), not only for supplying effective online financial services for both retail and corporate customers, but also for intelligent credit risk management and decision making for financial enterprises and partners.

1 Introduction

The financial services industry is in transition. Traditionally, financial services have been structured into four broad industry sectors: banking, insurance, equities/stock brokerage, and transaction services [17].

The exponential growth of the Internet during the past decade has essentially altered the landscape of the financial services industry with the initiation of continuously available e-trading services and the adoption of e-transactions [6].

Online trading systems can automatically match buy and sell orders with different trading mechanisms[13]. However, if we look at the overall process of trading in financial markets, online trading has only replaced telephones with the Web and provided a universal interface for individual customers to participate in the financial markets [9].

E-finance encompasses all financial products and services which are available to the consumer through the Internet. The Internet has enabled the expansion of financial services by making it possible to integrate the complex variety of financial data and services, and by providing new delivery channels such as

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 346–359, 2007.

mobile, online banking and investment, but using it to its full potential is often ignored [25].

The existing infrastructures for financial services are overwhelmingly closed, monolithic, and inward-directed. More dynamic, flexible and intelligent solutions on behalf of consumers and/or business operators are needed. In response to these circumstances, new financial model is beginning to emerge to support the integration of the financial processed across industry and national boundaries.

As the huge and multiple data sources are coupled with and geographic distribution of data, users, systems, services, and resources in the specific types of financial industries, the Grid platform is needed as a powerful middleware for developing e-finance portals [10]. Furthermore, workflow management systems address the enterprise process automation problems, which refer to a formal, executable description of a business process [1,2]. This paper concentrates on how to develop an mining-grid centric e-finance portal (MGCFP), not only for supplying effective online financial services for both retail and corporate customers, but also for intelligent risk management and decision making for financial enterprises and partners.

The rest of the paper is organized as follows. Section 2 discusses the evolving financial model and the advantages and problems for implementing this model. Section 3 describes the architecture of an mining-grid centric e-finance portal (MGCFP) corresponding to the proposed model in Section 2 and its main features. Section 4 describes security issues related to developing MGCFP. Section 5 presents a case study for demonstrating the usefulness of the proposed architecture. Section 6 summarizes the related work. Finally, Section 7 gives conclusions and our future work.

2 The Proposed Approach

The finance model is changing. The role of Internet changes from commercial medium and transaction platform to a emerging virtual marketplace mirroring the physical one. Therefore, e-finance is currently facing revolutionary changes: New marketplaces are enabling new kinds of services and interactions among consumers, vendors and suppliers.

The evolving e-finance model we propose is shown in Fig. 1. The model is defined in terms of the actors in the marketplace - customers, portals, and suppliers - and the inter-relations between them. The portal is acting as a marketplace for buyers and sellers to meet, with exclusive access to an integrated set of proprietary products and services as well as links to suppliers/partners that offer complementary financial services and access to any other financial products available through the Internet.

Commercial banks face significant challenges on both the supply side and demand side, associated in particular with competition, product-service quality and differentiation, transaction security, cost efficiency, and demographic change. Portal is exploited as a channel to build, maintain, and develop long-term relationships through products and services.

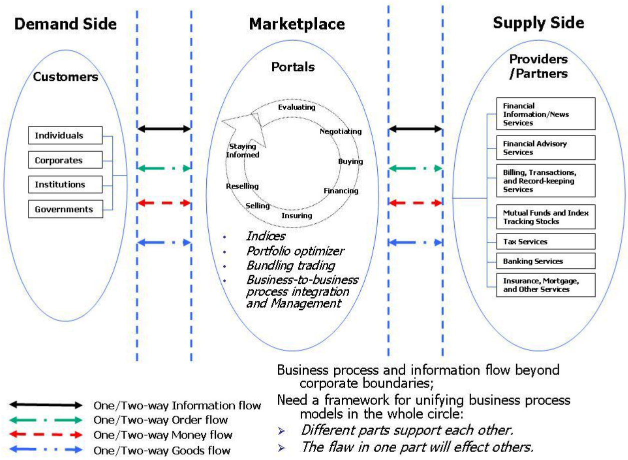


Fig. 1. A strategic financial bundling model

All business processes between the supply side and demand side are, in one-way or another, components comprising four flow processes:

– *Information flow*, primarily the demands
– *Order flow*
– *Money flow*, payments for goods or services
– *Material flow*, delivery of goods or services to fulfill the customers demands.

These are just the primary flows, and goods may actually flow via distributors/suppliers, and money flows may be via credit cards; these may be seen as refinements, alternatives or extensions of the primary flows. Business process model can indeed be mapped as a complex interrelated set of flows of money, materials and information and other resources.

Customers like to access the services within a single interface, usually Web browser. Traditional integrated banks, which exclusively distribute self-developed products via proprietary channels and fulfill all transaction and support services in-house, are no longer adequate for the changing demands of the environment. Despite the convenience and possibilities offered by the e-finance, the issues of trust, security of Internet transaction, lack of legislative framework and reluctance to change impede the growth of e-finance.

Following are some open issues for discussion:

– ***Business Model.*** Business models are usually represented by a mixture of informal textual, verbal, and ad-hoc graphical representations [12]. Web-based financial intermediaries will succeed by offering seamless integration of e-trading, e-banking and other business process.

– ***Enterprise Architecture.*** The service-oriented portal architecture built on software components offers many benefits to enterprises, such as data/service integration for enterprise-wide and cross-enterprise.
– ***Risk Management.*** Global risk information is usually difficult to obtain because it requires real-time data about deals made in different departments and supported by disparate systems.
– ***Security Threats.*** As the money has become the data and information moving around the Internet, the financial organizations have to pay more attention to security than other business sectors. In general, the attacks are aimed at the collection, exploitation, falsification, or destruction of data and/or transactions. As security threats change quickly and are difficult to anticipate, complete safety is impossible [23]. A flexible monitor and fast detection system with quick response to new security threats and adapting appropriate security measures is required for financial organizations.

3 Portal Architecture

Figure 2 shows the architecture of mining-grid centric e-finance portal (MGCFP) that has been developing by us. The architecture comprises a multi-tired, service-oriented solution that offers a high degree of modularity. The solution is available on the open industry standard platforms J2EE. The portal enables the financial enterprise to have a common infrastructure that encapsulates business rules, back-end connectivity logic and transaction behavior, enabling services to write-once, deploy-everywhere, across channels. The solution ensures a unified view of customer interactions to both the customers and the enterprises.

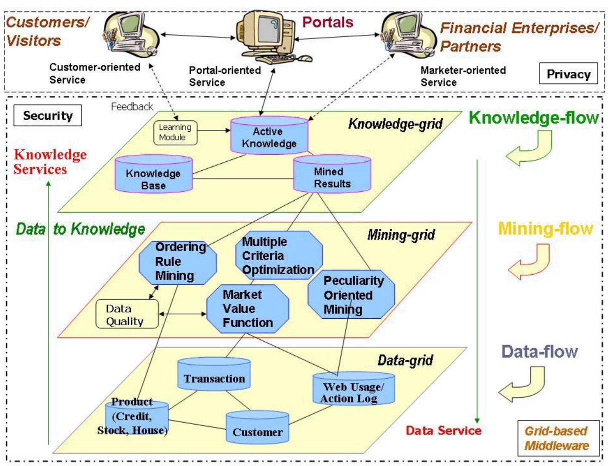


Fig. 2. The architecture of mining-grid centric e-finance portal (MGCFP), corresponding to the strategic e-finance portal shown in Fig. 1

For the individual customers, MGCFP provides them with the facility to check their accounts and do transactions on-line. MGCFP will provide all the bank facilities to its customers when their authentications(user id and password) match, including viewing account information, performing transfers, giving the customer an option of changing address, paying bills online, password retrieval, applying credit card, performing transactions, viewing transactions and locations of the bank and its branches. MGCFP could also support an online enrollment facility for credit card customers and should allow customers to view their personnel accounts and to pay bills online from their account.

In MGCFP, there are mainly four kinds of data sources deployed on the data-grid, namely customer, product, transaction, and Web usage dataset. Various data mining methods are employed as agents on the mining-grid for various service-oriented, multi-aspect data analysis [27]. Furthermore, the rules and hypotheses mined from multiple data sources are stored on the knowledge-grid, so that they will be refined into active knowledge by reasoning and inferring with the existing knowledge. The active knowledge is employed to provide personalized financial services for potential customers, portals, and financial enterprises through the three-layer Grid.

The status-based financial services in the MGCFP are dynamically organized by using the workflow management system. The workflows are divided into three levels, namely data-flow, mining-flow, and knowledge-flow, corresponding to the three-layer Grid, respectively. They are generated dynamically, based on the conditions (situations), data quality analysis, mining process, and available knowledge sources. In this model, lower level applications and services provide middleware support for higher level applications and services, thereby opening the door to developing more complex, flexible, and effective systems.

The architecture is deployed on the top of Grid middleware and services, i.e., it uses basic Grid services to build specific knowledge services. Following the integrated Grid architecture approach, these services can be developed in different ways using the available Grid toolkits and services. The current implementation is based on the Globus toolkit, so that the Grid-enabled applications can be accessed by any end users via a standard Web browser.

3.1 Data-Grid vs. Data Warehouse

There is a strong competitive advantage in acquiring solutions that integrate the channels used and enable a unified management view of customer interactions for all bank services. Traditional e-finance information system is shown in Fig. 3. Banking has traditionally been an integrated business, Where financial institutions exclusively distributed self-developed products via proprietary channels and fulfilled all transaction and support services in-house.

Most e-finance information systems are based on data warehouse. Using the popular definition of a data warehouse as a collection of subject-oriented, integrated, time variant, non-volatile data in support of management decisions. Along with the strengths of the data warehouse architecture are many limitations.

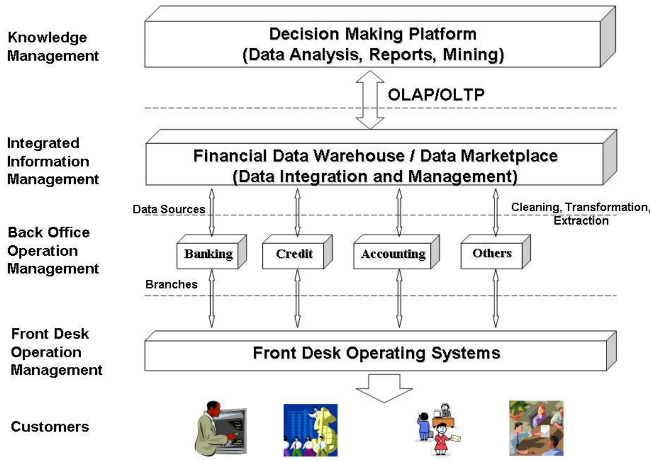


Fig. 3. A traditional e-banking architecture

The first is that it can be very inflexible. The Second limitation is the amount of development and maintenance effort needed to manage a star-oriented data warehouse. And once the data warehouse is built, the maintenance needed to add tables, columns, aggregates and indexes is an ongoing, time-consuming and extremely expensive process.

Therefore, we need a new distributed, flexible infrastructure to develop e-finance portal. As the huge and multiple data sources are coupled with and geographic distribution of data, users, systems, resources, and services in the typical types of enterprises, Grid platform is an ideal middleware or platform for e-finance portals development.

Data storing and retrieving are deployed on the Grid platform, like Globus, as a standard Grid service. OGSA-DAI is used to build database access applications [30]. The architecture is both scalable and transparent. MGCFP is scalable because its design allows for the distribution of different services (represented by objects) among different grid nodes. This is not only means that more nodes can be added to the portal as the number of users increases, but also means that services themselves can be distributed.

3.2 Mining-Grid vs. OLAP

E-finance portal is not simply about cheap transaction platform, although that will become more and more important. There is a whole range of all types of risks - credit, liquidity, interest rate risk and market risk - that need to be taken into account. In some ways, the Internet may heighten these risks.

From the top-down perspective, different data mining methods are deployed on the mining-grid as agents for mining services. On the mining-grid, different mining methods work just like agents, that is to say, they are working in an autonomic, distributed-cooperative mode. There are following three main reasons for developing multiple data mining agents on the mining-grid:

1. Businesses rely on data-driven analysis for decision making. Data mining is increasingly recognized as a key to analyzing, digesting, and understanding the large and growing data collected by financial applications.
2. We cannot expect to develop a single data mining method to solve all problems since complexity of the real world applications. Hence, various data mining agents need to be cooperatively used in the multi-step data mining process for performing multi-aspect analysis as well as multi-level conceptual abstraction and learning.
3. When performing multi-aspect analysis for complex problems, a data mining task needs to be decomposed into sub-tasks. Thus these sub-tasks can be solved by using one or more data mining agents that are distributed over different computers and multi-data repositories on the Grid. Thus the decomposition problem leads us to the problem of distributed cooperative system design.

3.3 Knowledge-Grid for Knowledge Management

Knowledge-grid allows an integrated management of risks such as credit and market risks. From the top-down perspective, the knowledge level, also the application-oriented level, is supported by both the mining level and data level for serving the customers, portals, and enterprise marketers. From the bottom-up perspective, the data level supplies data services for the mining level, the mining level produces new rules and hypotheses for the knowledge level to generate active knowledge.

In general, several kinds of rules and hypotheses can be mined from different data sources by multi-aspect mining. The results cannot be utilized for knowledge services until they are combined and refined into more general ones to form *active knowledge*, by meta-learning and reasoning. Distributed Web inference engines on the knowledge-grid will employ such active knowledge with various related knowledge sources together to implement knowledge services and business intelligence activities [24,26].

4 Security Concerns

Operational risk, including security risk, is of course one of the more frequently mentioned risks in connection with electronic banking. Security is not a new risk. We are all familiar with the various security issues that banks are facing on a day-today basis, e.g. robberies, thefts of ATM machines, frauds. However, banking transactions over the Internet do pose new issues.

Given the open nature of the Internet, transaction security is likely to emerge as the biggest concern among the e-finance's customers (actual and potential).

Since transaction risk would create a significant barrier to market acceptance, its management and control are crucial for business reputation and the promotion of consumer confidence as well as operational efficiency [18]. The customers must be assured that the confidentiality in their transactions must be maintained.

4.1 Security Objective

Accordingly, the fundamental objectives that e-finance portal security arrangements should try to achieve are to:

- restrict access to the system to those users who are authorized;
- authenticate the identity and authority of the parties concerned to ensure the enforceability of transactions conducted through the Internet;
- maintain the secrecy of information while it is in passage over the communications network;
- ensure that the data has not been modified either accidentally or fraudulently while in passage over the network; and
- prevent unauthorized access to the central computer system and database (intrusion).

4.2 Three-Level Security

Security should be integrated into an organization in such a way as to enhance and safeguard each facet in the least intrusive yet most effective way possible at a given time [15]. Therefore, the e-finance portal have to incorporate three-level security measures to protect the transactions from being abused or hacked.

- *Transactional Level Security.* Transactional level security refers to the ability of two entities on the Internet to conduct a transaction; privately and with authentication. SSL (Secure Socket Layer) provides encryption on all data transmitted between Internet, which helps ensure privacy of the data and authentication of the session while preserving the integrity of the message. Moreover, Grid middleware has supplied GSI (Grid Security Infrastructure) layer for security solution in the Grid environment.
- *System Level Security.* System level security helps to protect against corruption of service, and control user access to portal resources. Firewall and password are usually employed in this security level.
- *Application Level Security.* Besides the transaction level and system level security concerns, more concerns are coming from the upper-level applications (such as session management in Web forms) which is highly dependent on the portal architecture and design.

The dynamics at play seem to confirm the aphorism that “security is a process, not a technology.” Instead of trying to arrive at a set of universally applicable, absolutely bulletproof security practices, organizations should apply flexible security policies that best serve their goals using whatever information technologies are available to them at the time [15].

5 A Case Study: Credit Card Risk Management

Credit card transactions continue to grow in number, taking an ever-larger share of the e-finance system and leading to a higher rate of stolen account numbers and subsequent losses by banks [3]. Large-scale data mining techniques can improve on the sate of the art in commercial practice. Scalable techniques to analyze massive amounts of transaction data that efficiently compute fraud detectors in a timely manner is an important problem, especially for e-finance.

In this section, we present a case study on credit card analysis for demonstrating how to use the model proposed above in an e-finance portal. Data mining for credit card portfolio management decisions is to classify the different cardholder behaviors in terms of their payment to the credit card companies, such as banks and mortgage loan firms. In realty, the common categories of the credit card variables are balance, purchase, payment and cash advance. Some credit card company may consider residence state category and job security as special variables. In the case of FDC (First Data Corporation), there are 38 original variables from the common variables over the past seven months. Then, a set of 65-80 derived variables is internally generated from the 38 variables to perform the precise data mining [20,21,22].

We looked at the behavior of the cardholder - how much they spent each month and on what, how they paid their bills, how often they revolved (did not pay the full amount each month), and other behavioral information. We then used a set of mining methods developed in our group to predict how much money could be expected to be made or lost from any one account [14,16].

Figure 4 shows a framework of the behavior-based credit card portfolio management, corresponding to the three-layer Grid architecture shown in Fig. 2, for fast and effective online customer segmentation, and performing multi-level targeted marketing strategies. From this figure, we can see that the mining process can be divided into the following three phases:

1. Two kinds of datasets, $profile$ and *purchasing*, are deployed on the data-grid, and their relationship is connected by the data-flow. The profile dataset is generated by carefully cleaning the customer dataset. On the other hand, the purchasing dataset is generated from transaction dataset, which indicates the purchase number of a customer to a product in a time period.
2. Two kinds of mining agents, called POM (peculiarity oriented mining) and MVF (targeted market value function), deployed on the mining-grid, are employed to mine in the profile and purchasing datasets, respectively, under the mining-flow management;
3. The mined results by the POM and MVF agents are combined and refined into more general ones (i.e. combining customer's profile and purchasing behaviors) to form *active knowledge*, by meta-learning and reasoning with various related knowledge sources on the knowledge-grid.

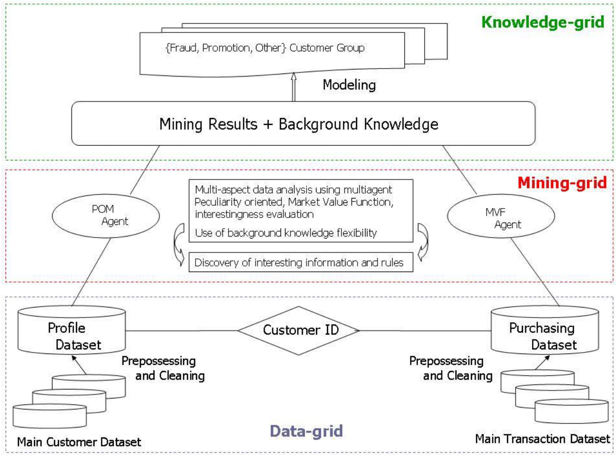


Fig. 4. A framework of the credit card risk management, corresponding to MGCFP shown in Fig. 2

5.1 POM Agent

The basis process is as follows, in the profile dataset, let A_1, A_2, ..., A_m represents the different products shown as columns. Let x_{ij} represents the attributes of customer i to product j, and n is the number of tuples. The peculiarity of x_{ij} can be evaluated by the *Peculiarity Factor*, $PF(x_{ij})$,

$$PF(x_{ij}) = \sum_{k=1}^{n} N(x_{ij}, x_{kj})^{\alpha} \tag{1}$$

where N denotes the conceptual distance, α is a parameter which can be adjusted by a user, and $\alpha = 0.5$ is used as default.

The peculiarity factor is calculated by the conceptual distances, $N(x_{ij}, x_{kj})$, with the following equation,

$$N(x_{ij}, x_{kj}) = |x_{ij} - x_{kj}| \tag{2}$$

The major method used for testing if the peculiar data exist or not (it is called *selection of peculiar data*) after the evaluation for the peculiarity factors is based on a threshold value in Eq. (3),

$$\begin{aligned} threshold \; &= \; mean\ of\ PF(x_{ij}) + \beta \times \\ &\qquad standard\ deviation\ of PF(x_{ij}) \end{aligned} \tag{3}$$

where β can be adjusted by users, and $\beta = 1$ as default. The threshold indicates that a data is a peculiar one if its PF value is much larger than the mean of the PF set. In other words, if $PF(x_{ij})$ is over the threshold value, x_{ij} is a peculiar data. The details about peculiarity oriented mining refer to [29].

5.2 MVF Agent

Targeted marketing involves the identification of customers having potential market value by studying the customers' characteristics and needs, and selects certain customers to promote. Underlying assumption is similar type of customers tend to make similar decisions and to choose similar services or products.

Formally, an information table is a quadruple:

$$S = (U, At, \{V_a | a \in At\}, \{I_a | a \in At\})$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, V_a is a nonempty set of values for $a \in At$, $I_a : U \to V_a$ is an information function for $a \in At$. Each information function I_a is a total function that maps an object of U of exactly one value in V_a. An information table represents all available information and knowledge. Objects are only perceived, observed, or measured by using a finite number of properties.

A market value function is a real-valued function from the universe to the set of real numbers, $r : U \to \Re$. For the targeted marketing problem, a market value function ranks objects according to their potential market values.

A linear market value function is of the form:

$$r(x) = \sum_{a \in At} w_a u_a(I_a(x)) \tag{4}$$

where w_a is the weight of attribute a, and u_a: $V_a \to \Re$ is a utility function defined on V_a for an attribute $a \in At$. x is the one of the elements in U.

Consider an attribute taking its value from V_a. For $v \in V_a$,

$$u_a(v) = \frac{\Pr(v|P)}{\Pr(v)} \tag{5}$$

where $\Pr(v|P)$ denotes the probability distribution of attribute value v in P. $\Pr(v)$ denotes the probability distribution of attribute value v in U.

Consider an attribute taking its value from V_a. For $v \in V_a$,

$$\omega_a = \sum_{v} \Pr(v|P) \log \frac{\Pr(v|P)}{\Pr(v)} \tag{6}$$

where $\Pr(v|P)$ denotes the probability distribution of attribute value v in P. $\Pr(v)$ denotes the probability distribution of attribute value v in U.

For each customer, the purchasing history can be recorded in the transaction dataset and then is transformed to generate the purchasing dataset. Using Eqs. (4) to (6), we can calculate each customer's market value based on each specific product and category. As stated in [28], the MVF agent is effective to sort customers based on some attributes, like, possibility to buy some product. The MVF agent can use not only demographic information, but also the past purchase information of the customers.

5.3 Learning Active Knowledge on the Knowledge-Grid

The mined results from two different mining agents are stored on the knowledge-grid, respectively, and they are combined and refined into more general ones to form *active knowledge* by meta-learning and reasoning. Once we obtain behavioral credit cardholder segmentation, any existing recommendation algorithm for cross-sell and up-sell can be employed for the targeted groups. Our segmentation is based on the propensity to consume. Hence, the products that the customer already owned should be filtered out to avoid seemingly trivial recommendations. In this case, association rules could be employed to find other related products.

6 Related Work

Rapid advances in IT and growing competition are causing fundamental changes in the world's financial services industry. Because the financial sector has a sizable presence in the IT market, these innovations and the subsequent institutional changes will also, in turn, strongly influence future IT development in areas such as distributed computing and Web-based application development [8].

Financial enterprises face significant challenges on both the supply side and demand side, associated in particular with competition, product-service quality and differentiation, transaction security, cost efficiency, and demographic change [18]. Traditional integrated financial enterprises, which exclusively distribute self-developed products via proprietary channels and fulfill all transaction and support services in-house, are no longer adequate for the changing demands of the environment.

The challenge Web-based financial intermediaries face is the transformation to profitability by expanding their relationship through the cross-selling of products and services. A new approach to system architecture is needed that reduces the complexity and costs of coupling information systems as well as increases flexibility to accommodate change.

Real-time integration of disparate data and applications is a key challenge faced by the financial services industry today [19]. The main challenge of Grid computing is the complete integration of heterogeneous computing systems and data resources with the aim of providing a global computing space through the use of standard protocols. Although most of Grid projects have focused on resource sharing in the distributed environment, researchers are beginning to touch about how to employ knowledge processing on the Grid [5].

How to develop an infrastructure focused on the design and implementation of an environment for geographically distributed high-performance knowledge discovery applications is discussed in [4], however, how to combine data mining and knowledge discovery with reasoning and using multiple information sources needs to be investigated at a unified way in depth.

Increasingly, it is becoming necessary to develop any higher level services that can automate the process and provide an adequate level of performance and reliability. Meanwhile, E. Deelman et al. discussed issues associated with workflow management in the Grid in general and provided a description of how to generate executable workflows on the Grid accordingly [7]. Furthermore, Y. Gil et al. used artificial intelligence planning techniques to compose valid end-to-end workflows on the Grid [11].

7 Conclusions

In this paper, we proposed a financial portal model with dynamic multi-level workflows corresponding to a multi-layer Grid architecture, for multi-aspect analysis in distributed, multiple data sources, and for dynamically organizing status-based financial services. It is clear that the value of e-finance technology cannot be fully realized unless we streamline all the interrelated processes and services in the marketplace effectively [9].

E-finance is progressing, with the enhancements of e-paying, e-trading, e-banking, e-commerce, and e-business. We illustrate how to use our model for building an e-finance portal. Our model is particularly appropriate for existing financial institutions with established resources and will help them transform themselves to compete successfully in the new financial environment.

References

1. G. Alonso, F. Casati, H. Kuno and V. Machiraju, Enterprise Application Integration, *Web Services - Concepts, Architectures and Applications*. Springer, 67-92, 2004.
2. J. Blythe, E. Deelman and Y. Gil, "Automatically Composed Workflows for Grid Environments", IEEE Intelligent Systems, 19(4), 16-23, 2004.
3. P. K. Chan, W. Fan, A. L. Prodromidis and S. J. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection", IEEE Intelligent Systems, 14(6), 67-74, 1999.
4. A. Congiusta, A. Pugliese, D. Talia and P. Trunfio, Designing Grid Services for Distributed Knowledge Discovery, *Web Intelligence and Agent Systems: An International Jounal*, 1, 91-104, 2003.
5. V. Curcin, M. Ghanem, Y. Guo, M. Köhler, A. Rowe, J. Syed and P. Wendel, Discovery Net: Towards a Grid of Knowledge Discovery, *Proc. KDD'02*, 658-663, 2002.
6. K. Dandapani, Success and Failure in Web-based Financial Services, *Communications of the ACM*, 47(5), 31-33, 2004.
7. E. Deelman, J. Blythe, Y. Gil and C. Kesselman, Workflow Management in GriPhyN, in J. Nabrzyski et al. (eds.), *Grid Resource Management*. Kluwer, 99-116, 2003.
8. M. Fan, J. Stallaert and A. B. Whinston, "A Web-Based Financial Trading System", IEEE Computer, 32(4), 64-70, 1999.
9. M. Fan, J. Stallaert and A. B. Whinston, The Internet and the Future of Financial Markets, *Communications of the ACM*, 43(11), 82-88, 2000.
10. I. Foster and C. Kesselman, Concepts and Architecture, in I. Foster and C. Kesselman (eds.), *The Grid 2: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, 37-64, 2003.
11. Y. Gil, E. Deelman, J. Blythe, C. Kesselman and H. Tangmunarunkit, Artificial Intelligence and Grids: Workflow Planning and Beyond, *IEEE Intelligent Systems*, 19(1), 26-33, 2004.

12. J. Gordijn and H. Akkermans, "Designing and Evaluating E-Business Models", IEEE Intelligent Systems, 16(4), 11-17, 2001.
13. S. H. Ha and S. C. Park, "Matching Buyers and Suppliers: An Intelligent Dynamic-Exchange Model", IEEE Intelligent Systems, 16(4), 28-40, 2001.
14. J. He, X. Liu, Y. Shi, W. Xu and N. Yan, Classifications of Credit Cardholder Behavior by using Fuzzy Linear Programming, *International Journal of Information Technology and Decision Making*, Vol. 3, 633-650, 2004.
15. C. L. Huntley, "A Developmental View of System Security", IEEE Computer, 39(1), 113-115, 2006.
16. G. Kou, X. Liu, Y. Peng, Y. Shi, M. Wise and W. Xu, Multiple Criteria Linear Programming to Data Mining: Models, Algorithm Designs and Software Developments, *Optimization Methods and Software*, Vol. 18, 453-473, 2003.
17. K. Kumar and J. Hillegersberg, New Architectures for Financial Services, *Communications of the ACM*, 47(5), 27-30, 2004.
18. Z. Liao and M. Cheung, "Challenges to Internet e-banking", Communications of the ACM, 46(12), 248-250, 2003.
19. A. Pan and A. Vina, An Alternative Architecture for Financial Data Integration, *Communications of the ACM*, 47(5), 37-40, 2004.
20. Y. Shi, Y. Peng, G. Kou and Z. Chen, Classifying Credit Card Accounts for Business Intelligence and Decision Making: A Multiple-Criteria Quadratic Programming Approach, *International Journal of Information Technology and Decision Making*, Vol. 4, 581-600, 2005.
21. Y. Shi, Y. Peng, W. Xu and X. Tang, Data Mining via Multiple Criteria Linear Programming: Applications in Credit Card Portfolio Management, *International Journal of Information Technology and Decision Making*, Vol. 1, 131-151, 2002.
22. Y. Shi, M. Wise, M. Luo and Y. Lin, Data Mining in Credit Card Portfolio Management: A Multiple Criteria Decision Making Approach, in M. Koksalan and S. Zionts (eds.), *Advance in Multiple Criteria Decision Making in the New Millennium*, Springer, Berlin, 427-436, 2001.
23. T. Slewe and M. Hoogenboom, Who Will Rob You on the Digital Highway?, *Communications of the ACM*, 47(5), 56-60, 2004.
24. K. Tomita, N. Zhong and H. Yamauchi, Coupling Global Semantic Web with Local Information Sources for Problem Solving, *Proc. SWMR'04*, 66-74, 2004.
25. R. T. Watson, G. M. Zinkhan and L. F. Pitt, Integrated Internet Marketing, *Communications of the ACM*, 43(6), 97-102, 2000.
26. N. Zhong, Developing Intelligent Portals by Using WI Technologies, in J. P. Li et al. (eds.), *Wavelet Analysis and Its Applications, and Active Media Technology.* World Scientific, 2, 555-567, 2004.
27. N. Zhong, J. L. Wu and Ch. N. Liu, Building a Data Mining Grid for Multiple Human Brain Data Analysis, *Proc. KGGI'03*, 36-45, 2003.
28. N. Zhong, Y. Y. Yao, Ch. N. Liu, Ch. X. Ou and J. J. Huang, Data Mining for Targeted Marketing, in N. Zhong and J. M. Liu (eds.), *Intelligent Technologies for Information Analysis.* Springer Monograph, 109-131, 2004.
29. N. Zhong, Y. Y. Yao and M. Ohshima, Peculiarity Oriented Multi-Database Mining, *IEEE TKDE*, 15(4), 952-960, 2003.
30. The OGSA-DAI project: *http://www.ogsadai.org.uk/.*

Knowledge Discovery from Click Stream Data and Effective Site Management

Katsutoshi Yada and Kosuke Ohno

Faculty of Commerce, Kansai University,
3-3-35 Yamate, Suita, 564-8680 Osaka, Japan
{yada,da20104}@ipcku.kansai-u.ac.jp
http://www2.ipcku.kansai-u.ac.jo/~yada/index-e.htm

Abstract. The aim of this paper is to discuss the development of a system for the discovery of valuable new knowledge and to create effective sales strategies based on that knowledge by using massive amounts of click stream data generated by site visitors. This paper discusses and clarifies the process as to how detailed consumer behavior patterns are extracted from click stream data of Internet mall retail site and how such patterns can be used as a source of new ideas for creating new marketing strategies. We will also discuss our successful use of an improved version of the genome analysis system called E-BONSAI to extract and analyze special character strings related to site visitor behavior indicated by distinctive click patterns.

Keywords: click stream data, site management, Internet mall, marketing strategy.

1 Introduction

With the spread of the use of the Internet, many consumers have come to enjoy shopping at virtual retail outlets. The share of total consumer economic activities accounted for by the Internet has been increasing every year and competition among Internet virtual site retail outlets has been growing more and more severe. The number of consumers accessing these sites has been rising and is being watched closely. However, in recent years, with competition so fierce, it has become necessary for individual sites to develop and use more effective site designs. Click stream data provides detailed information on consumer behavior related to accessing a given site and many companies are employing this data as a basic source of information for creating effective site designs and improving overall site management.

In the case of most sites, a simple analysis of this data is being carried out. However, there are very few sites using such data to discover new knowledge and to develop new, more effective sales strategies. This is mainly because these data are so massive in size and complicated in nature that it is not possible for most site analysts to make effective use of them. The basic aim of this paper is to clarify how these enormous amounts of click stream data can be used to

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 360–373, 2007.

discover new knowledge concerning the behavior of site users and also to clarify the process by which detailed visitor behavior patterns extracted from Internet mall click stream data can be used as a source of new and helpful ideas for creating new marketing strategies and improving site management.

In this paper, Internet mall outlet click stream data of several different categories of specialty outlets are used. Since this involves massive amounts of data, it is very difficult for site management personnel to use such data effectively. In this paper, by using the open source platform called MUSASHI [3] [5], we demonstrate how it becomes possible to effectively process very large data sets. MUSASHI includes not only a basic pre-processing module and analysis-related functions, but also several data mining algorithms. By using MUSASHI, we were able to process the click stream data effectively and discover new, useful knowledge of the type necessary for generating effective marketing strategies.

It was possible to clarify the unique features of the various Internet mall outlets investigated using the related Internet click stream data and understand the underlying meanings of the various click patterns, i.e. customer behavior. This led to the development of a process for creating new marketing strategies. The paper includes an explanation concerning how the genome character string pattern analysis software system called E-BONSAI [4], with improvements, could be used to analyze customer behavior patterns in detail. As a result,, new business opportunities were discovered.

The organization of this paper is as follows: First, we discuss the pre-processing and basic analysis of all the outlets in the Internet mall investigated. Next, we focus on outlets handling home appliance, analyze the customers in detail and extract important behavior patterns. Lastly, we clarify the implications for the business based on the new knowledge discovered.

2 Explanation of Data and Data Pre-processing

In this research, the used data [6] [7] was access log data related to an Internet virtual mall that included several different types of shops handling a wide variety of items. As shown in Fig. 1, there were seven shops in the mall that handled mainly electrical appliances and related products. Each of these shops is selling its own unique category of basically different products on its own sites. The various site pages use similar formats (structures) for their own product categories (Product Listing Sheets) and product explanation sheets that tend to have similar formats. Fig. 2 shows the typical site layout. The top pages of these shops are all linked directly to the Mall Directory. They list many different categories of product links ("ct") that, in turn, indicate tables of comparative function data for the various product groups ("ls"). For users who need more details of the products they are interested in, there are detailed product explanation pages ("dt") that can be referenced. Consumers buy the products they want by putting them in their "carts" ("kosik").

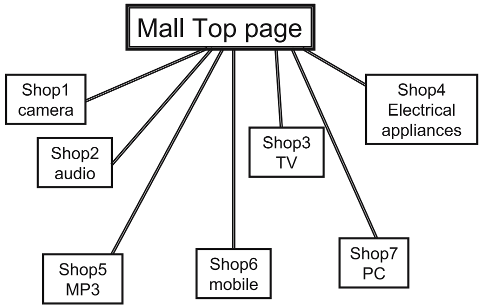


Fig. 1. Mall shops and their respective products

When potential customers visit mall sites, they spend different amounts of time at the various sites they visit and it is possible to track their going from one site to another. However, when these female customers switch from one shop site to another shop site, or when they close their browser and later return to a site, these visits are separate sessions. Therefore, in this study, the session was used as the unit of study of the customers (visitors) and it was possible to identify all specific respondent consumers and track individual visits to sites on separate days and to different shops. In addition, by using the data showing what products the buyer puts into her cart, it was possible to ascertain what each consumer purchased and at what outlet.

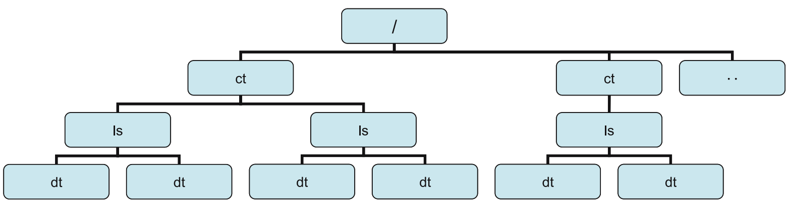


Fig. 2. Typical site layout of a mall shop

A session consists of one click or more. The data also included which pages in the site were clicked and in what order the pages were viewed. It was also possible to estimate the time spent viewing each page by measuring the time between clicks. This procedure of following the route of the clicks made within each site during a given session is called PATH. In addition, the data indicating a click on a given page of information also includes data viewed on the previous page. For example, if the visitor was looking at pages containing explanation of various TV sets, a list of all the previous pages involved is available. However, the visitor may just be giving a quick look at the entire list of what is available or, in reality, she may be carefully comparing the performance features of the different products and maybe, the pages previously viewed indicate the products that she has already looked at. Therefore, in this study, rather than the data from the previous pages, we gave precedence to the data for the page clicked just before the visitor clicked a given session page.

Next, the subset of data from the total data for analysis will be explained. The total data required very long sessions. However, about half of these sessions (49%) needed only one click. A single click does not constitute enough information to use it for analysis of consumer behavior. In addition, our objective was to obtain useful information for managing sites that can be obtained from the basic consumer behavior that is indicated from the PATH data of the visitors using the site. Thus, for the same reasons, PATH data including only a limited number of clicks was also not suitable for analysis.

In this paper, purchase was defined to have occurred when the customer, during a session, put the product in her cart. For sessions where purchase took place (i.e., included a click for the cart), the average number of clicks was 19. Sessions where purchase took place and which were 2-4 clicks or less, accounted for only 1% of total purchase sessions. In contrast, there were sessions that included very large numbers of clicks (for example, 15,000 clicks in a single session).

It is very difficult to analyze sessions that include such large numbers of clicks. Sessions that included 100 or more clicks comprised less than 1% of total sessions and thus, it follows that their overall importance is not much. Therefore, sessions with 100 or more clicks were excluded as targets of analysis for the purposes of this paper.

In summary, the sessions that were used for the analysis were those with more than 5 clicks and under 100 clicks. The sessions analyzed consisted of 144,327 sessions (That included 4,220 visitors who made purchases and 140,107 who did not).

3 Basic Analysis of All the Shops in the Projects

Based on past experience in analyzing consumer data, it is known that it is necessary, when new data is to be analyzed, to carry out the basic analysis several times before it is possible to understand the trends in the data and the parts of the findings that have application for business purposes. Thus, as a first step, the log data for all the shops was analyzed and the main features of this data were determined.

3.1 Purchase Behavior on the Sites

Fig. 3: The x-axis is the length of the session and the y-axis is the number of visitors (sessions) in that time frame category. It is clear from the data that the number of visitors who stay on a given site for two minutes or less is very large. Fig. 4 indicates these factors for visitors making purchases. The major sub-group of these purchasing visitors was in the 5-10 minute sector and overall, visits tended to be long. On occasions when purchases were made, it can be

surmised that products were compared or purchasing visitors spent time completing the procedures for making their purchases. The shortest route when purchase was involved (the number of clicks used until the cart appeared during a given session) was about 7 clicks. These purchasers were completing their purchases in 1-2 minutes.

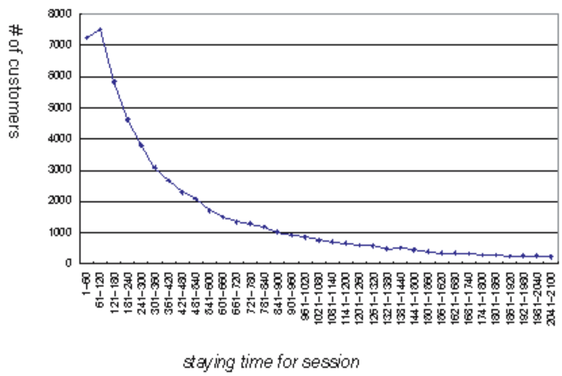


Fig. 3. Distribution of length of time spent on sites

The number of customers visiting these sites tended to be higher on weekdays. On weekends (Fridays, Saturdays and Sundays), visits were relatively limited. However, purchase probability (number of purchasing customers/ number of total site visitors) was almost the same for both weekdays and weekends. For session length, by comparison, although the number of visitors differed by time of the day (morning, noon, evening, late night), there was virtually no difference in purchase probabilities by time of the day.

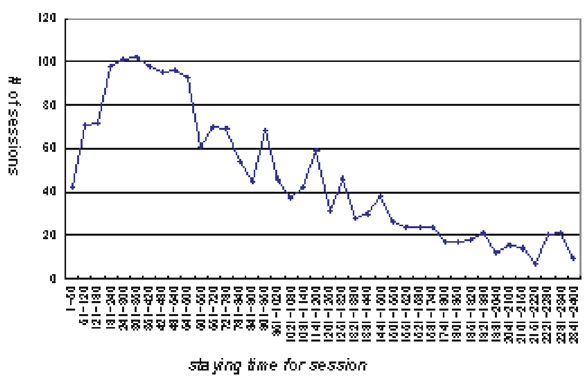


Fig. 4. Length of time spent on site by purchasing visitors

Table 1. Basic data for the various shops

shop	shop code	main product	# of session	avg. clicks/session	avg. stay/session
shop 1	10	camera	645	29.1	33
shop 2	11	audio	409	29.7	30
shop 3	12	TV	544	32.1	35
shop 4	14	electrical appliances	1624	31.4	29
shop 5	15	MP3	517	28.6	29
shop 6	16	mobile	259	22.4	18
shop 7	17	PC	222	26	32

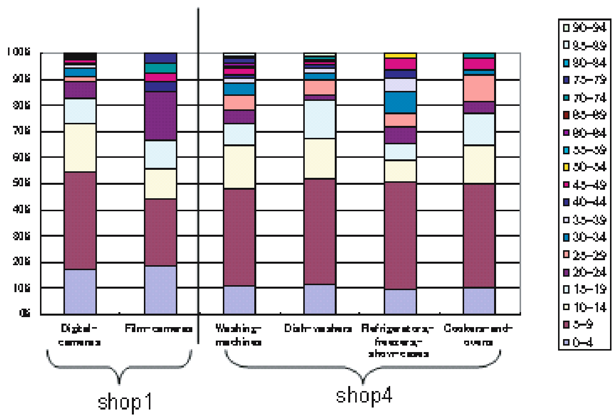


Fig. 5. Differences in number of clicks by product category

When the sites visited and categories of items they handled were used to examine visitor behavior, comparatively significant differences in visitor behavior were found. Table 1 indicates such basic data as number of sessions, average clicks per session, and the average length of time spent for one session for the various shop sites. The number of sessions for Shop #4 that handles electrical appliances, was the highest compared with all the other shops and the average number of clicks per session was also high. By comparison, Shop #6 that handles mobiles (phones, PDA's, etc.) and Shop #7 that handles PCs have lower number of sessions (visitors). However, Shop #6 sessions are short, but Shop #7 sessions are long. Like the case of off-highway/utility engine vehicles, the tendency of Net visitors to explore is not similar and tends to vary so that there are probably differences shop to shop. Fig. 5 indicates the differences in the number of clicks by product category and the distribution of these numbers. There are many customers who buy such items as digital camera and film-camera with a very low number of clicks. However, visitors who buy washing machines, dish-washers, refrigerators and ovens use a relatively high number of clicks. As can be seen from the above, depending on the product category, consumer behavior on the Net tends to vary.

3.2 Purchase Probability and Number of Sessions

At this point, the special characteristics of the various shops will be examined in terms of purchase probability and number of sessions. The probability that visitors will purchase something at the various sites they visit varies widely as seen from fig. 6. Compared with the other shops, shop #5 that handles MP3 products (digital audio) has a higher score for purchase probability. In fact, the average number of clicks used by visitors to this shop until the cart is reached is clearly lower than for the product categories.

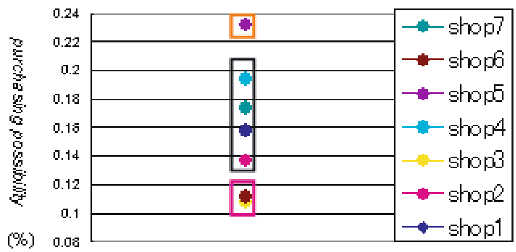


Fig. 6. Purchase probability by site

Therefore, in the case of this site, it is possible to infer that rmany of the prospects to this site have already decided what they intend to purchase before they visit the shop. By comparison, in the case of shop #3 that handles TVs and shop #6 that handles mobiles (mobile equipment items), the purchase probability is very low. Especially, in the case of shop #3, although the average number of clicks was the highest, the purchase probability of this shop was the lowest of all the seven shops investigated. The prospects who come to this site appear to have high interest in collecting information, but they seldom purchase and it seems they probably make their purchases in regular shops.

Next, we examine the sites by number of visitors. Fig. 7 shows the number of sessions for each site, including those that went to the cart stage (purchase) by product category. Shop #4, handling electric appliances, was accessed by the highest number of visitors. In contrast, shop #6, handling mobiles, and shop #6, handling PCs were accessed by very few visitors. Considering this situation in terms of visitor analysis, the sites that have the greatest number of prospects are the easiest to analyze and the sites that have the lowest number of visitors are the most difficult to analyze because of methodology problems.

Fig. 8 has been used to position the various sites in terms of number of sessions by product category. Here we discuss sales strategy for the shop group indicated in Fig. 8. In the case of Shop #5, since this shop already has a very high purchase probability, it is unlikely that it will be possible to anticipate an increase in sales from simply altering the site content. However, since the number of visitors to the shop is not sufficient, as a shop strategy, it is possible to suggest that the shop should focus directly on increasing the number of visitors to the site.

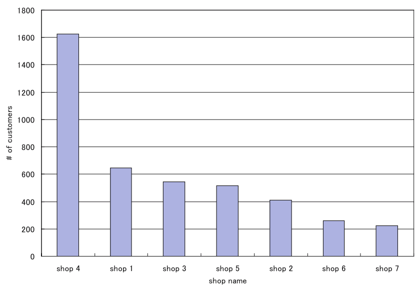


Fig. 7. Numbers of sessions by product category

This means that the shop should use banners and other types of advertising to lure visitors to the site. The shops with the second lowest purchase probability level are shop #3 and shop #6. Regarding these shops, it is possible that because of the special nature of the products handled, consumers are looking at the actual products and then purchasing them. In this kind of a situation, a strategy that involves a joint effort to bring customers into the real shop, or a joint effort of the "Click and Mortar" strategy type (real shop and Net mall shop cooperation) can be recommended.

From our point of view, recommending strategies for shops 1, 2, and 7 are the most difficult. Concerning analysis, using only the data provided, the number of visitor respondents is too few for proper analysis. Thus, strategies for these shops will have to wait for another opportunity in the future.

In the sections that follow, we will focus on shop #4. The reason for this is that shop #4 has a purchase probability level somewhere in the middle so that some work on the site content can probably be anticipated to bring about some improvement in sales. Since the number of visitors is large, it is relatively easier to carry out an effective analysis and the effectiveness of such analysis will probably be significant.

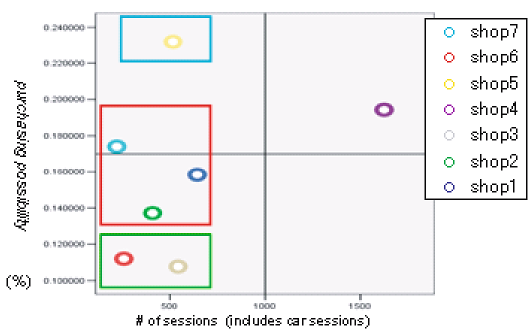


Fig. 8. Positioning of the shop sites

4 A Site That Can Bring in Wavering Potential Customers and Make Them Buy

In this section, shop #4, that is handling electrical appliances, is the focus. The intent is to use a detailed analysis of the log data to discover useful knowledge for creating a new shop strategy.

4.1 Defining the Target Customers

For the purposes of marketing, there is no strategy that will work for every single customer. Therefore, customer segmentation is carried out and it is necessary to find a strategy that takes into consideration all the segments. To this end, we chose to focus on the buying motives of the visitors to the site. For the customers who had already decided on the product they wanted, it was felt that changing the content of the site would not be very effective. In our view, it was important to focus on the customers who were wondering which product to buy and the customers who were wondering whether they should buy a product in category or not. It is important to pay attention to the off-highway/utility engine customers.

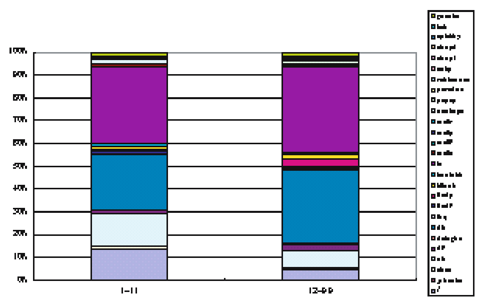


Fig. 9. The difference in the page content of the two groups of customers

The reason for this is that by putting more effort into access of the site, it can be anticipated that we can influence some of them to buy. Fig. 9 indicates customers who go to the cart on the 7th click and the other customers who go to the cart after more clicks and what type of pages they were looking at. As can be seen from the Figure, the customers who go to the cart after 11 clicks or less tend to look at the list of all the items (ls). As against this, those who went to the cart after 12 strokes or more, tended to go to the detailed product explanations (dt). The former looked at the total product list and tended to go straight to the cart. It can be assumed that the others tended to compare the various products and then go to the cart. We defined these customers as customers who were wavering concerning purchase and designated them as a segment that required analysis.

Our objective was to discover some type of new knowledge that would contribute to raising sales further and to achieve this we carried out analysis of the log data to see if could find a new strategy that would stimulate sales. At this point, for the analysis of the objective variables, for shop #4, among the visitors to the site that used 12 clicks or more, we separated and defined the visitors who bought something and those who made no purchases. Expressing this in another way, we extracted the characteristics from among the customers who were wavering whether to purchase a product or not and of the customers who did not make a purchase, i.e., to discover new knowledge. There were 657 customers who purchased a product and 2000 who did not.

4.2 Extracting the Features of the Decision Tree

In order to extract the special characteristics of the customer segment groups, we used a decision tree. The tool we used was the SPSS AnswerTree V3.1 [9] and we processed the explanatory variables from the log data. We used the average time spent on each page between clicks, the number of clicks made on each page and the purchased products. However, we removed the page data that had no direct connection to the purchase, the member log-in page (klient), and the cautions related to purchase (onakupu), etc. Regarding the products purchased, all products other than the top six products (washing machines, refrigerators-freezers, built-in-ovens, built-in-dish-washers, dish-washers, cookers-and-ovens) were classified as "Other".

We analyzed all the data available plus the data for each purchased product. This analysis showed that there was a similar tendency applying to all the products. Fig. 10 shows the decision tree for customers who read the page on refrigerators-freezers. There were 166 customers who looked at the material for the refrigerators-freezers and made a purchase. However, 346 customers who looked at the materials did not make a purchase. For customers who looked at product explanations (dt) page for periods of 11 seconds or less, almost none made purchases. Those who did tended to spend more time reading the product explanations. Besides, the customers who purchased looked at the product sheet (ls) for 2 seconds or so and there was a tendency for purchasers to make a list of the products they hoped to purchase.

The final rule that emerged was that customers who looked at the top page (A) 1-second or more had a high tendency to purchase. There appeared to be cases where visitors spend relatively long periods of time looking at the top page perhaps because they were not used to the site and could not find the products they wanted to buy; or perhaps because they could not find the category, or perhaps they had not decided in advance on the category of product they wanted, they stayed on the top page for long periods of time. However, using only the available data, it was not possible to pursue the matter further.

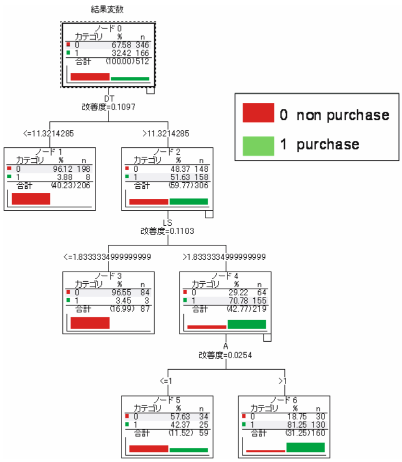


Fig. 10. Refrigerators-freezer purchase decision tree

4.3 Tracking Analysis Using the Character String Analysis System, E-BONSAI

The last stage of the analysis was PATH analysis carried out using the character string analysis system, E-BONSAI. E-BONSAI was originally developed to analyze DNA code. Since then, E-BONSAI has been improved and by expressing consumer behavior patterns as character strings, it can be used for extracting patterns from time-series category patterns as a data mining tool [2] [4] [5]. In the log data referred to in this paper, using the mapping table (as seen in Fig. 11(b)) variable conversions, characters from the internal site pages can be converted into different characters and the click PATH data (the data they referred to) for all visitors who were part of the project can be converted into character strings. For example, if a given visitor clicks on the top page, the category (ct), a (product) list (ls), and a product explanation, (dt) back to a list (ls), the other explanation (dt) and the popup (popup), her PATH would be expressed as the character string, "CLDLDp". In the way that E-BONSAI extracts DNA data concerning the cause of sickness in an ill patient, it is possible to extract special character strings.

When the data for the visitors who had purchased Refrigerators-Freezers referred to in the previous section was analyzed using E-BONSAI, results resembling those shown in Fig. reffig:12 were generated. The rules are complicated and difficult to interpret. Stating the overall tendencies roughly, in the data for the visitor examination of the materials on the site pages, amid the character strings, the strings: "1*7*7*" and "1*5*5*5*5*7*" were found. When these strings were present, it appeared that purchases were frequent. The "1" in this result is a "pop-up" and the "7" refers to a functional specification of a given product. When the key term (findp) was used for searchingA it was found that either "5" in the product list (ls) and product explanation (dt), one or the other was involved.

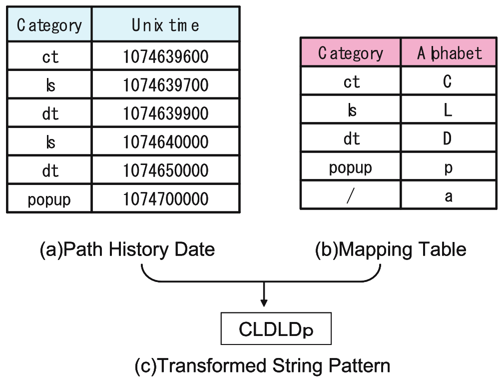


Fig. 11. Conversion list for internal site pages

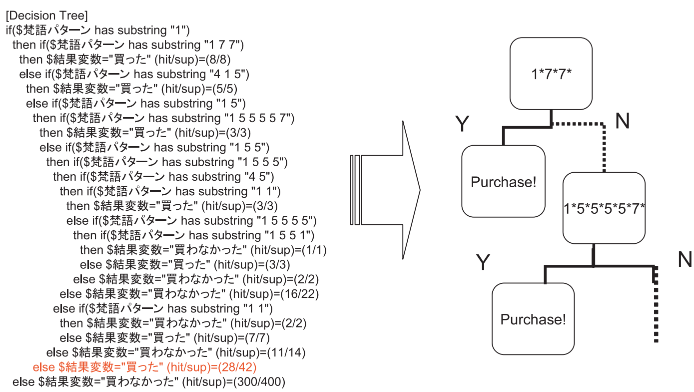


Fig. 12. Rules established by E-BONSAI

In other words, customers who purchased Refrigerators-freezers, after they clicked on the ”pop-up” tended to key in either a function or a specification and perform searches for product that had such features. In addition, it is clear that they clicked on the product list and product explanations. We were unable to obtain detailed information concerning the ”pop-up” so that it is not possible to interpret accurately what was happening, but it seems that these purchasing visitors were doing multiple searches and reading the detailed information.

4.4 Implications for Business

Here, we would like to indicate some of the implications for business related to the rules described above. Visitors who intended to purchase made a list of the products of interest to them and slowly read the detailed information concerning these products, probably as a matter of course. The factors that we paid special attention to were the multiple searches they made by keying in terms concerning the functions and specifications. There are two possible reasons for this behavior: The first is that the page design was bad so that it may have been hard to find an explanation without carrying out several searches in order to list the products they were interested in. Therefore, there may probably be a need to improve the design of the search function page.

A second possible cause was that the visitor was wavering whether he/she wanted to purchase one of the products or not and searched for many similar products. These visitors do need an easier-to-use site or a high-performance search function. They need a basis or system for making the decision whether they want to buy one of the products or not. It appears that they do not know which one of the products would actually fit their needs. Therefore, they are probably searching for many different functions and specifications, and giving themselves enough time to make up their mind.

We would like to suggest that word-of-mouth [1] data functions be made an additional function for these sites. In Japan, consumers tend to be heavily influenced by word-of-mouth sources and the power of this factor is growing. If previous purchasers could be asked about their actual experiences with various similar products in the past, and post-purchase data could be used, this would undoubtedly help establish a basis for judgment and decision-making for purchase.

Of course, there will be some negative opinions concerning product performance. However, if the shop sites are willing to provide potential customers with that kind of information, this would lead to a higher level of trust on the part of customers. If some shops are already using this type of function, it would appear that this function is not working in an integrated way with the search function. We would like to suggest that the use of an integrated form of word-of-mouth product information data and the search function might be an effective tool for building a strategy for maintaining good customer relations.

5 Conclusions

In this paper, by analyzing mall log data, it was possible to clarify the special characteristics of the visitors who accessed the shop sites and, in addition, by analyzing the visitor PATH data in detail, it was also possible to gain new knowledge. However, since there was not enough data available, comprehensive rules could not be generated and therefore, we were hard put to determine the implications for business operations. Moreover, since the unit of analysis was the session, it was not possible to get an understanding of consumer behavior on the Net in the form of time sequence data and the scale of the analysis was severely limited. If possible, we would like to have the opportunity to use data that can be identified as related to a given respondent and to be able to use such data in sufficient quantities in order to gain new and effective knowledge.

References

1. Brown, T. J., Barry, T. E., Dacin, P. A., Gunst, R. F.: Spreading the Word: Investigating Antecedents of Consumers' Positive Word-of-Mouth Intentions and Behaviors in a Retailing Context. Journal of the Academy of Marketing Science, Vol.33 No.2 (2005) 123–138.
2. Hamuro, Y., Katoh, N., Ip, E. H., Cheung, S. L. Yada, K.: Combining Information Fusion with String Pattern Analysis: A New Method for Predicting Future Purchase Behavior, V. Torra(ed.), Information Fusion in Data Mining, Studies in Fuzziness and Soft Computing, Vol.123 (2003) 161–187.
3. Hamuro, Y., Katoh, N., Yada, K., Washio, T.: MUSASHI: System for Knowledge Discovery in Large Business Data. Journal of the Japanese Society for Artificial Intelligence, Vol.20, No.1, (2005) 59–66. (in Japanese)
4. Hamuro, Y., Kawata, H., Katoh, N., Yada, K.: A Machine Learning Algorithm for Analyzing String Patterns Helps to Discover Simple and Interpretable Business Rules from Purchase History. Progress in Discovery Science, LNAI 2281 (2002) 565–575.
5. Katoh, N., Yada, K., Hamuro, Y.: Business Application for Sales Transaction Data by Using Genome Analysis Technology. Discovery Science, LNAI 2843 (2003) 208–219.
6. Naito, D., Yamamoto, K., Yada, K., Matsumura, N., Ohno, K., Tamura, H.: Does WEB Log Data Reveal Consumer Behavior? Proc. of the Workshop "Discovery Challenge" in conjunction with ECML/PKDD 2005 (2005) 43–54.
7. Yada, K., Hamuro, Y., Katoh, N., Fusamoto, I., Fujishima, D., Ikeda, T., Washio, T.: Data Mining Oriented CRM Systems Based on MUSASHI: C-MUSASHI. Active Mining, LNCS 3430 (2005) 152–173.
8. http://lisp.vse.cz/challenge/ecmlpkdd2005/chall2005.htm
9. http://www.spss.com/answertree/

Sampling-Based Stream Mining for Network Risk Management

Kenichi Yoshida

Graduate School of Business Science, University of Tsukuba
Otsuka 3-29-1, Bunkyo, Tokyo 112-0012, Japan
yoshida@gssm.otsuka.tsukuba.ac.jp

Abstract. Network security is an important issue in maintaining the Internet as an important social infrastructure. Especially, finding excessive consumption of network bandwidth caused by P2P mass flow, finding internet viruses, and finding DDoS attacks are important security issues. Although stream mining techniques seem to be promising techniques for network security, extensive network flow prevents the simple application of such techniques. Since conventional methods require non-realistic memory resources, a mining technique which works well using limited memory is required. This paper proposes a sampling-based mining method to achieve network security. By analyzing the characteristics of the proposed method with real Internet backbone flow data, we show the advantages of the proposed method, i.e. less memory consumption.

1 Introduction

Network security is an important issue in maintaining the Internet as an important social infrastructure. Especially, finding excessive consumption of network bandwidth caused by P2P mass flow, finding internet virus, and finding DDoS (Distributed Denial of Service) attacks are important security issues.

Figure 1 shows how we can use stream mining techniques to keep network secure. If few users use P2P software and consume huge amounts of network bandwidth, excessive consumption of network bandwidth increases the potential for network hazards. Thus, finding such excessive P2P traffic is an important task of network operators in order to protect internet security. To hide their existence from network operators, current P2P software randomly changes its port number. The random generation of port numbers makes network operators' task difficult. However, P2P software has to keep the same destination IP address and destination port number so that their companion starts communication. The combination of the same destination IP address and destination port number is an important index for their companions to start communication with them. Even if P2P software can generate their destination port number randomly, they have to use the same combination of the destination port number and destination IP address after their initializations. The stream mining technique seems to be a promising approach to find the combination of randomly generated destination port numbers and destination IP addresses.

We can also find computers infected with internet viruses by using stream mining techniques. Although new internet viruses are found everyday, they use combinations

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 374–386, 2007.

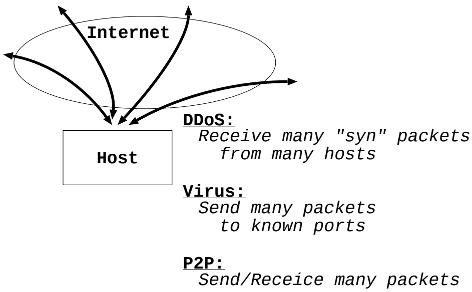


Fig. 1. Characteristics of IP packets

of already known security holes. Even if the combinations are new, the security holes that are used are limited. Thus computers infected by internet viruses will send a lot of known packets. We can find such computers by mining combinations of same source IP addresses and destination port numbers. Finding a DDoS attack is also a straight application of stream mining. Since the victim of DDoS receives a lot of “syn” packets from various sources, we can find such DDoS attacks by mining the frequent combinations of destination IP addresses and “syn” flags.

All these network security tasks try to find frequently appearing combinations of IP addresses, port numbers and IP flags in the packet header. We can use stream mining techniques to perform these tasks. Thus stream mining techniques seem to be promising for keeping networks secure. However, vast network flow prevents the simple application of such techniques. A mining technique that works well with limited memory is required. Actually, recent implementation of frequent mining requires non-realistic memory resources (See Section 3 for details).

In this paper, we propose a sampling-based mining method to obtain network security. By analyzing the characteristics of the proposed method with real Internet backbone flow data, we show the advantages of the proposed method, i.e. less memory consumption.

Section 2 of this paper briefly surveys related work. Section 3 analyzes the characteristics of network data in order to clarify the limitations of conventional data mining techniques, and clarifies the motivation for this research. Section 4 explains our approach, and Section 5 reports on the experimental results. Section 6 shows how we can use the proposed method to find Internet Viruses. Finally, Section 7 concludes our findings.

2 Related Work

The monitoring of internet traffic is an extensively studied area, e.g. [1,2,3,4,5]. IETF’s IPPM working group proposes a framework of IP performance metrics [4]. Their work is important in providing a baseline to compare the measured results by standardizing the attributes to be measured. Surveyor [5] is a project to create measurement infrastructure. NLANR [3] has a project to develop a large-scale data collection system as the base infrastructure for various data analysis. MAWI [2] also provides the archive of real internet traffic data. CAIDA [1] is making various tools to analyze network data. Their visualization tools cover various analyses of network data.

Analysis of measured data is also studied [6,7]. Some studies, e.g. [8], try to use data mining techniques to automate analysis. When considering the importance of data mining performance, various methods for frequent item finding have been proposed. Among them, Coarse counting [9], Sticky sampling, Lossy Counting [10], hash-based approaches [11,12], and the use of group tests [13] are important methods. These methods can quantify frequently appearing items without any omissions. However, most of them treat frequent items. Their performances on frequent itemsets are not examined comprehensively.

Frequent itemset mining implementations repository [14] provides various implementations of frequent itemset finding algorithms and test data sets. Among them, the best implementation award was granted to LCM-v2[15] in the FIMI'04 workshop. Although LCM-v2 and other implementations in the repository show excellent performance on the data set in the repository, we found their performance on the network data to be poor. The characteristics of network data causes new problems to which data mining communities have not paid attention.

The contribution of this paper is the use of real internet traffic data to protect network security. Although the conventional data mining technique requires too much memory resources, our proposed method can analyze internet traffic data with much less memory resources. We also show the performance of the proposed method on the standard test data to clarify the characteristics of the proposed method.

3 Requirements on Stream Mining

In this section, we analyze the characteristics of network data in order to clarify the limitations of conventional data mining techniques and clarify the motivation for this research. Three important characteristics of network data are:

- Zipfian-like distribution
- Large amounts of data
- Short transactions

Figure 2 shows the frequency of IP flows. An IP flow is a stream of data transfer between two host computers. Thus IP packets belong to the same IP flow share same combination of 1) source IP address, 2) destination IP address, 3) source port number, and 4) destination port numbers. As shown in Figure 2, the distribution of IP flows roughly follows the Zipfian distribution. Here, the data is obtained from the MAWI repository [2]. Dots show the real data. The line shows the simulated Zipfian distribution whose trend is fitted from the real data.

Note that the distribution is not pure Zipfian. Most frequent part (See top 1000 in Figure 2) has a slightly loose trend. Less frequent part (bellow 1000) has slightly tighter trend.

This small difference on distribution trends makes the finding of P2P traffic far more difficult. Figure 3 shows the cumulative frequency of the same data. Dots show the real data. The line shows the simulated Zipfian distribution. If data follows the pure

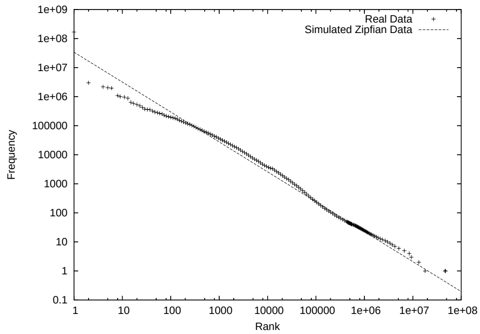


Fig. 2. Distribution of Internet flow data

Zipfian distribution, top 100 IP flows roughly consumes 50% of network bandwidth (Solid line in Figure 3). However, in the real data, the same bandwidth is filled by, roughly speaking, the top 10000 IP flows (Dots in Figure 3).

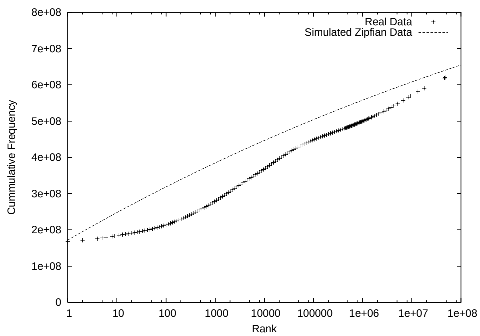


Fig. 3. Cumulative Distribution of Internet flow data

The analysis of major network resource consumption is important to find P2P traffic. If we use conventional frequent itemset implementations, this analysis requires huge memory resources. Figure 4 shows one such example. Here we use LCM-v2 [15], which won the best implementation award in FIMI'04 workshop. LCM-v2 are used to find IP flows which has more than 5000 IP packets in the data shown in Figure 2. This roughly corresponds to the analysis of top 10000 IP flows.

Since LCM-v2 is an off-line program, data of the first 10M, 20M 30M and 40M packets are used in different experiments. And memory consumption is observed. Since we used a personal computer with Pentium 4 2.4GHz CPU and 2G byte main memory, experiments with 40M packets which use 1,878M byte main memory were the maximum setting we could perform with LCM-v2.

The maximum size of data set, i.e. 40M packets, is unacceptable from a practical point of view. Table 1 shows how many packets a single 1 Gbps internet line can transfer. Although 1 Gbps line is considered as an access line, it can transfer 1M packets

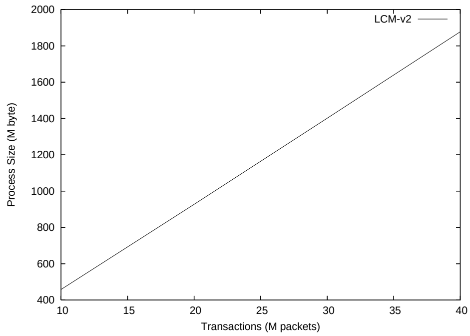


Fig. 4. Memory Resource Requirement

per second. For the network management tasks, at least 1 hour of data should be analyzed. However, if the memory consumption is in proportion to the amount of packets as shown in Figure 4, 200G byte of main memory is required. Needless to say, 200Gbyte of main memory is unacceptable as a resource for access line analysis. To analyze 10G or more bps backbone network, Tera byte class main memory is necessary. It is common that operators have to analyze weekly-based bandwidth consumption of backbone network. The conventional frequent itemset implementations seem to require Peta class memory for such analysis. In other words, the use of conventional frequent itemset implementations is not practical to keep networks secure.

Table 1. Amount of Network Packets

	Number of Packets	Kinds of Packets
1 second	1M	1K
1 minute	60M	4M
1 hour	4G	210M
1 day	86G	4G
1 week	605G	35G

The standard way to solve the memory resource problem is by "sampling." By reducing the total amount of data to be analyzed, sampling can decrease the requirements on the memory resources. However, sampling alone is not sufficient. For example, if a backbone network operator has to find P2P users who constantly use P2P software for more than weeks, the operator has to analyze more than 10^{12} packets. Since the practical limitation of LCM-v2 or similar implementation is 10^7 packets, the sampling rate becomes 10^{-5}. This sampling rate of 10^{-5} is too radical. Since 10^5 packets can transfer about 140Mbyte data, the use of a sampling rate 10^{-5} ignores the data transfer of less than 140M byte. Although this sampling rate might find really huge P2P traffic, large parts of P2P traffic are simply ignored in the sampling phase. We need some method that can ease this problem.

Although the first two characteristics, i.e. Zipfian-like distribution and huge data amount, make the problem more difficult, the last characteristics, i.e. short transactions, make the problem easier. Each IP packet has variety of information in the IP header and forms a transaction. Here, IP addresses, port numbers, and various IP flags form items. Although various IP flags exist, the transactions extracted from IP header are shorter than those of standard data such as retail data (See Section 5). In the next section, we will explain an attempt to make best use of these characteristics.

4 Sampling-Based Stream Mining

Since the analysis of all network data requires too much memory resources, we have to investigate sampling based methods to reduce the resource requirements. During the investigation, the task of keeping the sampling rate reasonable was the key of our study. As discussed in the previous section, naive combinations of sampling and conventional frequent itemset mining methods are not practical. In this section, we first explain a frequent "item" mining algorithm which works well with small memory resources. Next we expand it to the frequent "itemset" mining algorithm.

Figure 5 shows the basic idea of our frequent "item" mining algorithm. It simply performs sampling and counts the frequency of sampled items. A fixed size cache was used to store the frequency. If the size of cache is large enough, this algorithm only samples data and counts the frequency of the data. Although the basic idea/algorithm is extremely simple, we found this algorithm works well with a specific memory management strategy.

```
Basic Algorithm
begin
    while (input Item) do
        if (sample)
            i = index of Item in heap;
            increment heap_cnt[i] by 1;
            if (heap_cnt[i]>thresh_hold)
                Mark as Frequent;
    done
end
```

Fig. 5. Basic Algorithm

Figure 6 shows the memory management strategy we used. It first calculates "n" hash values of a given item. "n" hash functions are used for this purpose. Next it generates "n" indexes from "n" hash values. If the item to be stored is already in the cache, one of the indexes refers the entry for the item. If the item is a new one, then hash2 selects the index that refers to the least frequent entry out of "n" entries referred by "n" indexes. Then the old item stored in the cache at the index will be replaced by the new item.

Figure 7 shows how we can apply the basic idea to find frequent "N itemsets." From each sampled transaction, it first selects top N+E frequent "items." Then it removes the items which do not appear more than F from the selected items. After it removes items from the selected items, it generates combinations of items and stores the generated itemsets. Here, E and F are the tuning parameters.

```
Memory Management Function
Input
    Item: Data to be stored in Cache
Variable
    Hash[]: Table of Hash Values
    Idx[]: Table of Cache Index
begin
    Calculate "n" hash values from Item
            and store them into Hash[]
    Idx[] = Hash[] % Cache Size
    if (one of entry refereed by Idx[] stores Item)
    then return Idx that refers the entry
    else  Select Idx that refers least frequent entry
          heap_cnt[Idx] = 0
          return Idx
end
```

Fig. 6. Pseudo code of Memory Management

```
Algorithm for Itemsets
begin
   while (input Transaction) do
      if (sample)
         Items = Select Top N+E frequent items
                  from Transaction
         Remove items which appear less than F
                  from Items
         Itemset[] = Generate Combination of Items;
         for each Itemset do
            i = index of Itemset in heap;
            increment heap_cnt[i] by 1;
            if (heap_cnt[i]>thresh_hold)
                  Mark as Frequent;
      done
   done
end
```

Fig. 7. Algorithm for Itemsets

Figure 8 shows the data structure for this algorithm. Note that the proposed method only requires the memory resources for the counter and the input data itself.[1] Although the conventional frequent itemset mining methods require various auxiliary data structures, such as an FP-tree, the proposed methods do not use any auxiliary data. In the conventional implementations, various hash tables are used to store the positions of various data. However, the proposed method does not require such hash tables by directly re-calculating the position of the data (See Figure 6). This feature also reduces the memory requirement of the proposed method.

[1] To be exact, the data structure shown in Figure 8 does not store the data itself. It stores the hash value of data, and our implementation uses different data structure to translate hash values to the original data. However, this implementation tends to require less memory resources than directory storing the data itself.

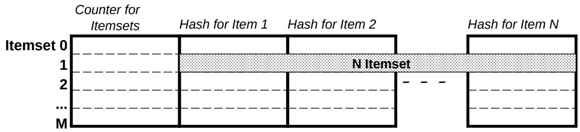


Fig. 8. Data structure

As we can see above, our algorithm is simple. But in the next section, we show that this simple idea works well on network security issues. Other important characteristics of the proposed method are:

– The proposed method does not guarantee to find all the frequent itemsets. It has the possibility to miss the frequent itemsets when the cache size is too small. However, if large amounts of data require sampling, sampling-based methods cannot be exempted from this defect. Their results become inaccurate because of the missing data in the sampling phase. However, the mechanism of memory management shown in Figure 6 statistically selects frequent data in general. In our experiments, the characteristics of network data seem to help this selection.
– The proposed method is a single pass algorithm that is suitable for on-line analysis. Most of the conventional frequent itemset mining algorithms require multiple scans of the data. And the multiple scans of data reduces the on-line performance of such methods. Sliding window-based on-line methods, such as [16], ease this problem by restricting the multiple scans on the window. However, the proposed method is simpler and is suited to on-line analysis.

5 Experimental Results

In this section, we analyze an IP header log with our proposed method and conventional methods. The IP header log is obtained from MAWI repository [2]. It was recorded on a monitoring point of a Internet backbone on 2005/Jan/07, and is a 24-hour-long trace on 100Mbps trans-Pacific line. The number of packets recorded in the log is about 620M. However we only use first 40M packets because of the limitation of conventional implementations (See later).

Here we analyze the performance of each method in finding P2P traffic. More precisely, the required result of each method is the frequently appearing combinations of source IP addresses, destination IP addresses, source port numbers, and destination port numbers. The combinations that appear more than 5,000 times are marked as frequent.

We selected LCM-v2 [15] and an implementation of apriori [17] as the conventional methods. We selected apriori as the most standard implementation. We also select LCM-v2 since it was selected as the best implementation in the FIMI'04 workshop. The default parameters for these implementations are used in the experiments. For the proposed sampling-based method, the cache of 100,000 entries was used. We used 4 as "n" for the memory management parameter. Although we did not extensively seek the best "n," 4 tended to make reasonably good results in various experiments. We set E as 4 and F as 100.

Figure 9 compares the memory resources used by each method. Since LCM-v2 and apriori are off-line programs, data of first 10M, 20M 30M and 40M packets are used in the different experiments. Memory consumption was also observed. Since we used a personal computer with Pentium 4 2.4GHz CPU and 2G byte main memory, experiments with 40M packets which use 1,878M byte main memory was the maximum setting we could perform with LCM-v2. Apriori requires slightly less memory resources. However both methods require over 1G byte of memory when the number of packets exceeds 30M. In contrast, the proposed sampling-based method requires only 5.8M byte main memory. Since the main memory size for the proposed sampling-based method is defined by the cache-size, 5.8M byte was the maximum that the proposed sampling method required. Thus the proposed sampling-based method has a firm advantage of memory resources.

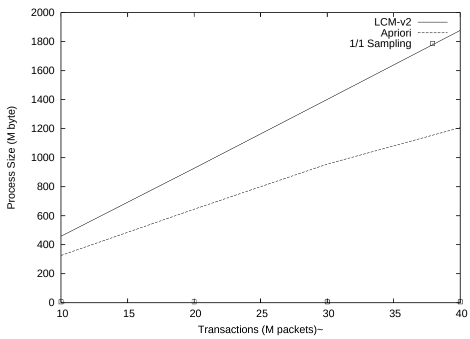


Fig. 9. Memory Resource Comparison

In the experiments shown in Figure 9, we used a sample rate of 1/1. In other words, without using sampling technique, all the data was analyzed by each method in Figure 9. Figure 10 compares how other sampling rates change the accuracy of the analysis. In the experiments shown in Figure 10, we used sampled data. When we used a sampling rate of 1/S and tried to find data that appeared more than F times, each method tried was found to appear more than F/S times of that in sampled IP header log. This sampling causes two types of errors. Accidentally, data that appears more than F times in the original log may not appear F/S times in the sampled log. In contrast, data which does not appears F times in the log may appear more than F/S times in the sampled log. Figure 10 shows the sum of both errors.

Since sampling has same effect on LCM-v2 and apriori, we just compared the error of LCM-v2 and our sampling-based method (See Figure 10 solid line and dot respectively). We also concentrate on the results of IP flow. In other words, we only compare the results on frequently appearing combinations of source IP addresses, destination IP addresses, source port numbers, and destination port numbers. Since the IP flow frequency is the basic information for network management, the accuracy on the IP flow frequency is important from the practical view points.

Figure 10 shows the percentage of error which each method has with various sampling rates. Although there exists a small difference, LCM-v2 and our sampling-based

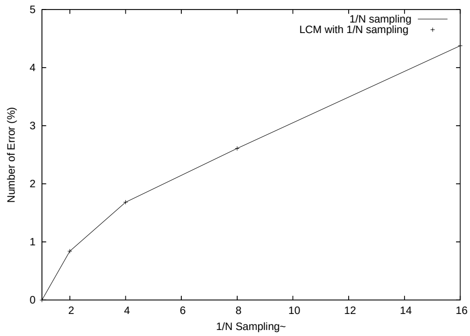


Fig. 10. Sampling Errors

method have the same trends. Error rates of both methods increase when the sampling rate S increases. Note that, the proposed sampling-based method does not guarantee to find all the frequent itemsets. However, the difference of LCM-v2 and the proposed method is negligible. In fact, we cannot see any difference through Figure 10. The change of sampling rate causes far larger effects.

Figure 11 compare the CPU time of LCM-v2, apriori, and our proposed method. Among them, LCM-v2 is fastest. Sampling-based method with sample rate 1/1 is next, and apriori is the slowest. Although LCM-v2 is the fastest implementation, its CPU time required to process 40M packets, i.e. 86 second, is not fast enough for on-line processing. Since 1Gbps access line transfers 1M packets per second, at least 1/2 sampling is necessary for LCM-v2, from the view point of CPU time.

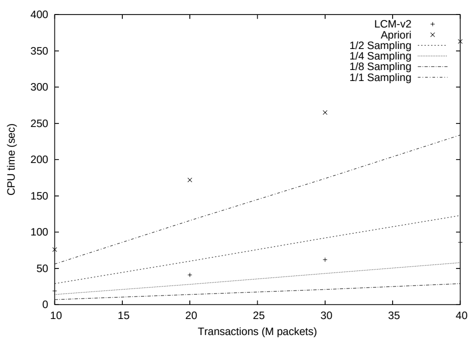


Fig. 11. CPU Time Comparison

From the view point of CPU time, the proposed method requires the sampling rate 1/6 for the on-line analysis. Under this condition, its error rate is estimated to be 2.5% (See Figure 10) and keeps acceptable accuracy from the practical viewpoints. Among 3 methods described in this section, the proposed sampling-based method is the only method that can realize the real time analysis of network data. The conventional methods seem to require too much memory resources.

6 Discussion

The previous section describes how we can use the proposed sampling-based method to find P2P traffic from network data with experimental results. This section shows how we can use the same method to find Internet Viruses.

To protect privacy, the traffic data in the MAWI repository are partly scrambled. Although other information is retained, both source IP addresses and destination IP addresses are changed so that we cannot get the original IP addresses. Since each original IP address is converted into the same scrambled IP address, this scrambling does not affect the attempt for the virus detection. Here we try to find virus infected hosts by mining frequently appearing combinations of destination port numbers and source IP addresses. Thus if each original IP address is converted into the same IP address, we can find the scrambled source IP address of a suspicious host that might be infected by the internet virus. However, we cannot confirm our suspicion by actually checking the suspicious hosts. Thus the discussion in this section is the first step toward achieving internet security.

Although further research is necessary, Table 2 shows promising results. It shows top frequent itemsets of length 2. To make this table, whole MAWI repository of 620M packets were analyzed by the proposed sampling-based method with a sampling rate 1/1. Since we can start various investigations with these results, we found these results very promising for network security.

Table 2. Top Frequent Itemsets of Length 2

No.	Frequency	Item.1	Item.2	Comments
1	174907672	d80	S129.254.151.98	Mass Flow
2	170162834	d80	D24.138.66.93	
3	170162834	S129.254.151.98	D24.138.66.93	
4	168195425	d80	s32865	
5	168195172	S129.254.151.98	s32865	
6	168195172	D24.138.66.93	s32865	
7	31856129	d80	D148.120.46.107	WWW Server
8	16536914	s53	S194.157.88.115	Name Server
9	8426763	s80	S195.95.0.116	WWW Server
10	5753676	d80	S194.84.83.198	Crawler
...				
755	238234	d1433	S214.146.52.17	

For example, the first 6 itemsets are fragments of the same mass flow (i.e. an itemset of length 4: source IP address: 129.254.151.98, destination IP address 24.138.66.93, source port number 32865, and destination port number 80). Here "d" means the destination port number. "s" means the source port number. "D" means the destination IP address. And "S" means the source IP address. Although destination port number 80 is the reserved port number of a WWW service, this flow is strange for normal WWW traffic. It is too large for normal WWW traffic. Although the pattern is an up stream flow of WWW, a down stream flow that normal WWW traffic tends to have, does not exist in the log. We suspect that this flow is not regular WWW traffic. It seems to be a data transfer that is using non-standard techniques.

The data of No. 7 seems to be an up stream flow of a WWW server. Since there are a variety of IP addresses that receive down stream WWW data from host 148.120.46.107, we believe host 148.120.46.107 to be a normal WWW server. The data of No. 10 shows the typical pattern of a WWW crawler. And more importantly, the results include a typical footprint of port scanning by an internet virus. The data of No. 755 is one such example. Host 214.146.52.17 seems to search Microsoft SQL servers and scans variety of destination hosts with port 1433. Since host 214.146.52.17 does not receive any response, it might be infected by an internet virus. Or it might be a host with incorrect installation. Both cases mean there is an important security incident.

Note that network management tasks are important in protecting internet security. Network operators have to check what is happening in their network. Above analysis of network usage makes up an important part of network management tasks, and contributes to the finding of security incidents such as the spreading of viruses. Frequent itemset mining techniques can help such analysis by checking various combination of IP header information, such as destination port number and source IP address (Table 2, No 1 and 10), source port number and source IP address (Table 2, No 8 and 9), and etc.

7 Conclusion

Network security is an important issue in maintaining the Internet as an important social infrastructure. Especially, finding excessive consumption of network bandwidth caused by P2P mass flow, finding internet viruses, and finding DDoS attacks are important security issues.

Although stream mining techniques seem to be promising techniques for network security, vast network flows prevent the simple application of such techniques. Since conventional methods require non-realistic memory resources, a mining technique that works well with limited memory is required. In this paper, we propose a sampling-based mining method to achieve internet security.

By analyzing the behavior of the proposed sampling-based method and conventional frequent itemset mining methods with real Internet backbone flow data and other conventional test data, we show:

– The proposed sampling-based method is the only method that can realize the real time analysis of network data.
– The naive combinations of sampling and conventional frequent itemset mining methods require too much memory resources.

References

1. http://www.caida.org/
2. http://www.wide.ad.jp/wg/mawi/
3. http://moat.nlanr.net/
4. Rfc2330, framework for ip performance metrics
5. http://www.advanced.org/surveyor/
6. Mirkovic, J., Prier, G., Reiher, P.L.: Attacking ddos at the source. In: Proc. of the 10th IEEE International Conference on Network Protocols. (2002) 312–321

7. Ohsita, Y., Ata, S., Murata, M., Murase, T.: Detecting distributed denial-of-service attacks by analyzing tcp syn packets statistically. In: Proc. of IEEE Globecom 2004. (2004)
8. Demaine, E.D., Lopez-Ortiz, A., Munro, J.I.: Frequency estimation of internet packet streams with limited space. In: Proc. of the 10th Annual European Symposium on Algorithms. (2002)
9. Fang, M., Shivakumar, N., Garcia-Molina, H., Motwani, R., Ullman, J.D.: Computing iceberg queries efficiently. In: Proc. 24th Int. Conf. Very Large Data Bases, VLDB. (1998) 299–310
10. Manku, G., Motwani, R.: Approximate frequency counts over data streams. In: Proc. of the 28th International Conference on Very Large Data Bases, Hong Kong, China. (2002) 346–357
11. Charikar, M., Chen, K., Farach-Colton, M.: Finding frequent items in data streams (2002)
12. Jin, C., Qian, W., Sha, C., Yu, J.X., Zhou, A.: Dynamically maintaining frequent items over a data stream. In: Proc. of the twelfth international conference on Information and knowledge management. (2003) 287–294
13. Cormode, G., Muthukrishnan, S.: What's hot and what's not: Tracking frequent items dynamically. In: Proc. of Principles of Database Systems. (2003) 296–306
14. http://fimi.cs.helsinki.fi/
15. http://sunsite.informatik.rwth-aachen.de/publications/ceur-ws//vol-126/
16. Golab, L., DeHaan, D., Demaine, E., Lopez-Ortiz, A., Munro, J.I.: Identifying frequent items in sliding windows over on-line packet streams. In: Proc. of ACM SIGCOMM Internet Measurement Conference, Miami, USA (2003)
17. http://fuzzy.cs.uni-magdeburg.de/ borgelt/apriori.html

Relation Between Abductive and Inductive Types of Nursing Risk Management

Akinori Abe, Hiromi Itoh Ozaku, Noriaki Kuwahara, and Kiyoshi Kogure

ATR Media Information Science Laboratories
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288 Japan
{ave, romi, kuwahara, kogure}@atr.jp

Abstract. In this paper, we contrast inductive nursing risk management and abductive nursing risk management, point out the importance of the abductive type, and suggest cooperation between them. In general risk management, inductive management is usually adopted. If we computationally conduct inductive management, it is vital to collect a considerable number of examples to perform machine learning. For nursing risk management, risk management experts usually perform manual learning to produce textbooks. In the Accident or Incident Report Database home page, we can review various types of accidents or incidents. However, since reports are written by various nurses, the granularity and quality of reports are not sufficient for machine learning. We, therefore, explain the importance of conducting dynamic nursing risk management that can be achieved by abduction, then illustrate cooperation between abductive and inductive types of nursing risk management.

1 Introduction

Recently, the number of medical accidents due to nursing activities has increased. Expectations about the safety and quality of nursing care have, therefore, increased, and the range of nursing responsibilities has expanded. Accordingly, some hospitals have introduced medical risk management systems to reduce medical accidents or incidents. In general, risk management is based on a statistical model that analyzes accident and incident reports to generate frequently occurring patterns of accidents and incidents. That is, in general risk management, inductive management is usually adopted. If we computationally conduct inductive management, we need to collect as many examples as possible to obtain sufficient machine learned results. In the Accident and Incident Report Database home page [11], it is possible to review various types of accidents and incidents. However, since reports are written by various nurses, the granularity and quality of the reports are not suitable to a computational analysis. However, since human experts add detailed comments to the reports that contain their knowledge on the accidents or incidents, they can be used as supplemental information for analysis of nursing accidents or incidents. In fact, because the reports include only one example for one situation, it is rather difficult to conduct data mining.

T. Washio et al. (Eds.): JSAI 2006, LNAI 4384, pp. 387–400, 2007.

Therefore, we are now collecting data of nursing activities in the E-nightingale project [14].

If we collect sufficient data set of nursing activities, we can mine the data to generate general models for nursing activities and accidents. However, static risk management has certain limitations such that it cannot deal with new or rear situations, thus we think it would better to conduct dynamic nursing risk management. Computationally speaking, this can be achieved by abduction, because risk management means risk prediction. For computational prediction, abduction is the best selection, and for abduction, we need to model nursing activities or human behaviour for errors. Based on Vincent's model [26], we proposed abduction-based nursing risk management [4] and extended the model to a scenario violation model [6] that has notion of "time." These models are quite flexible ones for conducting risk management, but for better management we need to prepare other models built from actual data on nursing activities. This is because risk management should be conducted according to actual situations. Thus both types of risk management have pros and cons.

In this paper, we contrast the features of inductive and abductive styles of nursing risk management and highlight the importance of the abductive (dynamic) type. In addition, we point out the necessity of cooperation between both styles of management. Section 2 illustrates inductive risk management, then Section 3 explains abductive risk management. In Section 4 we discuss the relationship between abductive and inductive risk management, and conclude in Section 5.

2 Inductive Nursing Risk Management

2.1 Necessity of Risk Management and Data Mining in Hospitals

Despite recent high-quality nursing education and advanced medical treatments, the number of medical accidents due to nursing activities has not decreased. Rather, expectations about the safety and quality of nursing care have increased, and the range of nursing responsibilities has expanded. Therefore, to lower medical accidents, it is important to reduce the number and severity of nursing accidents and incidents to benefit both hospitals and patients. Accordingly, some hospitals have introduced a medical risk management section to analyze medical accidents, including nursing accidents. Analyzing nursing accident and incident reports is vital for reducing or eliminating the same or similar accidents or incidents in the future. Since the main purpose is to extract general rules from many examples, such risk management is based on inductive risk management. Human induction generates useful and precise risk management rules, but it is expensive in terms of time and effort for human experts to read many accident and incident reports to generate general rules.

In response to the above problem, computational approaches were recently proposed to analyze the huge number of nursing accident and incident reports [15]. Such methods usually use data such as selections from a list or a formatted input such as date and time. Of course accident and incident reports also include free-description information, but if those methods want to analyze it, they need to apply, for instance, text mining techniques. As a result, the extracted rules are rather typical tendencies or rules that have already been pointed out by human experts. In addition, these tendencies appear to be quite superficial. This is because they are only based on experiments to determine the effectiveness of data mining in analysis of accident and incident reports. In addition, those methods only analyzed selective input sections in the reports. However, since it is rather difficult to obtain complex relationships only from such simple data, it is necessary to analyze free-description information to obtain sufficient rules.

2.2 Features of Accident and Incident Report Databases

The published accident and incident reports can be freely reviewed. They are stored in the Accident and Incident Report Database home page [11], which contains information on various types of nursing accidents and incidents. The report are typically organized as follows:

- Title
- Abstract
  This contains brief information about nurses and patients etc. but is removed.
- Detailed information about an accident or incident
- Reasons for the accident or incident
- Improvements to be considered or to be done
- Experts' comments

It would be reasonable to use the detailed information included in the "detailed information about an accident or incident." However, the problem is that a typical example seems to be selected for a case. As a result, it is rather difficult to perform machine learning or data mining from the data. Moreover, since reports are written by various nurses, the granularity and quality of each report are quite different, thus they are not suitable for conducting machine learning or data mining. Currently, then, the best way is to manually analyze the accident or incident reports by using our background knowledge to obtain general models for nursing accidents or incidents. One advantage of using the accident and incident reports is that we can refer to experts' comments to build general models of nursing accidents or incidents. Those models sometimes provide the missing information from nurses' descriptions, though currently it can be rather difficult to automatically analyze such freely described comments. Through a brief review of the report, we previously pointed out that two of the most serious reasons for accidents or incidents are nurses' (unintentionally) being convinced they are correct even when they are not, and miscommunication among nurses [4]. We currently use those errors as assumptions for causes of nursing accidents or incidents to build risk management models. It seems to work well but we still need to obtain the other data to conduct data mining to automatically generate general models for nursing accidents or incidents.

2.3 Data Collection Through Experiments in Hospitals

It would be best to obtain particular knowledge about nursing activities from actual data obtained in actual working places. Kuwahara et al. proposed an integrated nursing activity monitoring system that couples ubiquitous apparatus with fixed apparatus [14] (E-nightingale Project). The system was proposed to monitor all of the nursing activities in order to make reports on their entire workflow in real-time. In addition, the aim of the system is to give just-in-time advice to nurses during their jobs. In the current (experimental) system, nurses carry wearable sensors that record their (self-)conversations or dialogues, the number of footsteps, body angle, etc. In addition, video monitoring systems are placed in each room to visually record part of their activities (no sound), while radio-frequency (RF) stations are installed to record their locations. We think we can extract particular knowledge (workflow etc.) on nursing activities if we conduct statistical analysis or data mining of the obtained data. In fact, in [16], Naya et al. analyzed the acceleration data of body parts by Fisher Linear Discriminant Analysis. Although individual tasks are clearly classified, we need to label all the possible tasks by reviewing video monitors, we can determine individual tasks by analyzing pattern data (symbol grounding problem). However, temporal analysis seems to be rather difficult. Therefore, it is rather difficult to obtain the workflow if we simply adopt this analysis method to the image data. A certain method such as time division would be needed.

To avoid the symbol grounding problem, it would be better to use dialogue and conversation data by nurses where nurses speak the actual names of their activities. Currently, we are trying to build a set of nursing corpora [20] and extract nursing workflow patterns by analyzing transcribed[1] nursing dialogues [21,22]. We have collected around 1000 hours of dialogue and conversation data and transcribed around 40 hours of data from them. In addition, we are manually adding the tags of nursing jobs to the transcribed nursing dialogues (Table 1 (Private information is modified.)). Types of tags are determined by referring to authorized job categories provided in Classification of Nursing Practice [12] and in Nursing Practice Classification Table [13]. They include such labels as "conference (18-106)" and "intravenous infusion (13-63-6A0502)."

Since we have not finished tagging, we did not apply any data mining method to the above tagged data. However, previously we did apply C4.5 [24] (Fig. 1) and Keygraph [17] (Fig. 2) to dialogue and conversation data without any tags to obtain particular working models of nurses. Names such as "Meeting" which are different from the above tags were actually spoken by nurses during their activities, which are parsed by using Chasen [10]. For C4.5 we only used time information and job information, while for Keygraph we used all the words parsed from nursing dialogues, part of which have been manually corrected. In fact, we used quite simple data (granularity of nursing activity seems rather rough) for analysis, but could still obtain results that can be regarded as a general

[1] We cannot automatically extract dialogue and conversation data from voice data, because we do not have sufficient dictionaries to perform speech recognition of voice data including nursing specialized terminologies.

Table 1. Example of labeled dialogues by nurses

Time	dialogue	Job Category
11:01:00	I'm going to a short conference (meeting or handover).	18-106 conference
11:20:48	The short conference is finished.	18-106 conference
11:28:11	I'm going to prepare a drip infusion for Abe-san.	13-63-6A0502 intravenous infusion
11:32:01	I have finished preparing the drip for Abe-san.	13-63-6A0502 intravenous infusion

workflow model of nursing activities. It is likely that we could have obtained a more detailed workflow model if we had used more complex data. Actually in [7], we presented the possibility of using nursing practice manuals provided in hospitals to generate nursing workflow that can be regarded as a set of scenarios[2]. Since the nursing practice manuals contain detailed flow of individual nursing activities, we can use the flow as supplement of roughly categorized nursing activities. In addition, we will be able to extract concrete model by analyzing nursing practice manuals and generate abstract model of nursing activities by applying text mining techniques.

```
HourP = 7L: Clinical record
            checking
HourP = 8F: Meeting
HourP = 8L: Env. management
HourP = 9F: Env. management
HourP = 9L: Taking temperature
HourP = 10F: Medicine Division
HourP = 10L: Medicine Division
HourP = 11F: Meeting
HourP = 11L: Break
HourP = 12F: Break
HourP = 12L: Break
HourP = 13F: CAN input
HourP = 13L: Meeting
HourP = 14F: Hair-bath
HourP = 14L: Talking
HourP = 15F: Meeting
HourP = 15L: Meeting
HourP = 16F: Clinical record writing
HourP = 16L: Clinical record writing
HourP = 17F: Conference room cleaning
HourP = 17L: Talking
```

Fig. 1. Analysis of nursing activity by C4.5

In the above, we can only generate general patterns of nursing activities. In fact, our aim is to produce risk management models, but a problem is that currently we do not have any information on accidents or incidents linked to the obtained data. Those data can be offered by hospitals, but for several reasons, we have not obtained any data yet. Once we have prepared sufficient data set, it will be possible to devise particular workflow models of nurses and particular models of nursing accidents or incidents by applying data mining methods. Then after constructing appropriate models, we can conduct inductive nursing risk management.

However, it might be rather difficult to obtain sufficient actual accident or incident report to link the actual nursing activity data for several reasons. It is of course possible to detect accident or incident to conduct automatic or manual

[2] Scenario is defined in 3.2.

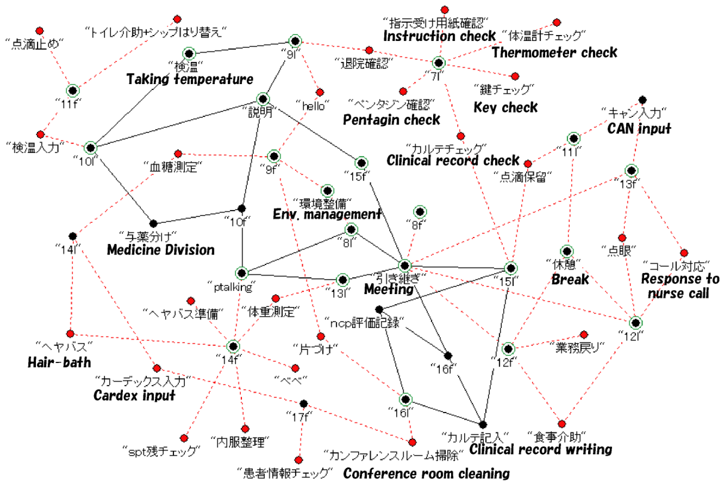


Fig. 2. Analysis of nursing activity by KeyGraph

survey of whole data such as monitored video images, but manual solution is not realistic and automatic one is currently rather difficult. Therefore, in addition to inductive risk management that uses statistic accident or incident knowledge, it is necessary to conduct dynamic risk management that dynamically determine accident or incident. In the next section, we introduce abductive risk management that is a dynamic risk management.

3 Abductive Nursing Risk Management

In previous papers, we pointed out the importance of dealing with possibly hidden (ignored or unconscious) events, factors, environmental elements, personal relationships, or matters likely to cause an unrecognized but serious accident in the future [3,4]. Such factors can be regarded as chances. In [3], we proposed an abductive framework for Chance Discovery [18]. In this section, we model risk management based on abduction. First, we briefly describe a pure abduction-based risk management model, then briefly illustrate a scenario violation model for risk management.

3.1 Abduction Model

In [4], we formalized nursing risk management with an abductive framework. In cases where we know all possible hypotheses and their ideal observations[3], it is possible detect malpractice beforehand because if someone selects the wrong hypothesis set or fails to generate a necessary hypothesis set, an ideal observation cannot be explained. When an ideal observation cannot be explained,

[3] If we use a workflow sheet for nurses or an electronic medical recording system, we can identify ideal observations.

accidents or incidents will occur. This mechanism enables us to logically determine the place where accidents or incidents are likely to occur before they occur. A simple logical framework to complete an activity is shown below (by using the framework of Theorist [23]):
If

$$F \cup h_1 \not\models ideal\_observation, \tag{1}$$

then find h_2 satisfying (2) and (3).

$$F \cup h_2 \models ideal\_observation \tag{2}$$

$$F \cup h_2 \not\models \Box, \tag{3}$$

$$h_1, h_2 \in H, \tag{4}$$

where F is a set of facts that are always consistent, and h_1 and h_2 are hypotheses that are not always consistent with the set of facts and other hypothesis sets. Hypotheses are generated (selected) from hypothesis base H; $\Box$ is an empty set. Therefore, the last formula means that F and h_2 are consistent.

If we can complete formula (2), the activity is successfully completed. On the other hand, if we cannot generate enough hypothesis sets to complete formula (2), particular problems will disturb the completion of the activity. Thus, we can determine beforehand the possibility of risk by abduction. That is, when we cannot explain an *ideal_observation* with a current hypothesis set, a certain error might occur. If objective (*ideal_observation*) cannot be explained, it cannot be completed. Such a situation is caused by particular accidents or incidents.

This is a very simple logical model that does not consider any effects of the order of hypothesis generation.

3.2 Scenario Violation Model

In a pure abduction model, it is not possible to deal with time information. For time information, Brusoni proposed a theoretical approach for temporal abduction [9], that shows abduction with absolute time information. However, we do not need to prepare strict models for temporal projection, rather, we need a simple strategy to express a series of events. For that, we introduced a scenario in abduction and proposed a scenario violation model [6]. As shown in [19], a scenario is a *time series* of events under a coherent context. Accordingly, by introducing a scenario, we can deal with time information in abduction. Thus, we introduced the effects of the order of hypothesis generation to the abduction model. In the scenario violation model, scenario violation means the possibility of error. This is a formalization of nursing risk management considering a series of events (time).

A simple logical model for checking a scenario violation is illustrated below. When all of the candidate scenarios are in a scenario base (SB), risk determination inference can be achieved as follows:

$$s_i \in SB \tag{5}$$

$$s_i = \sum_{j(in \quad chronological \ order)} e_{ij}, \tag{6}$$

where s_i is a scenario and e_{ij} is an event.

As shown above, to avoid accidents or incidents (by completing an activity), it is necessary to reach a final goal, and when it is achieved, we observe a particular result. Accordingly, we can set an observation as a result from the final goal. Our aim, therefore, is to explain the observation with sets of scenarios. Thus when no accident or incident occurs, the following formulae are completed:

$$F \cup \sum_{i(in \ chronological \ order)} O_i \models O, \tag{7}$$

$$F \cup s_i \models O_i, \tag{8}$$

where F is background knowledge, and O_i and O are observations (results of nursing activities). O_i can be regarded as a sub-observation of O. Of course, in some cases we do not need to consider sub-observations.

Formulae (7) and (8) show abduction (hypothetical reasoning) that determines whether a scenario is completed. The most important difference from the hypothetical reasoning model is that formula (7) requires verification of the chronological order of scenarios (hypotheses).

When

$$F \cup s_j \not\models O'_j, \tag{9}$$

$$F \cup s_j \models O_j, \tag{10}$$

$$O_j \neq O'_j, \tag{11}$$

and

$$F \cup \sum_{i(in \ chronological \ order)} O_i \not\models O, \tag{12}$$

certain scenarios appear to be violated, indicating the possibility of an error. The possibility of accidents or incidents occurring can logically be determined (explained) by abduction before they occur.

In this formalization, a scenario can be regarded as a structured and ordered hypothesis. In addition, each event can also be regarded as an ordered hypothesis. In [7], we extended the scenario violation model to control a scenario or event order.

4 Relation Between Abductive and Inductive Types of Risk Management

In the previous sections, we briefly illustrated inductive and abductive types of nursing risk management. They can be briefly contrasted as follows: Inductive risk management is a static style of management, whereas the abductive style is a dynamic type.

In this section, we compare both types of risk management and propose a way for them to cooperate for better risk management.

4.1 Features of Both Types of Nursing Risk Management

Inductive nursing risk management has been widely adopted in many hospitals. However, computational inductive nursing risk management is still in the experimental stage. The main reason is that we cannot prepare sufficient data for analysis. For instance, if accident and incident reports are used, that have been written by several nurses they are not well organized, and quality and granularity are not sufficient for analysis. In addition, because we must deal with natural language, it is quite difficult to computationally analyze the data as it is. We must, therefore, prepare sufficient background knowledge as risk management experts have. Previously, to generate nursing activity scenarios, we adopted C4.5 and Keygraph to the obtained nurses' dialogue and conversation data, manually parsing the transcribed data for analysis. As shown in the previous section, the results are not sufficient for conducting nursing risk management, but we did notice certain possibilities for generating models for nursing risk management.

Another problem of inductive risk management is that it is a static form of management. With static management, it is very difficult or impossible to deal with unknown or less known situations, because static risk management does not consider seldomly occurring situations. For frequently occurring accidents or incidents, however, static models work well. In those cases, static models are taught to nurses via textbooks and the nurses learn them to avoid frequently occurring accidents or incidents. However, for seldomly occurring accidents or incidents, if nurses only follow the static models, they cannot recognize such novel or rare events as triggers to accidents or incidents. Thus they tend to ignore such novel or rare events.

On the other hand, abductive nursing risk management does not seem to be widely accepted. However, it is a form of dynamic risk management, which can deal with unknown or lesser known situations, since even unknown or lesser known situations can be logically explained. For dynamic risk management, we prepare nursing activity models to logically determine the possibility of accidents or incidents occurring, with the idea being that when an ideal observation cannot be logically explained by the current hypotheses (environmental situation, event, task etc.), accidents or incidents will occur. That is, it is not necessary to prepare for all possible accidents or incidents. What we do need to prepare are workflow models of nursing activities, which are regarded as scenarios; in fact, if we follow hypothetical reasoning, we must prepare all the possible facts (models) and hypotheses. Some of the hypotheses can be obtained from the environment or an electronic medical recording system, though it is possible to adopt abduction such as CMS [25] and Abductive Analogical Reasoning (AAR) [1], which can generate missing hypotheses. Of course, as pointed out in [8], in the case of nursing risk management newly generated hypothesis cannot always be applied. In fact, we cannot adopt optional medicine as a hypothesis: we need additional mechanisms to guarantee such generated hypotheses. One solution was proposed in [8]. Thus for abductive risk management, particular models are necessary, and we are now trying to generate nursing scenarios from nursing practice manuals provided by hospitals and nursing academies, and transcribed dialogues and conversations among nurses [7,21]. We have not yet fully succeeded in obtaining general models of nursing activities, though, we think we will be able to obtain individual nursing scenarios after analyzing of those data. For instance, we can refer to a nursing practice manual to obtain the following procedure:

[Gas analysis of arterial blood (from A-line)]

1) After preparing necessary equipment, visit the patient.
2) Confirm that the blood collection part of the stop cocks of A-line is OFF.
3) Remove the cap of the stop cocks and disinfect it with a cotton-wool swab containing Hibiten alcohol.
4) Connect disposable injector (5 ml) to the stop cocks, and aspirate heparin saline into a route to fill it with arterial blood.
5) ....

Then, we can generate a (concrete) scenario on the gas analysis of arterial blood as follows:

e_1 = Prepare necessary equipment → e_2 = Visit the patient → e_3 = Check blood collection part of the stop cocks of A-line → e_4 = Remove cap of the stop cocks → e_5 = Disinfect the stop cocks with a cotton-wool swab with Hibiten alcohol → e_6 = Connect disposable injector (5 ml) to the stop cocks → e_7 = Aspirate heparin saline into a route to fill it with arterial blood → .....

Since abductive risk management logically infers the possibilities of occurring accidents or incidents, it can control the issuing of alarms to nurses, sounding alarms just before when nurses should be warned, something that inductive risk management cannot do. Inductive risk management only sounds alarms to nurses whenever the situation matches cases of well known accidents or incidents. Accordingly, there exists a problem that nurses tend to be accustomed to alarms.

We can thus compare the features of abductive and inductive risk management.

◇ Abductive risk management
  - dynamically determines accident or incident
  - can deal with rare or novel situations
  - needs sufficient models
  - collects or generates examples

◇ Inducitve risk management
  - statically determimes accident or incident
  - can only deal with frequently occurring situations
  - can generate models
  - needs many examples

4.2 Cooperation Between Both Types of Nursing Risk Management

It was mentioned earlier that inductive nursing risk management has been widely accepted and adopted, because it is rather easy for nurses to understand. Furthermore, since accidents or incidents covered by the inductive case are frequently occur, inductive risk management is useful in actual situations. Since its mechanism is rather concrete, after learning inductive models, it is rather simple to apply the knowledge to actual situations in order to reduce nursing accidents or incidents. A problem is that the provided models cannot be applied to other situations; especially, we cannot apply the models to unknown or lesser known situations. However, under normal conditions, since the same or similar accidents or incidents tend to occur, inductive nursing risk management would be sufficient.

On the other hand, abductive nursing risk management is rather abstract. In addition, the management is conducted using logical inferences, so it is rather difficult to understand. In addition, we need particular models to make inferences for risk management. However, after obtaining sufficient models, since it infers the possibility of accidents or incidents occurring according to particular situations, flexible risk management can deal with unknown or lesser known situations.

Thus both types of risk management are quite different in their procedures. From the viewpoint of understandability, inductive risk management is better, while from the perspective of effective and flexible risk management, abductive risk management is superior.

For effective risk management, then, it would be preferable to employ both risk management, while they cooperate with each other. For inductive risk management, we need to prepare as many examples as possible, which can be achieved by collecting data from several experiments in hospitals and results from abductive nursing risk management studies. On the other hand, for abductive risk management, we need to prepare sufficient nursing activity models. Those models cannot be easily obtained, but if we collect a considerable number of examples, it will be possible to build models. These models can be built through inductive risk management. In addition, after executing abductive nursing risk management, we can also collect missing examples to inductive risk management. In [2], we compared the features of inductive and abductive hypothesis and proposed integration of abduction and induction. A similar procedure to that proposed in [2] can be performed as shown below.

1) to generate propositional hypotheses by *abduction* under similar observations, and
2) to generate predicate rules by *induction* from abduced hypotheses. Then, put them into knowledge base as facts.

Then we can effectively collect abductive hypotheses and inductive hypotheses that can be used to build models for conducting abductive and inductive risk management. Of course this type of cooperation can be extended to actual risk management where both types of risk management help each other.

According to the above perspective, Fig. 3 illustrates a prototype image of cooperative or integrated risk management. As shown in the figure, with cooperation between abductive and inductive risk management, better risk management can be achieved.

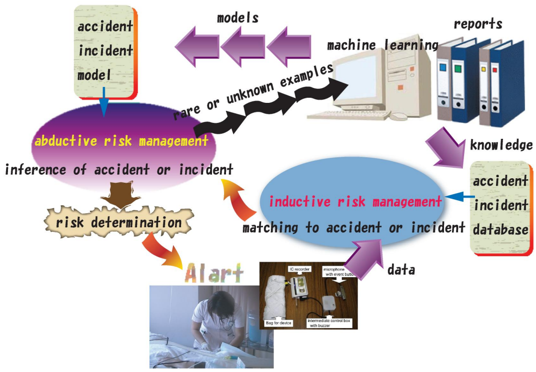


Fig. 3. Cooperation between both types of risk management

We previously pointed out that if the risk management system issues a warning even in a safe case, it might be harmful because nurses will ignore such meaningless messages in the future. Usually inductive risk management issues such a warning, but in the model shown in Fig. 3, even when inductive risk management is applied, abductive risk management confirms the possibility of occurring accident or incident. Then, the problem of ignoring message will be avoided. Thus cooperative or integrated risk management will provide

5 Conclusions

In this paper, we contrasted abductive and inductive nursing risk management, characterizing them as static and dynamic types of risk management. Static risk management is rather easy to understand, but cannot deal with novel or rare situations. On the other hand, dynamic risk management can deal with novel or rare situations, though the inference mechanism is not easy to understand. We conclude that both styles of risk management are necessary and should cooperate with each other to produce more effective and flexible risk management. We have therefore proposed a cooperative procedure between abductive and inductive risk management

Our current status is that we are constructing an abductive nursing risk management model by referring to cognitive models of human behaviour. The model is extended to deal with continuous or series of events using a concept of a scenario. For inductive risk management, we are now still collecting necessary data including nurses' (self-)conversations or dialogues, the number of footsteps, body angle, from hospitals. Then we are transcribing voice data, and cleaning the obtained data for data or text mining. To build nurses' workflow models, Naya et al. analyzed physical data such as footstep numbers and body angle etc. [16]. Certain results have been obtained, but there still seems to exist a symbol grounding problem. In addition, we are now focusing on voice data analysis because it contains actual names of activities. The transcribed voice date are manually tagged according to the authorized job categories for analysis. Although both data are from the same project(E-nightingale project), for several reasons, they are separately analysed.

Currently our first aim is to construct a set of nursing corpora for analysis of nursing activities [20]. In the E-nightingale framework, we also proposed integrated data mining that combines or complements necessary data obtained in the E-nightingale project [5]. In the future, both types of data should be combined or complemented for analysis to build better nursing activity models.

Acknowledgments

This research was supported in part by the National Institute of Information and Communications Technology (NICT).

References

1. Abe A.: Abductive Analogical Reasoning, *Systems and Computers in Japan*, Vol. 31, No. 1, pp. 11–19 (2000)
2. Abe A.: On The Relation between Abductive Hypotheses and Inductive Hypotheses, *in Abduction and Induction (Flach P. A. and Kakas A. C. eds.)*, pp. 169–180, Kluwer (2000)
3. Abe A.: The Role of Abduction in Chance Discovery, *New Generation Computing,* Vol. 21, No. 1, pp. 61–71 (2003)
4. Abe A., Kogure K., and Hagita N.: Determination of A Chance in Nursing Risk Management, *Proc. of ECAI2004 Workshop on Chance Discovery, pp. 222–231* (2004)
5. Abe A., Kogure K., Hagita N. and Iseki H.: Integrated Data Mining System for Analysis of Nursing Accidents and Incidents, *Proc. of SCI2004, Vol. VII, pp. 265–270* (2004)
6. Abe A., Ozaku H. I., Kuwahara N., and Kogure K.: Scenario Violation in Nursing Activities — Nursing Risk Management from the viewpoint of Chance Discovery, *Soft Computing Journal, Springer* (2006) to appear
7. Abe A., Ozaku H.I., Kuwahara N., and Kogure K.: Scenario-base Construction for Abductive Nursing Risk Management, *Proc. of IPUM2006, pp. 206–213* (2006)
8. Abe A., Ozaku H. I., Kuwahara N., and Kogure K.: What Should be Abducible in Abductive Nursing Risk Management?, *Proc. of KES2006, Vol III, pp. 22–29* (2006) to appear
9. Brusoni V., Console L., Terenziani P., and Dupré D.T.: An Efficient Algorithm for Temporal Abduction. *Proc. 5th Congress of the Italian Association for Artificial Intelligence, pp. 195–206* (1997)
10. Chasen homepage. `http://chasen.aist-nara.ac.jp/`
11. Database of accident and incident reports, `http://www.hiyari-hatto.jp/`. (in Japanese)

12. Japan Academy of Nursing Science eds.: Classification of Nursing Practice, *Japan Academy of Nursing Science* (2005)(in Japanese).
13. Japane Nursing Association eds.: Nursing Practice Classification Table, *Japane Nursing Association* (2004) (in Japanese).
14. Kuwahara N. et al.: Ubiquitous and Wearable Sensing for Monitoring Nurses' Activities, *Proc. SCI2004, Vol. VII, pp. 281–285* (2004)
15. Matsuoka K. and Yokoyama S.: Analysis of Incident Data by Medical Data Mining (No. 2), *Medical Treatment, Inspection Machine and Reagent, Vol. 25, No. 4, pp. 275–285* (2002) (in Japanese)
16. Naya F. et al.: Location and Activity Recognition using Hetrogeneous Sensors, *Proc. of KSEL2006, pp. 25–31* (2006)
17. Ohsawa Y., Benson N. E. and Yachida M.: KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor, *Proc. Advanced Digital Library Conference (IEEE ADL'98)*, pp. 12–18 (1998)
18. Ohsawa Y.: Chance Discovery for Making Decisions in Complex Real World, New Generation Computing, *Vol. 20, No. 2, pp. 143–163* (2002)
19. Ohsawa Y., Okazaki N., and Matsumura N.: A Scenario Development on Hepatics B and C, *Technical Report of JSAI, SIG-KBS-A301, pp. 177–182* (2003)
20. Ozaku H.I., Sagara K., Naya F., Kuwahara N., Abe A., and Kogure K.: Building Dialogue Corpora for Nursing Activity Analysis, *Proc. of LINC-2005 Workshop*, pp. 41–48 (2005)
21. Ozaku H.I., Abe A., Sagara K., Kuwahara N., and Kogure K.: A Task Analysis of Nursing Activites Using Spoken Corpora, *Advances in Natural Language Processing, Research in Computing Science*, Vol. 18, pp. 125–136 (2006)
22. Ozaku H.I., Sagara K., Kuwahara N., Abe A., and Kogure K.: Nursing Spoken Corpora for Understanding Nursing Assignments, *Proc. of NI2006, pp. 481–485* (2006)
23. Poole D., Goebel R., and Aleliunas R.: Theorist: A Logical Reasoning System for Defaults and Diagnosis, *The Knowledge Frontier: Essays in the Representation of Knowledge (Cercone N.J., McCalla G. Eds.), pp. 331–352, Springer-Verlag* (1987)
24. Quinlan J. R.: *C4.5: Programs for Machine Learning, Morgan Kaufman* (1993)
25. Reiter R. and de Kleer J.: Foundation of assumption-based truth maintenance systems: Preliminary report. *Proc. of AAAI87, pp.183–188* (1987)
26. Vincent C., Ennis M., and Audley R. J. eds.: *Medical Accidents,* Oxford University Press (1993)

Author Index