ADVANCES IN *SOFT COMPUTING* 46

Van-Nam Huynh
Yoshiteru Nakamori
Hiroakira Ono
Jonathan Lawry
Vkladik Kreinovich
Hung T. Nguyen (Eds.)

Interval / Probabilistic Uncertainty and Non-classical Logics

Advances in Soft Computing 46

Editor-in-Chief: J. Kacprzyk

Advances in Soft Computing

Further volumes of this series can be found on our homepage: springer.com

Frank Hoffmann, Mario Köppen,
Frank Klawonn, Rajkumar Roy (Eds.)
Soft Computing Methodologies and Applications, 2005
ISBN 978-3-540-25726-4

Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, Kryzysztof Trojanowski (Eds.)
Intelligent Information Processing and Web Mining, 2005
ISBN 978-3-540-25056-2

Abraham Ajith, Bernard de Bacts,
Mario Köppen, Bertram Nickolay (Eds.)
Applied Soft Computing Technologies: The Challenge of Complexity, 2006
ISBN 978-3-540-31649-7

Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, Kryzysztof Trojanowski (Eds.)
Intelligent Information Processing and Web Mining, 2006
ISBN 978-3-540-33520-7

Ashutosh Tiwari, Joshua Knowles,
Erel Avineri, Keshav Dahal,
Rajkumar Roy (Eds.)
Applications and Soft Computing, 2006
ISBN 978-3-540-29123-7

Bernd Reusch, (Ed.)
Computational Intelligence, Theory and Applications, 2006
ISBN 978-3-540-34780-4

Miguel López-Díaz, María ç. Gil,
Przemysław Grzegorzewski, Olgierd Hryniewicz, Jonathan Lawry
Soft Methodology and Random Information Systems, 2006
ISBN 978-3-540-34776-7

Ashraf Saad, Erel Avineri, Keshav Dahal,
Muhammad Sarfraz, Rajkumar Roy (Eds.)
Soft Computing in Industrial Applications, 2007
ISBN 978-3-540-70704-2

Bing-Yuan Cao (Ed.)
Fuzzy Information and Engineering, 2007
ISBN 978-3-540-71440-8

Patricia Melin, Oscar Castillo,
Eduardo Gómez Ramírez, Janusz Kacprzyk,
Witold Pedrycz (Eds.)
Analysis and Design of Intelligent Systems Using Soft Computing Techniques, 2007
ISBN 978-3-540-72431-5

Oscar Castillo, Patricia Melin,
Oscar Montiel Ross, Roberto Sepúlveda Cruz,
Witold Pedrycz, Janusz Kacprzyk (Eds.)
Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, 2007
ISBN 978-3-540-72433-9

Katarzyna M. Węgrzyn-Wolska,
Piotr S. Szczepaniak (Eds.)
Advances in Intelligent Web Mastering, 2007
ISBN 978-3-540-72574-9

Emilio Corchado, Juan M. Corchado,
Ajith Abraham (Eds.)
Innovations in Hybrid Intelligent Systems, 2007
ISBN 978-3-540-74971-4

Marek Kurzynski, Edward Puchala,
Michal Wozniak, Andrzej Zolnierek (Eds.)
Computer Recognition Systems 2, 2007
ISBN 978-3-540-75174-8

Van-Nam Huynh, Yoshiteru Nakamori,
Hiroakira Ono, Jonathan Lawry,
Vladik Kreinovich, Hung T. Nguyen (Eds.)
Interval / Probabilistic Uncertainty and Non-classical Logics, 2008
ISBN 978-3-540-77663-5

Van-Nam Huynh, Yoshiteru Nakamori,
Hiroakira Ono, Jonathan Lawry,
Vladik Kreinovich, Hung T. Nguyen (Eds.)

Interval / Probabilistic Uncertainty and Non-classical Logics

Springer

Editors

Van-Nam Huynh
Yoshiteru Nakamori
Hiroakira Ono

Japan Advanced Institute of Science
and Technology
1-1 Asahidai, Nomi
Ishikawa 923-1292
Japan
E-mail: {huynh,nakamori,ono}@jaist.ac.jp

Jonathan Lawry
Department of Engineering Mathematics,
University Of Bristol
Queen's Building, University Walk
Bristol BS8 1TR
United Kingdom
E-mail: j.lawry@bristol.ac.uk

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968
USA
E-mail: vladik@utep.edu

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003-8001
USA
E-mail: hunguyen@nmsu.edu

ISBN 978-3-540-77663-5 e-ISBN 978-3-540-77664-2

DOI 10.1007/978-3-540-77664-2

Advances in Soft Computing ISSN 1615-3871

Library of Congress Control Number: 2007942801

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

5 4 3 2 1 0

springer.com

Preface

Large-scale data processing is important. Most successful applications of modern science and engineering, from discovering the human genome to predicting weather to controlling space missions, involve processing large amounts of data and large knowledge bases. The corresponding large-scale data and knowledge processing requires intensive use of computers.

Computers are based on processing exact data values and truth values from the traditional 2-value logic. The ability of computers to perform fast data and knowledge processing is based on the hardware support for super-fast elementary computer operations, such as performing arithmetic operations with (exactly known) numbers and performing logical operations with binary ("true"-"false") logical values.

In practice, we need to go beyond exact data values and truth values from the traditional 2-value logic. In practical applications, we need to go beyond such operations.

Input is only known with uncertainty. Let us first illustrate this need on the example of operations with numbers. Hardware-supported computer operations (implicitly) assume that we know the exact values of the input quantities. In reality, the input data usually comes from measurements. Measurements are never 100% accurate. Due to such factors as imperfection of measurement instruments and impossibility to reduce noise level to 0, the measured value $\widetilde{x}$ of each input quantity is, in general, different from the (unknown) actual value x of this quantity. It is therefore necessary to find out how this input uncertainty $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x \neq 0$ affects the results of data processing.

Probabilistic approach to uncertainty. The need to take into account the uncertainty of input data has been known for centuries. In the early 19 century,

Gauss developed basic statistical techniques for processing such uncertainty. These techniques are based on the assumption that we know the probabilities of different values of the measurement error Δx. Usually, we assume that the distribution is normal (Gaussian), with 0 mean and a known standard deviation σ. Such *probabilistic* methods are actively used in engineering and scientific practice.

Interval approach to uncertainty. There are practical situations in which we do not know the probabilities of different values of the measurement error Δx. In many such situations, we only know the upper bound Δ on the (absolute value of the) measurement error, i.e., the value Δ for which $|\Delta x| \leq \Delta$. In such situations, after we perform the measurement and get the measured value $\widetilde{x}$, the only information that we have about the actual (unknown) value x is that this value belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

Techniques for processing data under such *interval* uncertainty can be traced back to ancient scientists such as Archimedes. Their active development started by T. Sunaga (Japan), by M. Warmus (Poland), and especially by R. Moore (USA) in the 1950s, when the arrival of modern computers led to the practical need for such development.

Need to combine probabilistic and interval uncertainty. At present:

- we have well-developed techniques for handling situations in which we know the exact probability distribution for the measurement error Δx, and
- we have well-developed techniques for handling situations in which we have no information about the probabilities – and we only know the upper bound on the measurement error.

In real life, we often encounter intermediate situations in which we have *partial* information about the probabilities.

This information is frequently described in interval-related terms: e.g., instead of knowing the exact value p of the probability, we may only know the interval $[\underline{p}, \overline{p}]$ of possible values of this probability. To handle such situation, it is desirable to combine probabilistic and interval approaches to uncertainty.

Several formalisms have been developed for such combination, such as imprecise probabilities, Dempster-Shafer approach, approaches related to rough sets, and many others.

First objective of the workshop. One of the main objectives of this workshop was to bring together researchers working on interval, probabilistic, and combined methods, so as to promote collaboration and further applications.

Need for non-classical logics. Another aspect in which we need to go beyond hardware-supported computer operations is logic.

In the computer, only operations from the traditional 2-valued ("true"-"false") logic are supported. However, in practice, usually, experts are not 100% sure about the truth of the statement included in the knowledge bases.

An important aspect of this uncertainty is that the experts' statements can be *vague* (*fuzzy*). For example, an expert can be 100% sure that the temperature in a certain region will be high, but he or she is not 100% sure whether this "high" necessarily means greater than 40°.

Fuzzy logic and type-2 fuzzy logic. Several logical techniques have been proposed to provide a more adequate description of expert knowledge. Historically, one of the first such techniques was the (standard) fuzzy logic, in which we use numbers from the interval $[0, 1]$ to describe the expert's degree of certainty in a statement:

- 1 means 100% certainty,
- 0 means no certainty at all, and
- intermediate values represent different grades of uncertainty.

An even more adequate representation of the expert's uncertainty comes from type-2 fuzzy logic, in which we take into account that just like an expert usually cannot describe his or her knowledge about some quantity by a single number, this same expert cannot describe his or her certainty by a single number: this certainty can be described by an interval of possible values, or, even more generally, by a fuzzy subset of the interval $[0, 1]$.

Probabilistic logics. In fuzzy logic, degrees from the interval $[0, 1]$ represent subjective degrees. In some cases, we can describe these degrees in a more objective way: e.g., in the weather example, as a probability that in the past, in similar situations, the temperature was above 40°. Such cases are handled by *probabilistic* logic.

Algebraic approach to logic. Yet another alternative comes from the fact that the most natural way for an expert to describe his/her knowledge is by using words from natural language. So, instead of quantizing these values, we may want to describe possible values of certainty as the set of such words and word combinations, and define appropriate "and"- and "or"-operations on this set – which would make it a logic.

This approach was one of the motivations behind the development of *algebraic* logics, i.e., logics described not by a specific implementation but rather by the properties of the corresponding logical operations.

Modal logics. For an individual event, such as temperature exceeding 40°, it is reasonable to ask whether this event will happen or not – and what is the expert's confidence that this event will happen. In practice, expert statements usually refer not to individual events, but rather to *repeating* events. For such events, it is also reasonable to ask whether it is *possible* that the event will happen, whether it is *necessary* that this event will happen – and if yes, to what degree.

Such statements about possibility and necessity form *modal logic*, another non-classical logic actively used in processing knowledge.

Other non-classical logics. There exist many other non-classical logics, e.g., logics used to describe induction (making generalizations based on several facts) and abduction (making conclusions about the causes of an observed event).

Need to combine uncertainty analysis with non-classical logics. In many practical problems, we need to process both measurement data and expert knowledge.

- We have already mentioned that to adequately process measurement data, we need to take into account probabilistic and interval uncertainty – and combine these two types of uncertainty.
- We have also mentioned that to adequately describe expert knowledge, we need to use various non-classical logic techniques – and sometimes we need to combine different non-classical logic methods.

It is therefore desirable to combine uncertainty analysis with non-classical logic.

Such combination was the main objective of this workshop – and of these proceedings.

Objectives of the workshop. Specifically, the main objectives of the workshop were:

- to bring together researchers working on uncertainty formalisms in information and knowledge systems;
- to attract researchers working in social sciences (economics, business, and environmental sciences) who are interested in applying uncertainty-related tools and techniques;
- to promote the cross-fertilization between the fundamental ideas connected with various approaches used in the study of non-classical logics;
- to bring together researchers from various fields on non-classical logics and applications in order to foster collaboration and further research, and
- to present and discuss open research problems and challenges.

Papers presented in these Proceedings describe different aspects of these problems.

We hope that this workshop will lead to a boost in the much-needed collaboration between the uncertainty analysis communities and the non-classical logic communities.

Acknowledgments. This workshop was partially supported by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08, JAIST 21st COE Program entitled "Technology Creation Based on Knowledge Science", JSPS Grant-in-Aid for Scientific Research [KAKENHI(C) #19500174], and JAIST 21st COE Program entitled "Verifiable and Evolvable e-Society". We are very thankful to JAIST for all the help.

We would also like to thank:

- all the authors for their hard work,
- all the members of the program committee and the external referees for their thorough analysis of submitted papers,
- all the participants for their fruitful discussions, and
- last but not the least, Professor Janusz Kacprzyk, the series editor, for his encouragement and support.

Let the collaborations continue and blossom!

JAIST, Japan
March 2008

Van Nam Huynh
Yoshiteru Nakamori
Hiroakira Ono
Jonathan Lawry
Vladik Kreinovich
Hung T. Nguyen

Contents

List of Contributors

Seiki Akama
1-20-1, Higashi-Yurigaoka, Asao-ku
Kawasaki 215-0012 Japana
kama@jcom.home.ne.jp

Sten F. Andler
School of Humanities and Informatics
University of Skövde, Sweden
sten.f.andler@his.se

Pedro Barahona
Centro de Inteligência Artificial
Universidade Nova de Lisboa
Portugal
pb@di.fct.unl.pt

Yaxin Bi
University of Ulster
Newtownabbey Co. Antrim
BT37 0QB, UK
y.bi@ulster.ac.uk

Tru H. Cao
Faculty of Computer Science and Engineering
Hochiminh City University of Technology Vietnam
tru@cse.hcmut.edu.vn

Elsa Carvalho
Centro de Inteligência Artificial
Universidade Nova de Lisboa
Portugal
elsac@uma.pt

Xi Chen
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, China
chenhbs@126.com

Kwai-Sang Chin
City University of Hong Kong
Hong Kong, China
mekschin@cityu.edu.hk

Jorge Cruz
Centro de Inteligência Artificial
Universidade Nova de Lisboa
Portugal
jc@di.fct.unl.pt

Fabio Cuzzolin
INRIARhône-Alpes
655 avenue de l'Europe
38334 SAINT ISMIER CEDEX
France
Fabio.Cuzzolin@inrialpes.fr

Scott Ferson
Applied Biomathematics
Setauket, NY 11733, USA
scott@ramas.com

Min Guo
Huazhong University of Science and Technology
Wuhan, China
guominlike@sina.com.cn

Rolf Haenni
Bern University of Applied Sciences
2501 Biel, Switzerland
rolf.haenni@bfh.ch

Ai Han
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, China
hanai06@mails.gucas.ac.cn

John Harding
New Mexico State University
Las Cruces, NM 88003, USA
jharding@nmsu.edu

Tu Bao Ho
Japan Advanced Institute of Science and Technology Ishikawa, 923-1292, Japan
bao@jaist.ac.jp

Eberhard P. Hofer
University of Ulm
89069 Ulm, Germany
eberhard.hofer@uni-ulm.de

Chenyi Hu
University of Central Arkansas
Conway AR 72035, USA
CHu@uca.edu

Van Nam Huynh
Japan Advanced Institute of Science and Technology
Ishikawa 923-1292, Japan
huynh@jaist.ac.jp

Masahiro Inuiguchi
Osaka University
Osaka 560-8531, Japan
inuiguti@sys.es.osaka-u.ac.jp

Ronnie Johansson
School of Humanities and Informatics
University of Skövde, Sweden
ronnie.johansson@his.se

Alexander Karlsson
School of Humanities and Informatics
University of Skövde, Sweden
alexander.karlsson@his.se

Olga Kosheleva
University of Texas at El Paso
El Paso, TX 79968, USA

Vladik Kreinovich
University of Texas at El Paso
El Paso, TX 79968-0518, USA
vladik@utep.edu

Yasuo Kudo
Muroran Institute of Technology
Muroran 050-8585, Japan
kudo@csse.muroran-it.ac.jp

Jonathan Lawry
University of Bristol
Bristol BS8 1TR, UK
J.Lawry@bris.ac.uk

Karin Mehr
University of Ulm
89069 Ulm, Germany
karin.mehr@uni-ulm.de

David Picado Muiño
University of Manchester
Manchester M13 9PL, UK
davidpicado@yahoo.com

Tetsuya Murai
Hokkaido University
Sapporo 060-0814, Japan
murahiko@main.ist.hokudai.ac.jp

Yoshiteru Nakamori
Japan Advanced Institute of Science and Technology
Ishikawa 923-1292, Japan
nakamori@jaist.ac.jp

Ma Nam
Faculty of Computer Science and Engineering
HoChiMinh City University of Technology Vietnam
manam@cse.hcmut.edu.vn

Hoa Nguyen
HoChiMinh City Open University
HoChiMinh City, Vietnam
hoa-hanh@hcm.vnn.vn

Hung T. Nguyen
New Mexico State University
Las Cruces, NM 88003-8001, USA
hunguyen@nmsu.edu

Canh Hao Nguyen
Japan Advanced Institute of Science and Technology
Ishikawa, 923-1292, Japan
canhhao@jaist.ac.jp

Hiroakira Ono
Japan Advanced Institute of Science and Technology
Ishikawa 923-1292, Japan
ono@jaist.ac.jp

Jeff B. Paris
University of Manchester
Manchester M13 9PL, UK
jeff.paris@manchester.ac.uk

Tho T. Quan
Faculty of Computer Science and Engineering
HoChiMinh City University of Technology Vietnam
qttho@cse.hcmut.edu.vn

Jan-Willem Romeijn
University of Groningen
9712 GL Groningen, The Netherlands
J.W.Romeijn@rug.nl

Michael Rosefield
University of Manchester
Manchester M13 9PL, UK
rosyatrandom@gmail.com

Mina Ryoke
University of Tsukuba
Tokyo 112-0012, Japan
ryoke@mbaib.gsbs.tsukuba.ac.jp

Xuhui Shen
Institute of Earthquake Science
Beijing 100036 China
shenxh@seis.ac.cn

Noboru Takagi
Toyama Prefectural University
Toyama 939-0398, Japan
takagi@pu-toyama.ac.jp

Milan Vlach
Charles University
118 00 Praha 1, Czech Republic
milan.vlach@mff.cuni.cz

Carol L. Walker
New Mexico State University
Las Cruces, NM 88003, USA
hardy@nmsu.edu

Elbert A. Walker
New Mexico State University
Las Cruces, NM 88003, USA
elbert@nmsu.edu

Hongwei Wang
Huazhong University of Science and Technology
Wuhan, China

Gregory Wheeler
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
grw@fct.unl.pt

Jon Williamson
University of Kent
Canterbury CT2 7NF, UK
j.williamson@kent.ac.uk

Shengli Wu
University of Ulster
Newtownabbey Co. Antrim
BT37 0QB, UK
s.wu@ulster.ac.uk

Gang Xiang
University of Texas at El Paso
TX 79968, USA
gxiang@acm.org

Shanying Xu
Chinese Academy of Sciences
Academy of Mathematics and
Systems Science, China
xsy@iss.ac.cn

Hongbin Yan
Japan Advanced Institute of Science
and Technology
Ishikawa 923-1292, Japan
hongbinyan@jaist.ac.jp

Jian-Bo Yang
The University of Manchester
Manchester M15 6PB, UK
jian-bo.yang@umist.ac.uk

Yukihiro Yoshioka
Osaka University
Osaka 560-8531, Japan
yoshioka@inulab.sys.es.osaka-u.ac.jp

Part I

Keynote Addresses

An Algebraic Approach to Substructural Logics – An Overview

Hiroakira Ono

Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
ono@jaist.ac.jp

We will give a state-of-the-art survey of the study of substructural logics. Originally, substructural logics were introduced as logics which, when formulated as Gentzen-style systems, lack some of the three basic *structural rules*, i.e. *contraction*, *weakening* and *exchange*. For example, relevance logics and linear logic lack the weakening rule, many-valued logics, fuzzy logics and linear logic lack the contraction rule, and hence all of them can be regarded as substructural logics. These logics have been studied extensively and various theories have been developed for their investigation. However their study has been carried out independently, mainly due to the different motivations behind them, avoiding comparisons between different substructural logics.

On the other hand, the general study of substructural logics has a comparative character, focusing on the absence or presence of structural rules. As such, at least in the initial stages of research, it was a study on how structural rules affect logical properties. This naturally led to a syntactic or proof-theoretic approach yielding deep results about properties of particular logics, provided that they can be formalized in systems, like cut-free Gentzen calculi. An obvious limitation of this study comes from the fact that not all logics have such a formulation.

Semantical methods, in contrast, provide a powerful tool for analyzing substructural logics from a more uniform perspective. Both Kripke-style semantics and algebraic semantics for some particular subclasses of substructural logics, e.g. relevant logics, were already introduced in the 70s and 80s and were studied to a certain extent.

An algebraic study of substructural logics for the last decade, based mainly on algebraic logic and universal algebra, has brought us a new perspective on substructural logics, which came from the observation that these logics all share the *residuation property*. Though this is usually not noticed, it is revealed explicitly in a sequent formulation by the use of extralogical symbols, denoted by commas. Precisely speaking, consider the following equivalence, concerning implication $\rightarrow$, that holds in most of sequent systems: *A formula γ follows from formulas α and β if and only if the implication $\alpha \rightarrow \gamma$ follows from β alone.*

Here 'follows from' is given by the particular substructural logic, and is usually denoted by $\Rightarrow$. Thus, the above equivalence can be restated as:

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 3–4, 2008.
springerlink.com

$$\alpha, \beta \Rightarrow \gamma \text{ is provable} \quad \text{iff} \quad \beta \Rightarrow \alpha \to \gamma \text{ is provable},$$

where provability is taken with respect to a sequent calculus for the particular substructural logic. If we replace the auxiliary symbol 'comma' by a new logical connective $\cdot$, called *fusion*, we have:

$$\alpha \cdot \beta \Rightarrow \gamma \text{ is provable} \quad \text{iff} \quad \beta \Rightarrow \alpha \to \gamma \text{ is provable}.$$

Algebraically, this can be expressed as

$$a \cdot b \leq c \quad \text{iff} \quad b \leq a \to c$$

which is known as the *law of residuation* in residuated (ordered) structures.

Thus, this observation naturally leads us to the thesis that *substructural logics are exactly logics of residuated structures*. This will explain why substructural logics, especially when formulated in Gentzen-style sequent systems, encompass most of the interesting classes of non-classical logics. For, implication, admittedly the most important logical connective, can be understood as the *residual* of fusion, a connective that behaves like a semigroup or groupoid operation. From a mathematical point of view, it is easier to discuss the latter than the former, just like developing a theory of multiplication of numbers is easier than developing a theory of division.

This is a starting point of our algebraic approach to substructural logics. Due to significant advances in the study of both residuated lattices and the abstract algebraic logic in the recent years, the research field is developing rapidly. Recent research reveals us that there are strong interplays between algebra and logic, and even between algebraic methods and proof-theoretic ones. These facts will provide us a deeper understanding of the subject and suggest us new directions of research.

The source of the present abstract is the book [1], jointly written with Galatos, Jipsen and Kowalski, where these topics are extensively discussed.

Reference

1. Galatos, N., et al.: Residuated Lattices: An Algebraic Glimpse at Substructural Logics. In: Studies in Logic and the Foundations of Mathematics, vol. 151, Elsevier, Amsterdam (2007)

On Modeling of Uncertainty Measures and Observed Processes

Hung T. Nguyen

Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003-8001, USA
hunguyen@nmsu.edu

Summary. This paper is about a short survey of some basic uncertainty measures in systems analysis arising from coarse data, together with new modeling results on upper semicontinuous random processes, viewed as random fuzzy sets. Specifically, we present the most general mathematical framework for analyzing coarse data, such as random fuzzy data, which arise often in information systems. Our approach is based upon the theory of continuous lattices. This probabilistic analysis is also useful for investigating upper semicontinuous random functions in stochastic optimization problems.

1 Introduction

Empirical data are necessary in almost all fields of science, especially in decision-making where theoretical criteria need to be validated from observed data. A typical situation is in mathematical finance where investment decisions are based upon risk measures (e.g. Levy, 2006 [8]). Future returns X, Y on two different prospects should be compared for selection. A partial order relation on random variables is defined in the spirit of Von Neumann and Morgenstern's expected utility theory as $X \succeq Y$ (X is preferred to Y) if and only if for all increasing function $u : \mathbb{R} \to \mathbb{R}$, $Eu(X) \geq Eu(Y)$. A characterization of $\succeq$ turns out to be $F(\cdot) \leq G(\cdot)$ where F, G are distribution functions of X, Y, respectively. This is a *theory* for risk assessment which is based upon distribution functions of random variables involved. But we never know these distributions! Thus, in order to make decisions (i.e. choosing investment prospects) we need to use empirical data to check whether this stochastic dominance order is valid. This is a hypothesis testing problem. In other words, statistics is needed to support decision-making. But statistical inference procedures depend heavily on which *type* of available data. For example, if available data are only observed within bounds. Such type of imprecise data is referred to as *coarse data* (i.e. data with low quality, e.g. Heitjan and Rubin, 1991 [7]), such as missing outcomes, censored data, partially observed data, interval data on outcomes or covariates are available, rather than point measurements, hidden Markov data in bioinformatics (e.g., Cappé et al., 2005 [1]), indirect observed data (e.g., in auction theory, Paarsch and Hong, 2006 [14]).

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 5–15, 2008.
springerlink.com

It is precisely the type of observed data which dictates inference procedures. Thus, when random data are imprecise, such as only known to lie within bounds, we face interval statistics where interval computations should be called upon, say, to extend standard statistical procedures (e.g. estimation and testing) to the new type of data. More generally, the theory of random sets, e.g., Matheron, 1975 [9]; Nguyen, 2006 [11], is necessary for inference with random set data. In intelligent systems building, we encounter perception-based information (Zadeh, 2002 [17]) which is fuzzy in nature (see also e.g., Nguyen and Wu, 2006 [12]). For each type of coarse data we need an appropriate mathematical modeling for the observed process in order to carry out any inference procedures, such as model building and forecasting.

In this paper, we focus on the case of coarse data which are both random and fuzzy, as in coarsening schemes of human perception-based information gathering processes. We provide a general and rigorous mathematical model for random fuzzy sets, extending Matheron's theory of random closed sets.

In analyzing coarse data, we come across various types on uncertainty, a survey of basic aspects of modeling will be given first.

2 Some Uncertainty Measures Derived from Coarse Data

First, let's look at a standard situation where coarse data are *set-valued observations*. While set-valued observations, i.e. outcomes of random experiments or records of natural phenomena, have different interpretations, depending on the goals of the analysis, such as tumor growth patterns in medical statistics, shape analysis, the specific situation related to coarse data is this. Let X be a random vector of interest. Either by performing a random experiment or observing X in a sample data $X_1, X_2, \ldots, X_n$, to discover, say, the distribution of X, we are unable to observe or measure this sample with accuracy. Instead, what we observe is a collection of sets $S_1, S_2, \ldots, S_n$ which contain the sample, i.e. $X_i \in S_i$, $i = 1, 2, \ldots, n$. The statistical problem is the same, but instead of using $X_1, X_2, \ldots, X_n$, we only have at our disposal the coarse sample $S_1, S_2, \ldots, S_n$. This clearly is a generalization of multivariate statistical analysis. In order to analyze the set-valued observations, we need to model the observation process. Since probability theory provides us with a fairly general setting, namely random elements in general measurable spaces of arbitrary nature, such as metric spaces, we can simply view $S_1, S_2, \ldots, S_n$ as a random sample from a *random set* S which contains X almost surely, i.e. X is an almost sure selector of S, or the other way around, S is a coarsening of X. Random set models for coarse data turn out to be useful in exhibiting various uncertainty measures in artificial intelligence.

2.1 Belief Functions

Consider the case where S is a coarsening of X on a finite set U, to avoid topological details. Let $A \subseteq U$ be an event. A is said to occur if $X(\omega) \in A$. But

if we cannot observe $X(\omega)$, but only $S(\omega)$, then clearly we are even uncertain about the occurrence of A. If $S(\omega) \subseteq A$, then clearly, A occurs. So, from a "pessimistic" viewpoint, we *quantify* our degrees of belief in the occurrence of an event A by $P(S \subseteq A)$ which is less than the actual probability that A occurs, namely $P(X \in A)$, since X is an a.s. selector of S . This is in fact the starting point for the now well-known Dempster-Shafer theory of evidence or of belief functions (e.g. Shafer, 1976 [15]).

Let $F : 2^U \to [0,1]$, where 2^U denotes the power set of U, be defined by $F(A) = P(S \subseteq A)$, which is the distribution function of the random set S, in the sense that F determines the probability law of S. Indeed, the set-function F satisfies the following basic properties:

(i) $F(\varnothing) = 0, F(U) = 1$
(ii) For any $n \geq 1$, and $A_1, A_2, ..., A_n$,

$$F(\bigcup_{i=1}^{n} A_i) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} F(\bigcap_{i \in I} A_i)$$

Clearly, (i) follows from the fact that S is a non-empty random set with values in 2^U. As for (ii), this is a weakening of Poincare's equality for probability measures. Observe that since U is finite,

$$F(A) = P(S \subseteq A) = \sum_{B \subseteq A} f(B) \text{ where we set } f(B) = P(S = B)$$

Let $J(B) = \{i \in \{1, 2, ..., n\} : B \subseteq A_i\}$. Clearly, we have

$$\{B : J(B) \neq \varnothing\} \subseteq \{B : B \subseteq \cup_{i=1}^{n} A_i\}$$

Now

$$\begin{aligned} P(\{B : J(B) \neq \varnothing\}) &= \sum_{B \subseteq U, J(B) \neq \varnothing} f(B) \\ &= \sum_{B \subseteq U, J(B) \neq \varnothing} f(B)[\sum_{\varnothing \neq I \subseteq J(B)} (-1)^{|I|+1}] \\ &= \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} \sum_{B \subseteq \cap_{i \in I} A_i} f(B) \\ &= \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i) \end{aligned}$$

To see that the axiomatic theory of belief functions is precisely the axiomatization of distributions of random sets, exactly like the case of random variables, it suffices to show the converse. Let $F : 2^U \to [0,1]$ such that

(i) $F(\varnothing) = 0, F(U) = 1$
(ii) For any $n \geq 1$, and $A_1, A_2, ..., A_n$,

$$F(\cup_{i=1}^{n} A_i) \geq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

Then there exist a probability space $(\Omega, \mathcal{A}, P)$ and a non-empty random set $S : \Omega \to 2^U$ such that $F(A) = P(S \subseteq A)$.

Indeed, by Möbius inversion, we have

$$f(A) = \sum_{B \subset A} (-1)^{|A \setminus B|} F(B)$$

which is a bona fide probability density function on 2^U.

2.2 Possibility Measures

Again, let S be a coarsening of X on a finite set U. If $S(\omega) \cap A \neq \varnothing$, then all we can say is that "*it is possible that A occurs*". A plausible way to *quantify* these degrees of possibility is to take $P(S(\omega) \cap A \neq \varnothing)$.

First, this seems to be consistent with the common sense that possibilities are larger than probabilities since possibilities tend to represent an "optimistic attitude" as opposed to beliefs.

This is indeed the case since, as an a.s. selector, we clearly have $\{X \in A\} \subseteq \{S \cap A \neq \varnothing\}$, and hence $P(\{X \in A\}) \leq P(\{S \cap A \neq \varnothing\})$.

Now observe that the set-function $T(A) = P(\{S \cap A \neq \varnothing\})$ is dual to the belief function F via $T(A) = 1 - F(A^c)$, the monotonicity of infinite order of F above implies the alternating of infinite order of T, namely

$$T(\bigcap_{i=1}^{n} A_i) \leq \sum_{\varnothing \neq I \subseteq \{1,2,...,n\}} (-1)^{|I|+1} T(\bigcup_{i \in I} A_i)$$

which still characterizes the distribution of a random set. However, not all such T can be used to model possibility measures, since possibility measures need to be truth-functional. According to Zadeh [16], a subjective concept of "possibility distributions" is primitive: $\pi : U \to [0, 1]$, just like a membership function of a fuzzy concept. From π, possibilities of events could be derived. Now, in our coarsening scheme, there exists a *canonical random set* which does just that! This is completely analogous to the situation in *survey sampling* in applied statistics! The canonical random set has its roots in an early work of Goodman [5] concerning relations between fuzzy sets and random sets.

From $T(A) = P(\{S \cap A \neq \varnothing\})$, we see that when $A = \{u\}, T(\{u\}) = P(u \in S) = \pi(u)$, the covering function of the random set S, where $\pi : U = \{u_1, ..., u_k\} \to [0, 1]$. Given π, there exist many random sets S admitting $\pi(\cdot)$ as their common covering function. Indeed, let $V_i(\omega) = S(\omega)(u_i),\ i = 1, 2, \ldots, k$, where, again, for $A \subseteq U$, we write $A(u)$ for the value of the indicator function of the set A at u. Each V_i is a $\{0, 1\}$-valued random variable with

$$P(V_i = 1) = P(u_i \in S) = \pi(u_i)$$

The distribution of the random vector $V = (V_1, V_2, \ldots, V_k)$ is completely determined by P_S and vice versa. Indeed, for any $x = (x_1, x_2, \ldots, x_k) \in \{0, 1\}^k$,

we have $P(V = x) = P(S = B)$, where $B = \{u_i \in U : x_i = 1\}$. The distribution function of V_i is

$$F_i(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 - \pi(u_i) & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

Thus, given the marginals $F_i,\ i = 1, 2, \ldots, k$, the joint distribution function F of V is the form

$$F(y_1, y_2, \ldots, y_k) = C(F_1(y_1), F_2(y_2), \ldots, F_k(y_k))$$

where C is an *k-copula*, according to Sklar's theorem (e.g., Nelsen, 1999 [10]).

For $C(y_1, y_2, \ldots, y_k) = \prod_{i=1}^{k} y_i$, we obtain the well-known Poisson sampling design (in sampling from a finite population) with density (of the random set S, the probability sampling plan):

$$f(A) = \prod_{j \in A} \pi(j) \prod_{j \in A^c} (1 - \pi(j))$$

For $C(y_1, y_2, \ldots, y_k) = \bigwedge y_j$, we get

$$f(A) = \sum_{B \subseteq A} (-1)^{|A-B|} [1 - \max\{\pi(j) : j \in B^c\}]$$

which is the density of the following nested random set (canonical):

Let $\alpha : \Omega \to [0, 1]$ be a random variable, uniformly distributed. Define $S(\omega) = \{u \in U : \pi(u) \geq \alpha(\omega)\}$. Then $\pi(u) = P(u \in S)$ and moreover,

$$P(S \cap B \neq \emptyset) = P\left(\omega : \alpha(\omega) \leq \max_{u \in B} \pi(u)\right) = \max_{u \in B} \pi(u)$$

Thus, using this canonical random set (a coarsening scheme) we arrive at Zadeh's axioms for possibility measures.

3 Canonical Borel-σ Fields and Continuous Lattices

In standard statistics, we model observed data as random variables, vectors or random functions by specifying canonical σ-fields on their range spaces. These are measurable mappings, defined on an abstract probability space $(\Omega, \mathcal{A}, P)$, with values in $\mathbb{R}$, $\mathbb{R}^d$, $C[0, 1]$ (space of continuous functions defined on $[0, 1]$), respectively. The canonical σ-field on $\mathbb{R}^d$ is taken to be the σ-field generated by the ordinary topology of $\mathbb{R}^d$, and that of $C[0, 1]$ is the one generated by the topology of the sup-norm. When observed data are discrete sets, closed sets (of some topological space) or upper semicontinuous functions, it is natural to inquire about canonical topologies on these spaces of sets, in some unified manner, in order to consider associated Borel σ-fields to model rigorously these random elements.

So far, as far as we know, σ-fields on these spaces are defined in some ad-hoc manner. Discrete random sets (e.g., on $\mathbb{Z}^2$) are defined in Goutsias [6], random closed sets are defined by using the hit-or-miss topology in Matheron [9], and compact random fuzzy sets are defined in, e.g., Gil et al. [4]. Now observe that these new types of observed data have natural order structures, and as such, we should look for topologies generated by such natural order structures. It turns out that these order structures are all of a very special types if we look at them in the right "angle"! Specifically, they form continuous lattices, and as such, canonical topologies on them should provide desirable σ-fields for modeling these observed data as random elements appropriately.

For ease of reading, let us recall the essentials of continuous lattices. For further details on continuous lattices, we refer the reader to Gierz et al. [2, 3]. It is known that there is a canonical Hausdorff and compact topology (called the Lawson topology) on every continuous lattice, and the space of closed sets of a Hausdorff and locally compact space is a continuous lattice. We give here essential background details for LCHS (locally compact, Hausdorff and second countable) space X establishing that the space $\mathcal{F}$ is a compact, Hausdorff and second countable space whose Lawson topology coincides with the hit-or-miss topology. As we will see in the next section, the space $USC(X)$ of upper semi-continuous functions is also a continuous lattice, and hence its Lawson topology is a natural topology to consider.

Recall that if $(L, \leq)$ is a poset, then x is said to be *way below* y, denoted as $x \ll y$, iff for all directed sets $D \subseteq L$ for which $\sup D$ exists, the relation $y \leq \sup D$ always implies $\exists d \in D$ such that $x \leq d$. Note that in a complete lattice, $x \ll y$ iff for any $A \subseteq L$, $y \leq \sup A$ implies the existence of a finite subset $B \subseteq A$ such that $x \leq \sup B$.

A lattice $(L, \leq)$ is called a continuous lattice if L is complete and satisfies the *axiom of approximation:*

$$x = \sup \Downarrow x, \text{ where } \Downarrow x = \{u \in L : u \ll x\} \text{ for all } x \in L.$$

On the continuous lattice L, the Lawson topology has as subbase the sets of the form

$$\Uparrow x = \{y \in L : x \ll y\}$$

or

$$L \backslash \uparrow x = \{y \in L : x \not\leq y\}, x \in L$$

The sets $\Uparrow x$ form a base for the Scott topology.

4 Discrete Random Sets

In image processing using random set approach instead of random fields (Goutsias, 1997 [6]), it is necessary to provide a rigorous foundation for discrete random sets, e.g., random sets on $\mathbb{Z}^2$. Now closed sets of discrete spaces, with their discrete topologies, are power sets which turn out to be continuous lattices under set inclusion. As such, there exist canonical σ-fields for defining discrete random

sets. The σ-field proposed in Goutsias [6] for the discrete space $\mathbb{N}$, for example, is precisely the canonical σ-field generated by the Lawson topology on the continuous lattice $(2^{\mathbb{N}}, \subseteq)$.

Indeed, consider the LCHS space $(\mathbb{N}, 2^{\mathbb{N}})$ with the discrete topology where $2^{\mathbb{N}}$ is the space of closed sets of $\mathbb{N}$. We see that $(2^{\mathbb{N}}, \subseteq)$ is a continuous lattice. This is so because any $B \subseteq \mathbb{N}$ can be written as $B = \cup\{A \subseteq \mathbb{N}\colon A \ll B\}$ which is obvious since for any $b \in B$, $\{b\} \ll B$ and $B = \cup_{b\in B}\{b\}$. As such, we look at its Lawson topology. Now observe that if $A \subseteq \mathbb{N}$ is an infinite subset, then

$$\Uparrow A = \{B \subseteq \mathbb{N} : A \ll B\} = \varnothing$$

Next, if A is finite, then

$$\Uparrow A = \{B \subseteq \mathbb{N} : A \subseteq B\} = \cap_{x\in A}\{B : \{x\} \subseteq B\}$$

Thus, the sets of the forms $\{B \subseteq \mathbb{N} : \{x\} \subseteq B\}$, $x \in \mathbb{N}$, form a base for the Scott topology. The other type of sets in the subbase of the Lawson topology is $\{B : A \nsubseteq B\}$, $A \subseteq \mathbb{N}$. The σ-field $\lambda(2^{\mathbb{N}})$ generated by the Lawson topology is also generated by $\{B \subseteq \mathbb{N} : A \subseteq B\}, A \subseteq \mathbb{N}$, since σ-fields are closed under set complement. Also, if $A \subseteq \mathbb{N}$, we have

$$\{B \subseteq \mathbb{N} : A \subseteq B\} = \cap_{x\in A}\{B : \{x\} \subseteq B\}$$

and by closure under countable intersections, $\lambda(2^{\mathbb{N}})$ is seen to be generated by $\{B \subseteq \mathbb{N} : \{x\} \subseteq B\} = \{B \subseteq \mathbb{N} : x \in B\}$, $x \in \mathbb{N}$. Thus, $\lambda(2^{\mathbb{N}})$ is generated by $\{B \subseteq \mathbb{N} : F \cap B \neq \varnothing\}$ $(= \cup_{x\in F}\ \{B \subseteq \mathbb{N} : x \in B\})$ for all finite F, i.e. $\lambda(2^{\mathbb{N}})$ coincides with Goutsias's σ-field for discrete random sets.

5 The Space of Closed Sets as a Continuous Lattice

In viewing the space $\mathcal{F}(X)$ of closed sets of a LCHS space X as a continuous lattice with respect to the partial order $\supseteq$, we show that the hit-or-miss topology coincides with the Lawson topology on $\mathcal{F}(X)$. Moreover, since X is second countable, $\mathcal{F}(X)$ is a compact, Hausdorff and second countable topological space, and hence metrizable.

First, note that $(\mathcal{F}(X), \subseteq)$ is a complete lattice but not continuous in general, we have

Proposition 1. *$\bigwedge\{F_i : i \in I\} = \bigcap\{F_i : i \in I\}$, and $\bigvee\{F_i : i \in I\} =$ the closure of $\bigcup\{F_i : i \in I\}$.*

To see that, take $X = \mathbb{R}$, we notice that if $A \ll \mathbb{R}$, then for any subset B of A, i.e $B \subseteq A$, we also have $B \ll \mathbb{R}$ (just use the equivalent condition of the way-below relation). Then any singleton closed set, e.g. $\{0\}$, is not way-below $\mathbb{R}$. Indeed,

$$\bigvee_{n\in\mathbb{N}}\{(-\infty, -1/n] \cup [1/n, \infty)\} = \mathbb{R}$$

but we can not find any finite subset A of $\{(-\infty, -1/n] \cup [1/n, \infty)\}_{n\in\mathbb{N}}$ such that $\{0\} \subseteq \bigvee A$. Therefore, the only closed set that is way-below $\mathbb{R}$ is the empty set. Then

$$\sup\{A \in \mathcal{F}(\mathbb{R}) : A \ll \mathbb{R}\} = \sup\{\emptyset\} = \emptyset \neq \mathbb{R}$$

However, for locally compact X, $(\mathcal{F}(X), \supseteq)$ is a continuous lattice. Indeed, for any $F \in \mathcal{F}(X)$, it is enough to show that $F \leq \sup\{A \in \mathcal{F}(X) : A \ll F\}$, i.e. $F \supseteq \bigcap\{A \in \mathcal{F}(X) : A \ll F\}$ or $F^c \subseteq \bigcup\{A^c \in \mathcal{F}(X) : A \ll F\}$. For any $x \in F^c$: open, since X is locally compact, there exists a compact set $Q_x \subseteq F^c$ such that its interior W_x containing x. Let $A = W_x^c \in \mathcal{F}(X)$, then $A^c = W_x \subseteq Q_x \subseteq F^c$. It follows that $A \ll F$, and therefore, $x \in \cup\{A^c \in \mathcal{F}(X) : A \ll F\}$. For more details, see Nguyen and Tran [13].

The following result shows that the hit-or-miss topology coincides with the Lawson topology on $(\mathcal{F}(X), \supseteq)$.

Proposition 2. *The Lawson topology of $\mathcal{F}(X)^{op}$, denoted by τ_F, has a subbase consisting of sets of the following form*

$$\{F \in \mathcal{F}(X) : F \cap K = \emptyset\} \text{ and } \{F \in \mathcal{F}(X) : F \cap U \neq \emptyset\}$$

where $K \in \mathcal{K}$ and $U \in \mathcal{G}$.

Proof. We only need to verify that for any $A \in \mathcal{F}(X)$,

$$\mathcal{F}(X)\backslash \uparrow A = \{F \in \mathcal{F}(X) : F \cap A^c \neq \emptyset\}$$

and then we just let $U = A^c$.

Indeed, since $\uparrow A = \{F \in \mathcal{F}(X) : A \leq F\} = \{F \in \mathcal{F}(X) : F \subseteq A\}$, we have

$$\mathcal{F}(X)\backslash \uparrow A = \{F \in \mathcal{F}(X) : F \not\subseteq A\} = \{F \in \mathcal{F}(X) : F \cap A^c \neq \emptyset\}$$

In fact, $\{F \in \mathcal{F}(X) : F \cap K = \emptyset\}_{K\in\mathcal{K}}$ is closed under finite intersection, so the Lawson topology of $\mathcal{F}(X)^{op}$ has as a base the sets of the form

$$\{F \in \mathcal{F}(X) : F \cap K = \emptyset \text{ and } F \cap U_i \neq \emptyset, i = 1, \ldots, n\}$$

where $K \in \mathcal{K}$ and $U_i \in \mathcal{G}$.

Moreover, for a LCHS space $(X, \mathcal{G})$, the space $\mathcal{F}(X)^{op}$ is compact, Hausdorff and second countable (and hence metrizable).

6 The Space of USC Functions as a Continuous Lattice

As emphasized by Zadeh [17], most of the information used by humans in control and decisions are based upon perception. Now perception-based information are imprecise and uncertain, and as such, can be modeled by probability theory and fuzzy sets theory.

The process underlying human perception-based gathering can be modeled as random fuzzy sets of a particular type, namely random fuzzy sets taking values in some fuzzy partition of the measurement space.

Again, X is LCHS and $USC(X)$ is the space of all usc functions from X to $[0,1]$, i.e. fuzzy subsets of X generalizing closed sets of X. The Lawson topology on it will provide a canonical σ-field for defining random fuzzy sets as random elements extending Matheron's theory of random closed sets.

Similarly to the special case of closed sets, $USC(X)$ is a complete lattice but not continuous with the point-wise order $\leq$, i.e. $f \leq g$ iff $f(x) \leq g(x), \forall x \in X$, and

$$\bigwedge_{j\in J} f_j = \inf_{j\in J} f_j, \text{ where } f_j \in USC(X),\ j \in J$$

Let $f = \inf_{j\in J} f_j$, to show $USC(X)$ is a complete lattice it is enough to show that $f \in USC(X)$, i.e. for any $r \in [0,1]$, $\{x : f(x) < r\}$ is open.

Indeed,

$$\{x : f(x) < r\} = \bigcup_{j\in J} \{x : f_j(x) < r\}$$

and since each $\{x : f_j(x) < r\}$ is open, $\{x : f(x) < r\}$ is also open.

Note 1. For any $f \in USC(X)$,

$$f = \inf_{r,K\ (\text{compact})} \{g_{r,K} : f(y) < r, \forall y \in K\}$$

where $g_{r,K}(x) = r$ if $x \in \overset{\circ}{K}$ and $= 1$ otherwise.

The proof of the following results can be found in Nguyen and Tran [13].

Theorem 1. *$L = (USC(X), \leq^{op})$ is a continuous lattice, where $f \leq^{op} g$ iff $f(x) \geq g(x), \forall x \in X$.*

Remark 1. For any $f, g \in L$, then $g \ll f$ implies $\forall x \in X, \exists r, K$ such that $x \in \overset{\circ}{K}$ and $f(y) < r \leq g(y), \forall y \in K$.

Theorem 2. *For any $r \in (0,1]$ and $K(compact) \subseteq X$, we have*

$$\{f \in L : f(y) < r,\ \forall y \in K\} = \bigcup_{\overset{\circ}{K}_i \supseteq K} \{f \in L : g_{r,K_i} \ll f\}$$

where g_{r,K_i} is defined as above.

Theorem 3. *The Scott topology $\tau(L)$ has as a subbase the sets $\{f : f(y) < r,\ \forall y \in K\}$, where $r \in (0,1]$ and $K(compact) \subseteq X$. In other words, the Scott topology $\tau(L)$ has as a base the sets*

$$\bigcap_{i=1}^{n} \{f : f(y) < r_i,\ \forall y \in K_i\}$$

where $r_i \in (0,1]$, $K_i(compact) \subseteq X$, and $n \in \mathbb{N}$.

Corollary 1. *The Lawson topology $\Lambda(L)$ has as a subbase the sets $\{f : f(y) < r, \forall y \in K\}$, where $r \in (0,1]$ and $K(compact) \subseteq X$ together with the sets $\{f : \exists x \in X$ such that $g(x) < f(x)\}$, where $g \in L$.*

Remark 2. It is known that $(L, \sigma(L))$ is second countable iff $(L, \Lambda(L))$ is second countable (Gierz et al. [3]).

Thus, to show that $(L, \Lambda(L))$ is second countable, it suffices to show that $(L, \sigma(L))$ has a countable base. See Nguyen and Tran [13] for details.

Remark 3. In view of the above results, by a *random fuzzy (closed) set* on a LCSHS space X, we mean a random element with values in the measurable space $(USC(X), \sigma(\Lambda))$, where $\sigma(\Lambda)$ is the Borel σ-field associated with the Lawson topology of the continuous lattice $USC(X)$ (with reverse order $\geq$). With the Lawson topology, $USC(X)$ is a compact, Hausdorff and second countable (hence metrizable). This falls neatly in the framework of separable metric spaces in probability theory.

7 Concluding Remarks

Empirical science is relied upon available data. Data exhibit error in a variety of forms. Measurement error in covariates is a well-known phenomenon in classical statistics. Research efforts are directed to providing robust statistical procedures in the case of low quality of data such as this. In complex systems, coarse data present several new aspects of "error" which need to be modeled for information processing. This paper focused on modeling of uncertainties in observed coarse data as well as the data themselves. Although data are typically modeled as random outcomes of phenomena, different types of uncertainty arise of which some are related to probabilistic uncertainty (belief function, possibility measures) , and others are of a different nature (e.g. fuzziness). Specifying various uncertainty measures involved and modeling of complex observed data form a firm step toward developing robust inference procedures for applications.

References

1. Cappé, O., Moulines, E., Rydén, T.: Inference in Hidden Markov Models. Springer, NY (2005)
2. Gierz, G., et al.: A Compendium of Continuous Lattices. Springer, NY (1980)
3. Gierz, G., et al.: Continuous Lattices and Domains. Cambridge University Press, Cambridge (2003)
4. Gil, M.A., Lopez-Diaz, M., Ralescu, D.A.: Overview on the development of fuzzy random variables. Fuzzy Sets and Systems 157, 2546–2557 (2006)
5. Goodman, I.R.: Fuzzy sets as equivalence classes of random sets. In: Yager, R. (ed.) Fuzzy Set and Possibility Theory, pp. 327–342. Pergamon Press, Oxford (1982)
6. Goutsias, J.: Morphological analysis of random sets. In: Goutsias, J., Mahler, R., Nguyen, H.T. (eds.) Random Sets: Theory and Applications, Springer, NY (1997)

7. Heitjan, D.F., Rubin, D.B.: Ignorability and coarse data. Annals of Statistics 19, 2244–2253 (1991)
8. Levy, H.: Stochastic Dominance: Investment Decision Making under Uncertainty, 2nd edn. Springer, NY (2006)
9. Matheron, G.: Random Sets and Integral Geometry. John Wiley, NY (1975)
10. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics, vol. 139. Springer, Heidelberg (1999)
11. Nguyen, H.T.: An Introduction to Random Sets. Chapman and Hall/CRC, Boca Raton, Florida (2006)
12. Nguyen, H.T., Wu, B.: Random and fuzzy sets in coarse data analysis. Computational Statistics and Data Analysis 51, 70–85 (2006)
13. Nguyen, H.T., Tran, H.: On a continuous lattice approach to modeling of coarse data in system analysis. Journal of Uncertainty Systems 1, 62–73 (2007)
14. Paarsch, H., Hong, H.: An Introduction to the Structural Econometrics of Auction Data. MIT Press, Cambridge (2006)
15. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, New Jersey (1976)
16. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1, 2–28 (1978)
17. Zadeh, L.A.: Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. Journal of Statistical Planning and Inference 102, 233–264 (2002)

Part II

Statistics under Interval Uncertainty and Imprecise Probability

Fast Algorithms for Computing Statistics under Interval Uncertainty: An Overview

Vladik Kreinovich[1] and Gang Xiang[1,2]

[1] University of Texas at El Paso, El Paso, TX 79968, USA
vladik@utep.edu
[2] Philips Healthcare Informatics
gxiang@acm.org

Summary. In many areas of science and engineering, it is desirable to estimate statistical characteristics (mean, variance, covariance, etc.) under interval uncertainty. For example, we may want to use the measured values $x(t)$ of a pollution level in a lake at different moments of time to estimate the average pollution level; however, we do not know the exact values $x(t)$ – e.g., if one of the measurement results is 0, this simply means that the actual (unknown) value of $x(t)$ can be anywhere between 0 and the detection limit DL. We must therefore modify the existing statistical algorithms to process such interval data.

Such a modification is also necessary to process data from statistical databases, where, in order to maintain privacy, we only keep interval ranges instead of the actual numeric data (e.g., a salary range instead of the actual salary).

Most resulting computational problems are NP-hard – which means, crudely speaking, that in general, no computationally efficient algorithm can solve all particular cases of the corresponding problem. In this paper, we overview practical situations in which computationally efficient algorithms exist: e.g., situations when measurements are very accurate, or when all the measurements are done with one (or few) instruments.

1 Computing Statistics Is Important

In many engineering applications, we are interested in computing statistics. For example, in environmental analysis, we observe a pollution level $x(t)$ in a lake at different moments of time t, and we would like to estimate standard statistical characteristics such as mean, variance, autocorrelation, correlation with other measurements.

For each of these characteristics C, there is an expression $C(x_1, \ldots, x_n)$ that enables us to provide an estimate for C based on the observed values $x_1, \ldots, x_n$. For example:

- a reasonable statistic for estimating the mean value of a probability distribution is the population average $E(x_1, \ldots, x_n) = \dfrac{1}{n} \cdot (x_1 + \ldots + x_n)$;
- a reasonable statistic for estimating the variance V is the population variance $V(x_1, \ldots, x_n) = \dfrac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E)^2$.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 19–31, 2008.
springerlink.com

Comment. The population variance is often computed by using an alternative formula $V = M - E^2$, where $M = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2$ is the population second moment.

Comment. In many practical situations, we are interested in an *unbiased* estimate of the population variance $V_u(x_1, \ldots, x_n) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - E)^2$. In this dissertation, we will describe how to estimate V under interval uncertainty; since $V_u = \frac{n}{n-1} \cdot V$, we can easily transform estimates for V into estimates for V_u.

2 Interval Uncertainty

In environmental measurements, we often only measure the values with interval uncertainty. For example, if we did not detect any pollution, the pollution value v can be anywhere between 0 and the sensor's detection limit DL. In other words, the only information that we have about v is that v belongs to the interval $[0, DL]$; we have no information about the probability of different values from this interval.

Another example: to study the effect of a pollutant on the fish, we check on the fish daily; if a fish was alive on Day 5 but dead on Day 6, then the only information about the lifetime of this fish is that it is somewhere within the interval $[5, 6]$; we have no information about the distribution of different values in this interval.

In non-destructive testing, we look for outliers as indications of possible faults. To detect an outlier, we must know the mean and standard deviation of the normal values – and these values can often only be measured with interval uncertainty; see, e.g., [38]. In other words, often, we know the result $\widetilde{x}$ of measuring the desired characteristic x, and we know the upper bound Δ on the absolute value $|\Delta x|$ of the measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ (this upper bound is provided by the manufacturer of the measuring instrument), but we have no information about the probability of different values $\Delta x \in [-\Delta, \Delta]$. In such situations, after the measurement, the only information that we have about the true value x of the measured quantity is that this value belongs to interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

In geophysics, outliers should be identified as possible locations of minerals; the importance of interval uncertainty for such applications was emphasized in [34, 35]. Detecting outliers is also important in bioinformatics [41].

In bioinformatics and bioengineering applications, we must solve systems of linear equations in which coefficients come from experts and are only known with interval uncertainty; see, e.g., [48].

In biomedical systems, statistical analysis of the data often leads to improvements in medical recommendations; however, to maintain privacy, we do not want to use the exact values of the patient's parameters. Instead, for each parameter, we select fixed values, and for each patient, we only keep the corresponding range. For example, instead of keeping the exact age, we only record

whether the age is between 0 and 10, 10 and 20, 20 and 30, etc. We must then perform statistical analysis based on such interval data; see, e.g., [23].

3 Estimating Statistics under Interval Uncertainty: A Problem

In all such cases, instead of the true values $x_1, \ldots, x_n$, we only know the intervals $\mathbf{x}_1 = [\underline{x}_1, \overline{x}_1], \ldots, \mathbf{x}_n = [\underline{x}_n, \overline{x}_n]$ that contain the (unknown) true values of the measured quantities. For different values $x_i \in \mathbf{x}_i$, we get, in general, different values of the corresponding statistical characteristic $C(x_1, \ldots, x_n)$. Since all values $x_i \in \mathbf{x}_i$ are possible, we conclude that all the values $C(x_1, \ldots, x_n)$ corresponding to $x_i \in \mathbf{x}_i$ are possible estimates for the corresponding statistical characteristic. Therefore, for the interval data $\mathbf{x}_1, \ldots, \mathbf{x}_n$, a reasonable estimate for the corresponding statistical characteristic is the range

$$C(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{C(x_1, \ldots, x_n) \,|\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

We must therefore modify the existing statistical algorithms so that they compute, or bound these ranges. This is the problem that we will be solving in this dissertation.

This problem is a part of a general problem. The above range estimation problem is a specific problem related to a combination of interval and probabilistic uncertainty. Such problems – and their potential applications – have been described, in a general context, in the monographs [30, 42]; for further developments, see, e.g., [4, 5, 6, 7, 16, 19, 32, 33, 39, 40, 43] and references therein.

4 Mean

Let us start our discussion with the simplest possible characteristic: the mean. The arithmetic average E is a monotonically increasing function of each of its n variables $x_1, \ldots, x_n$, so its smallest possible value $\underline{E}$ is attained when each value x_i is the smallest possible ($x_i = \underline{x}_i$) and its largest possible value is attained when $x_i = \overline{x}_i$ for all i. In other words, the range $\mathbf{E}$ of E is equal to $[E(\underline{x}_1, \ldots, x_n), E(\overline{x}_1, \ldots, \overline{x}_n)]$. In other words, $\underline{E} = \dfrac{1}{n} \cdot (\underline{x}_1 + \ldots + \underline{x}_n)$ and $\overline{E} = \dfrac{1}{n} \cdot (\overline{x}_1 + \ldots + \overline{x}_n)$.

5 Variance: Computing the Exact Range Is Difficult

Another widely used statistic is the variance. In contrast to the mean, the dependence of the variance V on x_i is not monotonic, so the above simple idea does not work. Rather surprisingly, it turns out that the problem of computing the exact range for the variance over interval data is, in general, NP-hard [17] which

means, crudely speaking, that the worst-case computation time grows exponentially with n. Specifically, computing the upper endpoint $\overline{V}$ of the range $[\underline{V}, \overline{V}]$ is NP-hard. Moreover, if we want to compute the variance range or $\overline{V}$ with a given accuracy ε, the problem is still NP-hard. (For a more detailed description of NP-hardness in relation to interval uncertainty, see, e.g., [22].)

6 Linearization

From the practical viewpoint, often, we may not need the exact range, we can often use approximate linearization techniques. For example, when the uncertainty comes from measurement errors Δx_i, and these errors are small, we can ignore terms that are quadratic (and of higher order) in Δx_i and get reasonable estimates for the corresponding statistical characteristics. In general, in order to estimate the range of the statistic $C(x_1, \ldots, x_n)$ on the intervals $[\underline{x}_1, \overline{x}_1], \ldots, [\underline{x}_n, \overline{x}_n]$, we expand the function C in Taylor series at the midpoint $\widetilde{x}_i \stackrel{\text{def}}{=} (\underline{x}_i + \overline{x}_i)/2$ and keep only linear terms in this expansion. As a result, we replace the original statistic with its linearized version $C_{\text{lin}}(x_1, \ldots, x_n) = C_0 - \sum_{i=1}^{n} C_i \cdot \Delta x_i$, where $C_0 \stackrel{\text{def}}{=} C(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, $C_i \stackrel{\text{def}}{=} \dfrac{\partial C}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, and $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$. For each i, when $x_i \in [\underline{x}_i, \overline{x}_i]$, the difference Δx_i can take all possible values from $-\Delta_i$ to Δ_i, where $\Delta_i \stackrel{\text{def}}{=} (\overline{x}_i - \underline{x}_i)/2$. Thus, in the linear approximation, we can estimate the range of the characteristic C as $[C_0 - \Delta, C_0 + \Delta]$, where $\Delta \stackrel{\text{def}}{=} \sum_{i=1}^{n} |C_i| \cdot \Delta_i$.

In particular, if we take, as the statistic, the population variance $C = V$, then $C_i = \dfrac{\partial V}{\partial x_i} = \dfrac{2}{n} \cdot (\widetilde{x}_i - \widetilde{E})$, where $\widetilde{E}$ is the average of the midpoints $\widetilde{x}_i$, and $C_0 = \dfrac{1}{n} \cdot \sum_{i=1}^{n} (\widetilde{x}_i - \widetilde{E})^2$ is the variance of the midpoint values $\widetilde{x}_1, \ldots, \widetilde{x}_n$. So, for the variance, $\Delta = \dfrac{2}{n} \cdot \sum_{i=1}^{n} |\widetilde{x}_i - \widetilde{E}| \cdot \Delta_i$.

It is worth mentioning that for the variance, the ignored quadratic term is equal to $\dfrac{1}{n} \cdot \sum_{i=1}^{n} (\Delta x_i)^2 - (\Delta E)^2$, where $\Delta E \stackrel{\text{def}}{=} \dfrac{1}{n} \cdot \sum_{i=1}^{n} \Delta x_i$, and therefore, can be bounded by 0 from below and by $\Delta^{(2)} \stackrel{\text{def}}{=} \dfrac{1}{n} \cdot \sum_{i=1}^{n} \Delta_i^2$ from above. Thus, the interval $[V_0 - \Delta, V_0 + \Delta + \Delta^{(2)}]$ is a guaranteed enclosure for $\mathbf{V}$.

7 Linearization Is Not Always Acceptable

In some cases, linearized estimates are not sufficient: the intervals may be wide so that quadratic terms can no longer be ignored, and/or we may be in a situation

where we want to guarantee that, e.g., the variance does not exceed a certain required threshold. In such situations, we need to get the exact range – or at least an enclosure for the exact range.

Since, even for as simple a characteristic as variance, the problem of computing its exact range is NP-hard, we cannot have a feasible-time algorithm that always computes the exact range of these characteristics. Therefore, we must look for the reasonable classes of problems for which such algorithms are possible. Let us analyze what such classes can be.

8 First Class: Narrow Intervals

The main idea behind linearization is that if the measurement errors Δx_i are small, we can safely ignore quadratic and higher order terms in Δx_i and replace the original difficult-to-analyze expression by its easier-to-analyze linear approximation. The accuracy of this techniques is determined by the size of the first term that we ignore, i.e., is of size $O(\Delta x_i^2)$. Thus, the narrower the intervals (i.e., the smaller the values Δx_i), the more accurate is the result of this linearization.

In real life, we want to compute the range with a certain accuracy. So, when the intervals are sufficiently accurate, the results of linearization estimation provide the desired accuracy and thus, we have a feasible algorithm for solving our problem. When the intervals become wider, we can no longer ignore the quadratic terms and thus, the problem becomes more computationally complex. In other words, when intervals are narrower, the problem of computing statistics under interval uncertainty becomes easier. It is therefore reasonable to consider the case of narrow intervals as the first case in which we can expect feasible algorithms for computing statistics of interval data.

How can we describe "narrowness" formally? The very fact that we are performing the statistical analysis means that we assume that the actual values $x_1, \ldots, x_n$ come from a probability distribution, and we want to find the statistical characteristics of this probability distribution. Usually, this distribution is continuous: normal, uniform, etc. Formally, a continuous distribution is a one for which a finite probability density $\rho(x)$ exists for every x. In this case, for every the real number a, the probability $p = \int_{a-\delta}^{a+\delta} \rho(x)\,dx$ to have a random value within an interval $[a-\delta, a+\delta]$ is approximately equal to $\rho(a) \cdot 2\delta$ and thus, tends to 0 as $\delta \to 0$. This means that for every value a, the probability to have a random value exactly equal to a is 0. In particular, this means that with probability 1, all the values $x_1, \ldots, x_n$ randomly selected from the original distribution are different.

The data intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$ contain these different values $x_1, \ldots, x_n$. When the intervals $\mathbf{x}_i$ surrounding the corresponding points x_i are narrow, these intervals do not intersect. When their widths becomes larger than the distance between the original values, the intervals start intersecting.

Thus, the ideal case of "narrow intervals" can be described as the case when no two intervals $\mathbf{x}_i$ intersect.

9 Second Class: Slightly Wider Intervals

Narrow intervals can be described as intervals which do not intersect at all. Namely, we have a set of (unknown) actual values $x_1 < x_2 < \ldots < x_n$, and we have intervals around each value which are so narrow that the neighboring intervals $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ do not intersect.

As the widths of the intervals increase, they start intersecting. At first, only the neighboring intervals $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ intersect, but intervals $\mathbf{x}_i$ and $\mathbf{x}_{i+2}$ still do not intersect. As the widths increase further, intervals $\mathbf{x}_i$ and $\mathbf{x}_{i+2}$ start intersecting, etc. When the intervals become very wide, all n intervals intersect.

We can therefore gauge the degree of narrowness by the number of intervals which have a common point.

Specifically, we define the case of slightly wider intervals as the situation when for some integer K, no set of K intervals has a common intersection. The case of narrow intervals correspond to $K = 2$, the next case is $K = 3$, etc. – all the way to the general case $K = n$.

As we have mentioned, the narrower the intervals, the easier the corresponding computational problem. Since the parameter K is a measure of this narrowness, it is therefore reasonable to expect that feasible algorithms exist in this case – at least for values of K which are not too large.

10 Third Class: Single Measuring Instrument

We have already mentioned that one of the most widely used engineering techniques for dealing with measurement uncertainty is linearlization. To be able to easily compute the range $\mathbf{C}$ of a statistic C by using linearization, we must make sure not only that intervals are relatively narrow, but also that they are approximately of the same size: otherwise, if, say, Δx_i^2 is of the same order as Δx_j, we cannot meaningfully ignore Δx_i^2 and retain Δx_j. In other words, the interval data set should not combine high-accurate measurement results (with narrow intervals) and low-accurate results (with wide intervals): all measurements should have been done by a single measuring instrument (or at least by several measuring instruments of the same type).

The traditional linearization techniques only provide us with an approximate range. However, as we will show, for some classes of problems, these approximate estimates can be refined into an efficient computation of the exact range. Because of this possibility, let us formulate, in precise terms, the class of problems for which linearization is possible, i.e., the class of problem for which all the measurements have been performed by a single measuring instrument.

How can we describe this class mathematically? A clear indication that we have two measuring instruments (MI) of different quality is that one interval is a proper subset of the other one: $[\underline{x}_i, \overline{x}_i] \subseteq (\underline{x}_j, \overline{x}_j)$.

This restriction only refers to not absolutely measurement results, i.e., to non-degenerate intervals. In addition to such interval values, we may also have machine-represented floating point values produced by very accurate measurements, so accurate that we can, for all practical purposes, consider these values

exactly known. From this viewpoint, when we talk about measurements made by a single measuring instrument, we may allow degenerate intervals (i.e., exact numbers) as well.

As we will see, the absence of such pairs is a useful property that enables us to compute interval statistics faster. We will also see that this absence happens not only for measurements made by a single MI, but also in several other useful practical cases. Since this property is useful, we will give it a name.

We say that a collection of intervals satisfies a *subset property* if $[\underline{x}_i, \overline{x}_i] \not\subseteq (\underline{x}_j, \overline{x}_j)$ for all i and j for which the intervals $\mathbf{x}_i$ and $\mathbf{x}_j$ are non-degenerate.

11 Fourth Class: Several MI

After the single MI case, the natural next case is when we have several (m) MI, i.e., when our intervals are divided into several subgroups each of which has the above-described subset property.

We have already mentioned that the case of a single MI is the easiest; the more MI we involve, the more complex the resulting problem – all the way to the general case $m = n$, when each measurement is performed by a different MI.

Since the parameter m is a measure of complexity, it is therefore reasonable to expect that feasible algorithms exist for the case of a fixed number m – at least for the values of m which are not too large.

12 Fifth Class: Privacy Case

In the previous text, we mainly emphasized that measurement uncertainty naturally leads to intervals. It is worth mentioning, however, that interval uncertainty may also come from other sources: e.g., from the desire to protect privacy in statistical databases. Indeed, often, we collect large amounts of data about persons – e.g., during census, or during medical experiments. Statistical analysis of this data enables us to find useful correlations between, e.g., age and effects of a certain drug, or between a geographic location and income level. Because of this usefulness, it is desirable to give researchers an ability to perform a statistical analysis of this data. However, if we simply researchers to receive answers to arbitrary queries and publish the results of their analysis, then these results may reveal the information from the databases which is private and not supposed to be disclosed.

One way to protect privacy is not to keep the exact actual values of the privacy-related quantities such as salary or age in the database. Instead, we fix a finite number of thresholds, e.g., 0, 10, 20, 30 years, and for each person, we only record the corresponding age range: from 0 to 10, or from 10 to 20, or from 20 to 30, etc. Since the actual values are not stored in the database anymore, no queries can disclose these values.

So, this idea solves the privacy problem, but it opens up another problem: how can perform statistical processing on this privacy-related interval data? Suppose that we are interested in the values of a statistical characteristic $C(x_1, \ldots, x_n)$.

If we knew the actual values $x_1, \ldots, x_n$, then we could easily compute the value of this characteristic. However, in case of privacy-related interval uncertainty, all we know is intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values of x_i. Different values $x_i \in \mathbf{x}_i$ lead, in general, to different values of $C(x_1, \ldots, x_n)$. So, a reasonable idea is to return the range of possible values of the characteristic $C(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i$.

From the algorithmic viewpoint, we get the same problem as with measurement-related interval uncertainty: find the range of the given characteristic $C(x_1, \ldots, x_n)$ on given intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The *difference* between this case and the two previous cases is that, in the first two cases, we *do not know the exact values*, while in this case, in principle, it is *possible to get the exact value*, but we do not use the exact values, because we want to protect privacy.

From the mathematical viewpoint, privacy-related intervals have the following property: they either coincide (if the value corresponding to the two patients belongs to the same range) or are different, in which case they intersect in at most point. Similarly to the above situation, we also allow exact values in addition to ranges; these values correspond, e.g., to the exact records made in the past, records that are already in the public domain.

We will call interval data with this property – that every two non-degenerate intervals either coincide or intersect in at most one point – *privacy case.*

Comment. For the privacy case, the subset property is satisfied, so algorithms that work for the subset property case work for the privacy case as well.

Comment. Sometimes, in the privacy-motivated situation, we must process interval data in which intervals come from several different "granulation" schemes. For example, to find the average salary in North America, we may combine US interval records in which the salary is from 0 to 10,000 US dollars, from 10,000 to 20,000, etc., with the Canadian interval records in which the ranges are between 0 to 10,000 Canadian dollars, 10,000 to 20,000 Canadian dollars, etc. When we transform these records to a single unit, we get two different families of intervals, each of which satisfies the subset property. Thus, to handle such situations, we can use algorithms developed for the several MI case.

13 Sixth Class: Non-detects

An important practical case is the case of non-detects. Namely, many sensors are reasonably accurate, but they have a detection limit DL – so they cannot detect any value below DL but they detect values of DL and higher with a very good accuracy.

In this case, if a sensor returns a value $\widetilde{x} \geq DL$, then this value is reasonably accurate, so we can consider it exact (i.e., a degenerate interval $[\widetilde{x}, \widetilde{x}]$). However, if the sensor does not return any signal at all, i.e., the measurement result $\widetilde{x} = 0$, then the only thing we can conclude about the actual value of the quantity is that this value is below the detection limit, i.e., that it lies in the interval $[0, DL]$.

In this case, every interval is either an exact value or a *non-detect*, i.e., an interval $[0, DL_i]$ for some real number DL_i (with possibly different detection limits for different sensors). Under this assumption, the resulting non-degenerate intervals also satisfy the subset property. Thus, algorithms that work for the subset property case work for this "non-detects" case as well.

Also, an algorithm that works for the general privacy case also works for the non-detects case when all sensors have the same detection limit DL.

14 Results

The main results are summarized in the following table:

Table 1. Computational complexity of statistical analysis under interval uncertainty: an overview

Case	E	V	L, U	S
Narrow intervals	$O(n)$	$O(n)$	$O(n \cdot \log(n))$	$O(n^2)$
Slightly wider narrow intervals	$O(n)$	$O(n \cdot \log(n))$	$O(n \cdot \log(n))$	?
Single MI	$O(n)$	$O(n)$	$O(n \cdot \log(n))$	$O(n^2)$
Several (m) MI	$O(n)$	$O(n^m)$	$O(n^m)$	$O(n^{2m})$
New case	$O(n)$	$O(n^m)$	?	?
Privacy case	$O(n)$	$O(n)$	$O(n \cdot \log(n))$	$O(n^2)$
Non-detects	$O(n)$	$O(n)$	$O(n \cdot \log(n))$	$O(n^2)$
General	$O(n)$	NP-hard	NP-hard	?

Here, E is a population mean, V is a population variance, $S \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - E)^3$ is the population skewness, and $L \stackrel{\text{def}}{=} E - k_0 \cdot \sigma$ and $U \stackrel{\text{def}}{=} E + k_0 \cdot \sigma$ are endpoints of the confidence interval, where a parameter k_0 is usually taken as $k_0 = 2$, $k_0 = 3$, or $k_0 = 6$.

Comment. For descriptions of the algorithms, and for proofs of the algorithm correctness, see [18, 46] and references therein; see also [1, 3, 12, 13, 14, 20, 21, 23, 24, 25, 26, 27, 28, 29, 31, 44, 45, 47].

Applications. There are several application areas in which it is possible to take into account interval uncertainty in statistical data processing:

- the seismic inverse problem in geophysics [2],
- the problem of estimating and decreasing the clock cycle in computer chips [36, 37],
- the problem of separating the core from the fragments in radar data processing [15], and
- the problem of inverse half-toning in image processing [11].

15 Conclusion

In many areas of science and engineering, it is desirable to estimate statistical characteristics (mean, variance, covariance, etc.) under interval uncertainty. Such a modification is necessary, e.g., to process data from statistical databases, where, in order to maintain privacy, we only keep interval ranges instead of the actual numeric data (e.g., a salary range instead of the actual salary).

Most resulting computational problems are NP-hard – which means, crudely speaking, that in general, no computationally efficient algorithm can solve all particular cases of the corresponding problem.

In this paper, we overview practical situations in which computationally efficient algorithms exist: e.g., situations when measurements are very accurate, or when all the measurements are done with one (or few) instruments.

Acknowledgments

This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Max Planck Institut für Mathematik, and by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08.

The authors are thankful to Dr. Van Nam Huynh for valuable suggestions.

References

1. Alo, R., Beheshti, M., Xiang, G.: Computing variance under interval uncertainty: A new algorithm and its potential application to privacy in statistical databases. In: IPMU 2006. Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, France, July 2–7, pp. 810–816 (2006)
2. Averill, M.G., et al.: Using expert knowledge in solving the seismic inverse problem. International Journal of Approximate Reasoning 45(3), 564–587 (2007)
3. Beck, J., Kreinovich, V., Wu, B.: Interval-valued and fuzzy-valued random variables: From computing sample variances to computing sample covariances. In: Lopez, M., et al. (eds.) Soft Methodology and Random Information Systems, pp. 85–92. Springer, Heidelberg (2004)
4. Berleant, D.: Automatically verified arithmetic with both intervals and probability density functions. Interval Computations (2), 48–70 (1993)
5. Berleant, D.: Automatically verified arithmetic on probability distributions and intervals. In: Kearfott, R.B., Kreinovich, V. (eds.) Applications of Interval Computations, Kluwer, Dordrecht (1996)
6. Berleant, D., Goodman-Strauss, C.: Bounding the results of arithmetic operations on random variables of unknown dependency using intervals. Reliable Computing 4(2), 147–165 (1998)
7. Berleant, D., et al.: Dependable handling of uncertainty. Reliable Computing 9(6), 407–418 (2003)
8. Berleant, D., Xie, L., Zhang, J.: Statool: A Tool for distribution envelope determination (DEnv), an interval-based algorithm for arithmetic on random variables. Reliable Computing 9(2), 91–108 (2003)

9. Berleant, D., Zhang, J.: Using Pearson correlation to improve envelopes around the distributions of functions. Reliable Computing 10(2), 139–161 (2004)
10. Berleant, D., Zhang, J.: Representation and problem solving with the distribution envelope determination (DEnv) method. Reliability Engineering and System Safety 85(1–3) (July–September 2004)
11. Cabrera, S.D., et al.: On inverse halftoning: Computational complexity and interval computations. In: CISS 2005. Proceedings of the 39th Conference on Information Sciences and Systems, March 16–18, p. 164. John Hopkins University (2005)
12. Ceberio, M., et al.: How to take into account dependence between the inputs: From interval computations to constraint-related set computations, with potential applications to nuclear safety, bio- and geosciences. In: Proceedings of the Second International Workshop on Reliable Engineering Computing, Savannah, Georgia, Febuary 22–24, pp. 127–154 (2006)
13. Dantsin, E., et al.: Population variance under interval uncertainty: A new algorithm. Reliable Computing 12(4), 273–280 (2006)
14. Dantsin, E., et al.: Detecting outliers under interval uncertainty: A new algorithm based on constraint satisfaction. In: IPMU 2006. Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, France, July 2–7, pp. 802–809 (2006)
15. Debroux, P., et al.: Using 1-D radar observations to detect a space explosion core among the explosion fragments: Sequential and distributed algorithms. In: Proceedings of the 11th IEEE Digital Signal Processing Workshop, August 1–4, pp. 273–277. Taos, New Mexico (2004)
16. Ferson, S.: RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers. CRC Press, Boca Raton, Florida (2002)
17. Ferson, S., et al.: Exact bounds on finite populations of interval data. Reliable Computing 11(3), 207–233 (2005)
18. Ferson, S., et al.: Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty, Sandia National Laboratories, Report SAND2007-0939 (May 2007)
19. Ferson, S., Myers, D., Berleant, D.: Distribution-Free Risk Analysis: I. Range, Mean, and Variance, Applied Biomathematics, Technical Report (2001)
20. Granvilliers, L., Kreinovich, V., Müller, N.: Novel approaches to numerical software with result verification. In: Alt, R., et al. (eds.) Dagstuhl Seminar 2003. LNCS, vol. 2991, pp. 274–305. Springer, Heidelberg (2004)
21. Kreinovich, V.: Probabilities, intervals, what next? optimization problems related to extension of interval computations to situations with partial information about probabilities. Journal of Global Optimization 29(3), 265–280 (2004)
22. Kreinovich, V., et al.: Computational Complexity and Feasibility of Data Processing and Interval Computations. Kluwer, Dordrecht (1997)
23. Kreinovich, V., Longpré, L.: Computational complexity and feasibility of data processing and interval computations, with extension to cases when we have partial information about probabilities. In: Brattka, V., et al. (eds.) CCA 2003. Proc. Conf. on Computability and Complexity in Analysis, Cincinnati, Ohio, USA, August 28–30, pp. 19–54 (2003)
24. Kreinovich, V., et al.: Outlier detection under interval uncertainty: Algorithmic solvability and computational complexity. Reliable Computing 11(1), 59–76 (2005)
25. Kreinovich, V., et al.: Interval versions of statistical techniques, with applications to environmental analysis, bioinformatics, and privacy in statistical databases. Journal of Computational and Applied Mathematics 199(2), 418–423 (2007)

26. Kreinovich, V., Nguyen, H.T., Wu, B.: On-line algorithms for computing mean and variance of interval data, and their use in intelligent systems. Information Sciences 177(16), 3228–3238 (2007)
27. Kreinovich, V., et al.: Probabilities, intervals, what next? extension of interval computations to situations with partial information about probabilities. In: Proceedings of the 10th IMEKO TC7 International Symposium on Advances of Measurement Science, St. Petersburg, Russia, June 30–July 2, vol. 1, pp. 137–142 (2004)
28. Kreinovich, V., Xiang, G., Ferson, S.: Efficient algorithms for computing mean and variance under dempster-shafer uncertainty. International Journal of Approximate Reasoning 42, 212–227 (2006)
29. Kreinovich, V., et al.: Towards combining probabilistic and interval uncertainty in engineering calculations: Algorithms for computing statistics under interval uncertainty, and their computational complexity. Reliable Computing 12(6), 471–501 (2006)
30. Kuznetsov, V.P.: Interval Statistical Models, Radio i Svyaz, Moscow (in Russian) (1991)
31. Langewisch, A.T., Choobineh, F.F.: Mean and variance bounds and propagation for ill-specified random variables. IEEE Transactions on Systems, Man, and Cybernetics, Part A 34(4), 494–506 (2004)
32. Lodwick, W.A., Jamison, K.D.: Estimating and validating the cumulative distribution of a function of random variables: Toward the development of distribution arithmetic. Reliable Computing 9(2), 127–141 (2003)
33. Moore, R.E., Lodwick, W.A.: Interval analysis and fuzzy set theory. Fuzzy Sets and Systems 135(1), 5–9 (2003)
34. Nivlet, P., Fournier, F., Royer, J.: A new methodology to account for uncertainties in 4-D seismic interpretation. In: SEG 2001. Proc. 71st Annual Int'l Meeting of Soc. of Exploratory Geophysics, San Antonio, TX, September 9–14, 2001, pp. 1644–1647 (2001)
35. Nivlet, P., Fournier, F., Royer, J.: Propagating interval uncertainties in supervised pattern recognition for reservoir characterization. In: SPE 2001. Proc. 2001 Society of Petroleum Engineers Annual Conf., September 30–October 3, New Orleans, LA, paper SPE-71327 (2001)
36. Orshansky, M., et al.: Interval-based robust statistical techniques for non-negative convex functions, with application to timing analysis of computer chips. In: SAC 2006. Proceedings of the ACM Symposium on Applied Computing, Dijon, France, April 23–27, pp. 1629–1633 (2006)
37. Orshansky, M., et al.: Interval-based robust statistical techniques for non-negative convex functions with application to timing analysis of computer chips. In: Proceedings of the Second International Workshop on Reliable Engineering Computing, Savannah, Georgia, February 22–24, pp. 197–212 (2006)
38. Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice. Springer, New York (2005)
39. Regan, H., Ferson, S., Berleant, D.: Equivalence of five methods for bounding uncertainty. Journal of Approximate Reasoning 36(1), 1–30 (2004)
40. Rowe, N.C.: Absolute bounds on the mean and standard deviation of transformed data for constant-sign-derivative transformations. SIAM Journal of Scientific Statistical Computing 9, 1098–1113 (1988)
41. Shmulevich, I., Zhang, W.: Binary analysis and optimization-based normalization of gene expression data. Bioinformatics 18(4), 555–565 (2002)

42. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman & Hall, NY (1991)
43. Williamson, R., Downs, T.: Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. International Journal of Approximate Reasoning 4, 89–158 (1990)
44. Xiang, G., et al.: New Algorithms for statistical analysis of interval data. In: Dongarra, J., Madsen, K., Wasniewski, J. (eds.) PARA 2004. LNCS, vol. 3732, pp. 189–196. Springer, Heidelberg (2005)
45. Xiang, G.: Fast algorithm for computing the upper endpoint of sample variance for interval data: Case of sufficiently accurate measurements. Reliable Computing 12(1), 59–64 (2006)
46. Xiang, G.: Fast Algorithms for Computing Statistics under Interval Uncertainty, with Applications to Computer Science and to Electrical and Computer Engineering, PhD Dissertation, Computer Science Dept., University of Texas at El Paso (July 2007)
47. Xiang, G., Kosheleva, O., Klir, G.J.: Estimating information amount under interval uncertainty: Algorithmic solvability and computational complexity. In: IPMU 2006. Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, France, July 2–7, pp. 840–847 (2006)
48. Zhang, W., Shmulevich, I., Astola, J.: Microarray Quality Control. Wiley, Hoboken, New Jersey (2004)

Trade-Off between Sample Size and Accuracy: Case of Static Measurements under Interval Uncertainty

Hung T. Nguyen[1] and Vladik Kreinovich[2]

[1] New Mexico State University, Las Cruces, NM 88003, USA
hunguyen@nmsu.edu
[2] University of Texas at El Paso, El Paso, TX 79968, USA
vladik@utep.edu

Summary. In many practical situations, we are not satisfied with the accuracy of the existing measurements. There are two possible ways to improve the measurement accuracy:

- first, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements;
- second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., a procedure that includes the use of a higher quality reagent.

In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument. What is the appropriate trade-off between sample size and accuracy? This is the general problem that we address in this paper.

1 General Formulation of the Problem

We often need more accurate measurement procedures. Measurements are never 100% accurate, there is always a measurement inaccuracy.

Manufacturers of a measuring instrument usually provide the information about the accuracy of the corresponding measurements. In some practical situations, however, we want to know the value of the measured quantity with the accuracy which is higher than the guaranteed accuracy of a single measurement.

Comment. Measurements are provided either by a *measuring instrument* or, in situations like measuring level of pollutants in a given water sample, by a *measuring lab*. Most problems related to measurement accuracy are the same, whether we have an automatic device (measuring instrument) or operator-supervised procedure (measuring lab). In view of this similarity, in the following text, we will consider the term "measuring instrument" in the general sense, so

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 32–44, 2008.
springerlink.com

that the measuring lab is viewed as a particular case of such (general) measuring instrument.

Two ways to improve the measurement accuracy: increasing sample size and improving accuracy. There are two possible ways to improve the measurement accuracy:

- first, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements;
- second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., the procedure that includes the use of a higher quality reagent.

In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument.

Problem: finding the best trade-off between sample size and accuracy. What guidance shall we give to an engineer in this situation? Shall she make repeated measurements with the original instrument? shall she instead purchase a more accurate measuring instrument and make repeated measurements with this new instrument? How more accurate? how many measurement should we perform? In other words, what is the appropriate trade-off between sample size and accuracy?

This is the general problem that we address in this paper.

2 In Different Practical Situations, This General Problem Can Take Different Forms

There are two different situations which, crudely speaking, correspond to engineering and to science.

In most practical situations – in *engineering*, *ecology*, etc. – we know what accuracy we want to achieve. In *engineering*, this accuracy comes, e.g., from the tolerance with which we need to guarantee some parameters of the manufactured object. To make sure that these parameters fit into the tolerance intervals, we must measure them with the accuracy that is as good as the tolerance. For example, if we want to guarantee, e.g., the resistance of a certain wire does not deviate from its nominal value by more than 3%, then we must measure this resistance with an accuracy of at least 3% (or better).

In *ecological* measurements, we want to make sure that the measured quantity does not exceed the required limit. For example, if we want to guarantee that the concentration of a pollutant does not exceed 0.1 units, then we must be able to measure this concentration with an accuracy somewhat higher than 0.1. In such situations, our objective is to minimize the cost of achieving this accuracy.

In *science*, we often face a different objective:

- we have a certain amount of funding allocated for measuring the value of a certain quantity;
- within the given funding limits, we would like to determine the value of the measured quantity as accurately as possible.

In other words:

- In engineering situations, we have a fixed accuracy, and we want to minimize the measurement cost.
- In scientific situations, we have a fixed cost, and we want to maximally improve the measurement accuracy.

3 A Realistic Formulation of the Trade-Off Problem

Traditional engineering approach. The traditional engineering approach to solving the above problem is based on the following assumptions – often made when processing uncertainty in engineering:

- that all the measurement errors are normally (Gaussian) distributed known standard deviations σ;
- that the measurement errors corresponding to different measurement are independent random variables; and
- that the mean value Δ_s of the measurement error is 0.

Under these assumptions, if we repeat a measurement n times and compute the arithmetic average of n results, then this average approximates the actual value with a standard deviation $\dfrac{\sigma}{\sqrt{n}}$. So, under the above assumptions, by selecting appropriate large number of iterations n, we can get make measurement errors as small as we want.

This approach – and more general statistical approach – has been actively used in many applications to science in engineering problems; see, e.g., [1, 2, 6, 8].

Limitations of the traditional approach. In practice, the distributions are often Gaussian and independent; however, the mean (= systematic error) Δ_s is not necessarily 0. Let us show this if we do not take systematic error into account, we will underestimate the resulting measurement inaccuracy.

Indeed, suppose that we have a measuring instrument about which we know that its measurement error cannot exceed 0.1: $|\Delta x| \le 0.1$. This means, e.g., that if, as a result of the measurement, we got the value $\widetilde{x} = 1.0$, then the actual (unknown) value x $(= \widetilde{x} - \Delta x)$ of the measured quantity can take any value from the interval $[1.0 - 0.1, 1.0 + 0.1] = [0.9, 1.1]$.

If the mean of the measurement error (i.e., the systematic error component) is 0, then we can repeat the measurement many times and, as a result, get more and more accurate estimates of x. However, if – as is often the case – we do not have any information about the systematic error, it is quite possible that

the systematic error is actually equal to 0.07 (and the random error is negligible in comparison with this systematic error). In this case, the measured value 1.0 means that the actual value of the measured quantity was $x = 1.0 - 0.07 = 0.93$. In this case, we can repeat the measurement many times, and every time, the measurement result will be equal to $\approx x + \Delta_s = 0.93 + 0.01 = 1.0$. The average of these values will still be approximately equal to 1.0 – so, no matter how many times we repeat the measurement, we will get the exact same measurement error 0.07.

In other words, when we are looking for a trade-off between sample size and accuracy, the traditional engineering assumptions can result in misleading conclusions.

A more realistic description of measurement errors. We do not know the actual value of the systematic error Δ_s – if we knew this value, we could simply re-calibrate the measuring instrument and thus eliminate this systematic error.

What we do know are the bounds on the systematic error. Specifically, in measurement standards (see, e.g., [7]), we are usually provided with the upper bound Δ on the systematic error – i.e., with a value Δ for which $|\Delta_s| \le \Delta$. In other words, the only information that we have about the systematic error Δ_s is that it belongs to the *interval* $[-\Delta, \Delta]$.

Resulting formulas for the measurement accuracy. Under these assumptions, what is the guaranteed accuracy of a single measurement made by the measuring instrument?

Although formally, a normally distributed random variable can take any value from $-\infty$ to $+\infty$, in reality, the probability of value which are too far away from the average is practically negligible. In practice, it is usually assumed that the values which differ from the average a by more than $k_0 \cdot \sigma$ are impossible – where the value k_0 is determined by how confident we want to be:

- 95% confidence corresponds to $k_0 = 2$,
- 99.9% corresponds to $k_0 = 3$, and
- confidence $100\% - 10^{-6}\%$ corresponds to $k_0 = 6$.

Thus, with selected confidence, we know that the measurement error is between $\Delta_s - k_0 \cdot \sigma$ and $\Delta_s + k_0 \cdot \sigma$. Since the systematic error Δ_s can take any value from $-\Delta$ to $+\Delta$, the smallest possible value of the overall error is $-\Delta - k_0 \cdot \sigma$, and the largest possible value of the overall error is $\Delta + k_0 \cdot \sigma$.

Thus, for a measuring instrument with a standard deviation σ of the random error component and a upper bound Δ on the systematic error component, the overall error is bounded by the value $\Delta + k_0 \cdot \sigma$, where the value k_0 is determined by the desired confidence level.

Resulting formulas for the accuracy of a repeated measurement. When we repeat the same measurement n times and take the average of n measurement results, the systematic error remains the same, while the standard deviation of

the random error decreases $\sqrt{n}$ times. Thus, after n measurements, the overall error is bounded by the value $\Delta + k_0 \cdot \dfrac{\sigma}{\sqrt{n}}$.

So, we arrive at the following formulation of the trade-off problem.

Trade-off problem for engineering. In the situation when we know the overall accuracy Δ_0, and we want to minimize the cost of the resulting measurement, the trade-off problem takes the following form:

$$\text{Minimize } n \cdot F(\Delta, \sigma) \text{ under the constraint } \Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} \le \Delta_0, \tag{1}$$

where $F(\Delta, \sigma)$ is the cost of a single measurement performed by a measuring instrument whose systematic error is bounded by Δ and whose random error has a standard deviation σ.

Trade-off problem for science. In the situation when we are given the limit F_0 on the cost, and the problem is to achieve the highest possible accuracy within this cost, we arrive at the following problem

$$\text{Minimize } \Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} \text{ under the constraint } n \cdot F(\Delta, \sigma) \le F_0. \tag{2}$$

4 Solving the Trade-Off Problem in the General Case

Mathematical comment. The number of measurement n is a discrete variable. In general, optimization with respect to discrete variables requires much more computations than continuous optimization (see, e.g., [4]). Since our formulation is approximate anyway, we will treat n as a real-valued variable – with the idea that in a practical implementation, we should take, as the actual sample size, the closest integer to the corresponding real number solution n_{opt}.

Towards resulting formulas. For both constraint optimization problems, the Lagrange multiplier method leads to the following unconstraint optimization problem:

$$n \cdot F(\Delta, \sigma) + \lambda \cdot \left(\Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} - \Delta_0 \right) \to \min_{\Delta, \sigma, n}, \tag{3}$$

where λ can be determined by one of the formulas

$$\Delta + k_0 \cdot \frac{\sigma}{\sqrt{n}} = \Delta_0, \quad n \cdot F(\Delta, \sigma) = F_0. \tag{4}$$

Equating the derivatives of the objective function (with respect to the unknowns Δ, σ, and n) to 0, we conclude that

$$n \cdot \frac{\partial F}{\partial \Delta} + \lambda = 0; \quad n \cdot \frac{\partial F}{\partial \sigma} + \lambda \cdot \frac{k_0}{\sqrt{n}} = 0; \quad F - \frac{1}{2} \cdot \lambda \cdot k_0 \cdot \frac{\sigma}{n^{3/2}} = 0. \tag{5}$$

Substituting the expression for λ from the first equation into the second one, we conclude that

$$n = k_0^2 \cdot \frac{(\partial F/\partial \Delta)^2}{(\partial F/\partial \sigma)^2}. \tag{6}$$

Substituting these expression into the other equations from (5) and into the equations (4), we get the following non-linear equations with two unknowns Δ and σ:

$$F + \frac{1}{2} \cdot \sigma \cdot \frac{\partial F}{\partial \sigma} = 0; \tag{7}$$

$$\Delta + \frac{\sigma \cdot (\partial F/\partial \sigma)}{\partial F/\partial \Delta} = \Delta_0; \quad k_0^2 \cdot \frac{(\partial F/\partial \Delta)^2}{(\partial F/\partial \sigma)^2} \cdot F = F_0. \tag{8}$$

So, we arrive at the following algorithm:

General formulas: results. For each of the optimization problems (1) and (2), to find the optimal accuracy values Δ and σ and the optimal sample size n, we do the following:

- First, we determine the optimal accuracy, i.e., the optimal values of Δ and σ, by solving a system of two non-linear equations with two unknowns Δ and σ: the equation (7) and one of the equations (8) (depending on what problem we are solving).
- After that, we determine the optimal sample size n by using the formula (6).

For practical engineering problems, we need more explicit and easy-to-use recommendations. The above formulas provide a general theoretical solution to the trade-off problem, but to use them in practice, we need more easy-to-use recommendations. In practice, however, we do not have the explicit formula $F(\Delta, \sigma)$ that determines how the cost of the measurement depends on its accuracy. Therefore, to make our recommendations more practically useful, we must also provide some guidance on how to determine this dependence – and then use the recommended dependence to simply the above recommendations.

5 How Does the Cost of a Measurement Depend on Its Accuracy?

Two characteristics of uncertainty: Δ and σ. In our description, we use two parameters to characterize the measurement's accuracy: the upper bound Δ on the systematic error component and the standard deviation σ of the random error component.

It is difficult to describe how the cost of a measurement depends on σ. The standard deviation σ is determined by the noise level, so decreasing σ requires a serious re-design of the measuring instrument. For example, to get a standard measuring instrument, one thing designers usually do is place the

instrument in liquid helium so as to eliminate the thermal noise as much as possible; another idea is to place the measuring instrument into a metal cage, to eliminate the effect of the outside electromagnetic fields on the measuring instrument's electronics.

Once we have eliminated the obvious sources of noise, eliminating a new source of noise is a creative problem, requiring a lot of ingenuity, and it is difficult to estimate how the cost of such decrease depends on σ.

The inability to easily describe the dependence of cost on σ may not be that crucial. The inability to easily handle the characteristic σ of the random error component may not be so bad because, as we have mentioned, the random error component is the one that can be drastically decreased by increasing the sample size – in full accordance with the traditionally used simplifying engineering assumptions about uncertainty.

As we have mentioned, in terms of decreasing the overall accuracy, it is much more important to decrease the systematic error component, i.e., to decrease the value Δ. Let us therefore analyze how the cost of a measurement depends on Δ.

How we can reduce Δ: reminder. As we have mentioned, we can decrease the characteristic Δ of the systematic error component by calibrating our measuring instrument against the standard one.

After N repeated measurements, we get a systematic error Δ_s whose standard deviation is $\approx \sigma/\sqrt{N}$ (and whose distribution, due to the Central Limit Theorem, is close to Gaussian). Thus, with the same confidence level as we use to bound the overall measurement error, we can conclude that $|\Delta_s| \le k_0 \cdot \sigma/\sqrt{N}$.

Calibration is not a one-time procedure. To properly take calibration into account, it is important to recall that calibration is not a one-time procedure. Indeed, most devices deteriorate with time. In particular, measuring instruments, if not periodically maintained, become less and less accurate. Because of this, in measurement practices, calibration is not a one-time procedure, it needs to be done periodically.

How frequently do we need to calibrate a device? The change of Δ_s with time t is slow and smooth. A smooth dependence can be represented by a Taylor series $\Delta_s(t) = \Delta_s(0) + k \cdot t + c \cdot t^2 + \ldots$ In the first approximation, we can restrict ourselves to the main – linear – term (linear trend) in this expansion, and thus, in effect, assume that the change of Δ_s with time t is linear.

Thus, if by calibrating the instrument, we guaranteed that $|\Delta_s| \le \Delta$, then after time t, we can only guarantee that $|\Delta_s| + k \cdot t \le \Delta$. Once the upper bound on Δ_s reaches the level that we want not to exceed, this means that a new calibration is in order. Usually (see, e.g., [7]), to guarantee the bound Δ throughout the entire calibration cycle, we, e.g., initially calibrate it to be below $\Delta/2$, and then re-calibrate at a time t_0 when $\Delta/2 + k \cdot t_0 = \Delta$. In such a situation, the time t_0 between calibrations is equal to $t_0 = \Delta/(2 \cdot k)$.

How the calibration-based reduction procedure translates into the cost of a measurement: the main case. As we have just mentioned, the way to decrease Δ is to calibrate the measuring instrument. Thus, the resulting additional cost of a measurement comes from the cost of this calibration (spread over all the measurement performed between calibrations).

Each calibration procedure consists of two stages:

- first, we transport the measuring instrument to the location of a standard – e.g., to the National Institute of Standard and Technology (NIST) or one of the regional standardization centers – and set up the comparison measurements by the tested and the standard instruments;
- second, the we perform the measurements themselves.

Correspondingly, the cost of calibration can be estimated as the sum of the costs of there two stages.

The standard measuring instrument is usually a very expensive operation. So, setting it up for comparison with different measuring instruments requires a lot of time and a lot of adjustment. Once the set-up is done, the second stage is fast and automatic – and therefore not that expensive.

As a result, usually, the cost of the first stage is the dominating factor. So, we can reasonably assume that the cost of the calibration is just the cost of the set-up – i.e., the cost of the first stage of the calibration procedure.

By definition, the set-up does not depend on how many times N we perform the comparison measurements. Thus, in the first approximation, we can simply assume that each calibration requires a flat rate f_0.

The interval between time calibrations is $t_0 = \Delta/(2 \cdot k)$, then during a fixed period of time T_0 (e.g., 10 years), we need

$$\frac{T_0}{t_0} = \frac{T_0}{\Delta/(2 \cdot k)} = \frac{2 \cdot k \cdot T_0}{\Delta}$$

calibrations. Multiplying this number by the cost f_0 of each calibration, we get the overall cost of all the calibrations performed during the fixed time T_0 as $\dfrac{2 \cdot k \cdot T_0 \cdot f_0}{\Delta}$. Finally, dividing this cost by the estimated number N_0 of measurements performed during the period of time T_0, we estimate the cost $F(\Delta)$ of an individual measurement as

$$F(\Delta) = \frac{c}{\Delta}, \tag{9}$$

where we denoted

$$c \stackrel{\text{def}}{=} \frac{2 \cdot k \cdot T_0 \cdot f_0}{N_0}. \tag{10}$$

Comment. The above formula was first described, in a somewhat simplified form, in [3].

This formula is in good accordance with chemistry-related measurements. It is worth mentioning that the dependence $c \sim 1/\Delta$ also occurs in

measurements related to chemical analysis. Indeed, in these measurements, the accuracy of the measurement result is largely determined by the quality of the reagents, i.e., mainly, by the concentration level δ of the unwanted chemicals (pollutants) in a reagent mix. Specifically, the maximum possible error Δ is proportional to this concentration δ, i.e., $\Delta \approx c_0 \cdot \delta$.

According to [9], the cost of reducing pollutants to a level δ is proportional to $1/\delta$. Since the accuracy Δ is proportional to δ, the dependence of the cost of the accuracy is also inverse proportional to Δ, i.e., $F(\Delta) = c/\Delta$ for some constant c.

This formula is in good accordance with actual prices of different measurements. This dependence is in good agreement by the experimental data on the cost of measurements of chemical-related measurements. For example, in a typical pollution measurement, a measurement with the 25% accuracy costs $\approx$ \$200, while if we want to get 7% accuracy, then we have to use a better reagent grade in our measurements which costs between \$500 and \$1,000. Here, the 3–4 times increase in accuracy (i.e., 3–4 times decrease in measurement error) leads to approximately the same (4–5) times increase in cost – which is indeed in good accordance with the dependence $F(\Delta) \approx c/\Delta$.

How the calibration-based reduction procedure translates into the cost of a measurement: cases of more accurate measurements. In deriving the formula $F(\Delta) \approx c/\Delta$, we assumed that the cost of actually performing the measurements with the standard instrument is much smaller than the cost of setting up the calibration experiment. This is a reasonable assumption if the overall number of calibration-related measurement N is not too large.

How many measurement do we need? After N measurements, we get the accuracy $\Delta = k_0 \cdot \sigma/\sqrt{N}$. Thus, for a measuring instrument with standard deviation σ, if we want to achieve the systematic error level Δ, we must use

$$N = k_0 \cdot \frac{\sigma^2}{\Delta^2} \tag{11}$$

measurements.

So, if we want to use the calibration procedure to achieve higher and higher accuracy – i.e., smaller and smaller values of Δ – we need to perform more and more calibration-related measurements. For large N, the duration of the calibration-related measurements exceeds the duration of the set-up. Since the most expensive part of the calibration procedure is the use of the standard measuring instrument, the cost of this procedure is proportional to the overall time during which we use this instrument. When N is large, this time is roughly proportional to N.

In this case, instead of a flat fee f_0, the cost of each calibration becomes proportional to N, i.e., equal to $f_1 \cdot N$, where f_1 is the cost per time of using the standard measuring instrument multiplied by the time of each calibration measurement. Due to the formula (11), the resulting cost of each calibration is equal to $f_1 \cdot k_0 \cdot \dfrac{\sigma^2}{\Delta^2}$. To get the cost of a single measurement, we must multiply

this cost by the number of calibrations $\frac{2 \cdot k \cdot T_0}{\Delta}$ required during the time period T_0, and then divide by the typical number of measurements performed during this period of time. As a result, the cost of a single measurement becomes $\frac{\text{const}}{\Delta^3}$.

The cost of measurements beyond calibration: general discussion. In many scientific cutting-edge experiments, we want to achieve higher accuracy than was possible before. In such situations, we cannot simply use the existing standard measuring instrument to calibrate the new one, because we want to achieve the accuracy that no standard measuring instrument has achieved earlier.

In this case, how we can increase the accuracy depends on the specific quantity that we want to measure.

The cost of measurements beyond calibration: example. For example, in radioastrometry – the art of determining the locations of celestial objects from radioastronomical observation – the accuracy of a measurement by a single radio telescope is $\Delta \approx \lambda/D$, where λ is the wavelength of the radio-waves on which we are observing the source, and D is the diameter of the telescope; see, e.g., [10]. For a telescope of a linear size D, just the amount of material is proportional to its volume, i.e., to D^3; the cost F of designing a telescope is even higher – it is proportional to D^4. Since $D \approx \text{const}/\Delta$, in this case, we have $F(\Delta) \approx \text{const}/\Delta^4$.

The cost of measurements beyond calibration: power laws. The above dependence is a particular case of the *power law* $F(\Delta) \approx \text{const}/\Delta^\alpha$. Power laws are, actually, rather typical descriptions of the dependence of the cost of an individual measurement on its accuracy.

In [5], we explain why in the general case, power laws are indeed reasonable approximation: crudely speaking, in the absence of a preferred value of the measured quantity, it is reasonable to assume that the dependence does not change if we change the measuring unit (i.e., that it is scale invariant), and power laws are the only scale-invariant dependencies.

Comment. The same arguments about scale invariance apply when we try to find out how the cost of a measurement depends on the standard deviation. So, it is reasonable to assume that this dependence is also described by a power law $F(\sigma) \approx \text{const}/\sigma^\beta$ for some constant β.

6 Trade-Off between Accuracy and Sample Size in Different Cost Models

Let us plug in the above cost models into the above general solution for the tradeoff problem and find out what is the optimal trade-off between accuracy and sample size in the above cost models.

Since the above cost models only describe the dependence of the cost of Δ and n, we will assume that the characteristic σ of the random error component is fixed, so we can only select the accuracy characteristic Δ and the sample size n.

Basic cost model: engineering situation. Let us start with the basic cost model, according to which $F(\Delta) = c/\Delta$. Within this model, we can explicitly solve the above system of equations. As a result, for the engineering situation, we conclude that

$$n_{\text{opt}} = \frac{9 \cdot k_0^2 \cdot \sigma^2}{4 \cdot \Delta_0^2}; \quad \Delta_{\text{opt}} = \frac{1}{3} \cdot \Delta_0. \tag{12}$$

Observation. In this case, the overall error bound Δ_0 is the sum of the bounds coming from two error components:

- the bound Δ_0 that comes from the systematic error component, and
- the bound $k_0 \cdot \dfrac{\sigma}{\sqrt{n}}$ that comes from the random error component.

In the optimal trade-off, the first component is equal to 1/3 of the overall error bound, and therefore, the second component is equal to 2/3 of the overall error bound. As a result, we conclude that when the error comes from several error components, in the optimal trade-off, these error components are of approximately the same size.

Heuristic consequence of this observation. As a result of this qualitative idea, it is reasonable to use the following heuristic rule when looking for a good (not necessarily optimal) trade-off: split the overall error into equal parts.

In the above example, this would mean taking $\Delta = (1/2) \cdot \Delta_0$ (and, correspondingly, $k_0 \cdot \dfrac{\sigma}{\sqrt{n}} = (1/2) \cdot \Delta_0$) instead of the optimal value $\Delta = (1/3) \cdot \Delta_0$. How non-optimal is this heuristic solution?

For the optimal solution $\Delta = (1/3) \cdot \Delta_0$, the resulting value of the objective function (1) (representing the overall measurement cost) is $\dfrac{27}{4} \cdot \dfrac{k_0^2 \cdot \sigma^2 \cdot c}{\Delta_0^2}$, while for $\Delta = (1/2) \cdot \Delta_0$, the cost is $8 \cdot \dfrac{k_0^2 \cdot \sigma^2 \cdot c}{\Delta_0^2}$ – only $\approx 20\%$ larger.

If we take into account that all our models are approximate, this means that the heuristic trade-off solution is practically as good as the optimal one.

Basic cost model: science situation. In the science situation (2), we get

$$n_{\text{opt}} = \left(\frac{F_0 \cdot k_0 \cdot \sigma}{2 \cdot c} \right)^{2/3}; \quad \Delta_{\text{opt}} = \frac{n_{\text{opt}} \cdot c}{F_0}. \tag{13}$$

Cases of more accurate and cutting-edge measurements. When $F(\Delta) = c/\Delta^\alpha$, for the engineering case, we get

$$n_{\text{opt}} = \frac{(\alpha+2)^2 \cdot k_0^2 \cdot \sigma^2}{4 \cdot \Delta_0^2}; \quad \Delta_0 = \frac{\alpha}{2+\alpha} \cdot \Delta_0.$$

For the science case,

$$n_{\text{opt}} = \left(\frac{F_0}{c} \right)^{2/(2+\alpha)} \cdot \left(\frac{k_0 \cdot \alpha}{2} \right)^{(2\alpha)/(2+\alpha)}; \quad \Delta_{\text{opt}} = \frac{\alpha}{2} \cdot k_0 \cdot \frac{\sigma}{\sqrt{n_{\text{opt}}}}.$$

In both cases, the error bound coming from the systematic error component is approximately equal to the error bound coming from the random error component.

7 Conclusion

In many practical situations, we are not satisfied with the accuracy of the existing measurements. There are two possible ways to improve the measurement accuracy. First, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements. Second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., a procedure that includes the use of a higher quality reagent. In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument.

What is the appropriate trade-off between sample size and accuracy? Traditional engineering approach to this problem assumes that we know the exact probability distribution of all the measurement errors. In many practical situations, however, we do not know the exact distributions. For example, we often only know the upper bound on the corresponding measurement (or estimation) error; in this case, after the measurements, we only know the interval of possible values of the quantity of interest. In the first part of this paper, we show in such situations, traditional engineering approach can sometimes be misleading, so for interval uncertainty, new techniques are needed. In the remainder of this paper, we describe proper techniques for achieving optimal trade-off between sample size and accuracy under interval uncertainty.

Acknowledgments. This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Max Planck Institut für Mathematik, and by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08.

The authors are very thankful to the anonymous referees for valuable suggestions.

References

1. Eng, J.: Sample Size Estimation: How Many Individuals Should Be Studied? Radiology 227, 309–313 (2003)
2. Hox, J.J.: Multilevel Analysis: Techniques and Applications. Lawrence Erlbaum Associates, Mahwah (2002)
3. Kreinovich, V.: How to compute the price of a measuring instrument? Leningrad Center for New Information Technology "Informatika", Technical Report, Leningrad (in Russian) (1989)

4. Kreinovich, V., et al.: Computational complexity and feasibility of data processing and interval computations. Kluwer, Dordrecht (1997)
5. Nguyen, H.T., Kreinovich, V.: Applications of continuous mathematics to computer science. Kluwer, Dordrecht (1997)
6. Obuchowski, N.A., Mcclish, D.K.: Sample Size Determination For Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices. Statistics in Medicine 16(13), 1529–1542 (1997)
7. Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice. Springer, New York (2005)
8. Skidmore, A.K.: Accuracy assessment of spatial information. In: Spatial Statistics for Remote Sensing, pp. 197–209. Springer, The Netherlands (2002)
9. Stevens, D.: Analysis of biological systems. In: Proceedings of the NIH MARC Winter Institute on Undergraduate Education in Biology, Santa Cruz, California, January 7–11, (2005)
10. Verschuur, G.L., Kellerman, K.I.: Galactic and Extragalactic Radio Astronomy. Springer, NY (1988)

Trade-Off between Sample Size and Accuracy: Case of Dynamic Measurements under Interval Uncertainty

Hung T. Nguyen[1], Olga Kosheleva[2], Vladik Kreinovich[2], and Scott Ferson[3]

[1] New Mexico State University, Las Cruces, NM 88003, USA
hunguyen@nmsu.edu
[2] University of Texas at El Paso, El Paso, TX 79968, USA
vladik@utep.edu
[3] Applied Biomathematics, Setauket, NY 11733, USA
scott@ramas.com

Summary. In many practical situations, we are not satisfied with the accuracy of the existing measurements. There are two possible ways to improve the measurement accuracy:

- first, instead of a *single* measurement, we can make *repeated* measurements; the additional information coming from these additional measurements can improve the accuracy of the result of this series of measurements;
- second, we can replace the *current* measuring instrument with a *more accurate* one; correspondingly, we can use a more accurate (and more expensive) measurement procedure provided by a measuring lab – e.g., a procedure that includes the use of a higher quality reagent.

In general, we can combine these two ways, and make *repeated* measurements with a *more accurate* measuring instrument. What is the appropriate trade-off between sample size and accuracy? In our previous paper, we solved this problem for the case of static measurements. In this paper, we extend the results to the case of dynamic measurements.

1 Formulation of the Problem

In some practical situations, we want to know the value of the measured quantity with the accuracy which is higher than the guaranteed accuracy of a single measurement. There are two possible ways to improve the measurement accuracy:

- first, instead of a *single* measurement, we can make several (n) measurements;
- second, we can replace the *current* measuring instrument with a *more accurate* one.

What is the appropriate trade-off between sample size and accuracy? In our previous paper [11], we analyzed this problem for the case when we measure a

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 45–56, 2008.
springerlink.com

static quantity, i.e., a quantity which does not change with time. In this paper, we extend the results from [11] to the general case of *dynamic* measurements, when the measured quantity changes over time.

For such dynamic quantities, we may have two different objectives:

- We may be interested in knowing the *average* value of the measured quantity, e.g., the average concentration of a pollutant in a lake or the average day temperature. In addition to to knowing the average, we may also want to know the standard deviation and/or other statistical characteristics.
- We may also want to know not only the average, but also the actual dependence of the measured quantity on space location and/or time.

For example:

- If we are interested in general weather patterns, e.g., as a part of the climatological analysis, then it is probably sufficient to measure the average temperature (or the average wind velocity) in a given area.
- On the other hand, if our intent is to provide the meteorological data to the planes flying in this area, then we would rather know how exactly the wind velocity depends on the location, so that the plane will be able to avoid locations where the winds are too strong.

In this paper, we analyze the trade-off between accuracy and sample size for both objectives.

2 First Objective: Measuring the Average Value of a Varying Quantity

Case of ideal measuring instruments: analysis. Let us start to analyze this situation with the case of an ideal measuring instrument, i.e., a measuring instrument for which the measurement errors are negligible.

By using this ideal instrument, we can measure the value of the quantity of interest at different points and at different moments of time. After we perform n measurements and get n measurement results $x_1, \ldots, x_n$, a natural way to estimate the desired mean value $x_0 = E[x]$ of x is to use the arithmetic average $E \stackrel{\text{def}}{=} \dfrac{x_1 + \ldots + x_n}{n}$ of these measured values. It is reasonable to assume that the differences $x_i - x_0$ are independent random variables, with a known standard deviation σ_0.

In this case, due to the Central Limit Theorem, for large n, the difference $\Delta x_0 \stackrel{\text{def}}{=} E - x_0$ between the estimate E and the desired value x_0 is approximately normally distributed with 0 average and standard deviation $\sigma_0/\sqrt{n}$.

So, even for measurements with the ideal measuring instrument, the result E of measuring x_0 is not exact; we can only guarantee (with the corresponding level of confidence) that the measurement error Δx_0 is bounded by the value $k_0 \cdot \sigma_0/\sqrt{n}$.

Comment. If we do not know this standard deviation, we can estimate it based on the measurement results $x_1, \ldots, x_n$, by using the standard statistical formulas, such as

$$\sigma_0 \approx \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - E)^2}.$$

Case of ideal measuring instruments: recommendations. In the case of ideal measuring instruments, if we want to achieve the desired overall accuracy Δ_0 with a given confidence, then the sample size n must be determined by the condition that $k_0 \cdot \sigma_0 / \sqrt{n} \le \Delta_0$, where k_0 corresponds to this confidence:

- 95% confidence corresponds to $k_0 = 2$,
- 99.9% corresponds to $k_0 = 3$, and
- confidence $1 - 10^{-6}\%$ corresponds to $k_0 = 6$.

The above condition is equivalent to $\sqrt{n} \ge \dfrac{k_0 \cdot \sigma_0}{\Delta_0}$, i.e., to $n \ge \dfrac{k_0^2 \cdot \sigma_0^2}{\Delta_0^2}$. To minimize the measurement costs, we must select the smallest sample size for which this inequality holds, i.e., select $n \approx \dfrac{k_0^2 \cdot \sigma_0^2}{\Delta_0^2}$.

Case of realistic measuring instruments: description. In practice, measuring instruments are not perfect, they have measurement errors. Usually, we assume that we know the standard deviation σ of the corresponding measurement error, and we know the upper bound Δ on the possible values of the mean (systematic) error $\Delta - s$: $|\Delta_s| \le \Delta$; see, e.g., [14].

Case of realistic measuring instruments: analysis. For realistic measuring instruments, for each measurement, the difference $\Delta x_i = \widetilde{x}_i - x_i$ between the measured and actual values of the quantity of interest is no longer negligible.

In this case, based on n measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$, we do not get the arithmetic average E of the *actual* values, we only get the average

$$\widetilde{E} = \frac{\widetilde{x}_1 + \ldots + \widetilde{x}_n}{n}$$

of the *measured* values. We are using this average $\widetilde{E}$ as an estimate for the desired average x_0. There are two reasons why $\widetilde{E}$ is different from x_0:

- first, due to measurement errors, $\widetilde{x}_i \ne x_i$, hence $\widetilde{E} \ne E$;
- second, due to the finite sample size, $E \ne x_0$.

As a result, the error Δx_0 with which this procedure measures x_0, i.e., the difference $\Delta x_0 \stackrel{\text{def}}{=} \widetilde{E} - x_0$, can be represented as the sum of two error components:

$$\widetilde{E} - x_0 = (\widetilde{E} - E) + (E - x_0). \tag{1}$$

If we use a measuring instrument whose mean (systematic) error is Δ_s and standard deviation is σ, then for the difference of arithmetic averages, the mean is the same value Δ_s (systematic error) and the standard deviation is $\sqrt{n}$ times smaller: it is equal to $\sigma/\sqrt{n}$. We have just described the difference $E - x_0$: it is a random variable with 0 mean and standard deviation $\sigma_0/\sqrt{n}$.

Since the mean value of $E - x_0$ is 0 (by definition of x_0 as the mean of x_i), the mean value of the sum (11) is equal to the mean value of the first error component, i.e., to Δ_s.

It is reasonable to assume that the measurement errors $\widetilde{x}_i - x_i$ (caused by the imperfections of the measurement procedure) and the deviations $x_i - x_0$ (caused by variability of the quantity of interest) are independent random variables. In this case, the variance of the sum (11) is equal to the sum of the corresponding variances, i.e., to

$$\frac{\sigma^2}{n} + \frac{\sigma_0^2}{n} = \frac{\sigma_t^2}{n},$$

where we denoted $\sigma_t \stackrel{\text{def}}{=} \sqrt{\sigma^2 + \sigma_0^2}$. Hence, the standard deviation of the total error is equal to $\sigma_t/\sqrt{n}$.

So, the measurement error $\widetilde{E} - x_0$ is approximately normally distributed, with the mean Δ_s (about which we know that $|\Delta_s| \le \Delta$) and the standard deviation $\sigma_t/\sqrt{n}$. Thus, we can conclude that with a selected degree of confidence, the overall error cannot exceed $\Delta + k_0 \cdot \dfrac{\sigma_t}{\sqrt{n}}$.

Case of realistic measuring instruments: recommendations. From the purely mathematical viewpoint, when the standard deviation σ of a measuring instrument is fixed, then, to determine Δ and n, we get exactly the same formulas as in the case of static measurements, with the only difference that:

- instead of the standard deviation σ of the random error component of the measuring instrument,
- we now have the combined standard deviation $\sigma_t = \sqrt{\sigma^2 + \sigma_0^2}$ of the measuring instrument and of the measured quantity.

So, all the recommendations that we have developed in [11] for static measurements are also applicable here.

Example. If we want to achieve a given accuracy Δ_0 with the smallest possible cost, then, according to [11], we should use the measuring instrument with accuracy $\Delta \approx (1/3) \cdot \Delta_0$. The sample size n is then determined by the formula $k_0 \cdot \dfrac{\sigma_t}{\sqrt{n}} = (2/3) \cdot \Delta_0$.

For measuring average, the optimal accuracy Δ if the same as for static measurements, but the optimal sample size is now determined by a new formula $n_{\text{opt}} = \dfrac{9 \cdot k_0^2 \cdot \sigma_t^2}{4 \cdot \Delta_0^2}$, with σ_t instead of σ. Since $\sigma_t > \sigma$, we will need a larger sample size n.

3 Second Objective: Measuring the Actual Dependence of the Measured Quantity on Space Location and/or on Time

Formulation of the problem. In many real-life situations, we are interested not only in the average value of the measured quantity x, we are also interested in the actual dependence of this quantity on space and/or time.

Within this general scheme, there are several possible situations:

- We may have a quantity that does not depend on a spatial location but does depend on time – e.g., we may be interested in the temperature at a given location. In this case, we are interested to learn how this quantity x depends on the time t, i.e., we are interested to know the dependence $x(t)$.
- We may be interested in a quantity that does not change with time but does change from one spatial location to the other. For example:
  - in a geographic analysis, we may be interested in how the elevation x depends on the 2-D spatial location $t = (t_1, t_2)$;
  - in a geophysical analysis, we may be interested how in the density depends on a 3-D location $t = (t_1, t_2, t_3)$ inside the Earth.
- Finally, we may be interested in a quantity that changes both with time and from one spatial location to the other. For example:
  - we may be interested in learning how the surface temperature depends on time t_1 and on the 2-D spatial location (t_2, t_3);
  - we may be also interested in learning how the general temperature in the atmosphere depends on time t_1 and on the 3-D spatial location (t_2, t_3, t_4).

In all these cases, we are interested to know the dependence $x(t)$ of a measured quantity on the point $t = (t_1, \ldots, t_d)$ in d-dimensional space, where the dimension d ranges from 1 (for the case when we have a quantity depending on time) to 4 (for the case when we are interested in the dependence both on time and on the 3-D spatial location).

Measurement inaccuracy caused by the finiteness of the sample. In practice, we can only measure the values of x at finitely many different locations, and we must use extrapolation to find the values at other locations. So, even if we use a perfect measuring instrument, for which the measurement error can be ignored, we still have an error cause by extrapolation.

For example, suppose that we have measured the values $x(t^{(i)})$ of the quantity x at moments of time $t^{(1)} < t^{(2)} < \ldots, < t^{(n)}$, and we want to describe the value $x(t)$ of this quantity at a different moment of time $t \neq t^{(i)}$, a moment of time at which no measurement has been made.

In practice, for most systems, we know the limit g on how fast the value of the quantity x can change with time (or from one spatial location to the other). So, when, e.g., $t^{(1)} < t < t^{(2)}$, we can conclude that $|x(t) - x(t^{(1)})| \leq g \cdot |t - t^{(1)}|$, i.e., that $x(t) \in [x(t^{(1)}) - g \cdot |t - t^{(1)}|, x(t^{(1)}) + g \cdot |t - t^{(1)}|]$. Thus, even when we have an ideal measuring instrument, the fact that we only have a finite sample $t^{(1)}, \ldots, t^{(n)}$ leads to uncertainty in our knowledge of the values $x(t)$ for $t \neq x^{(i)}$.

Estimate of the measurement uncertainty for a given measurement accuracy and given sample size. Let us consider a general situation when we perform measurements with a guaranteed accuracy Δ, and when we measure the quantity x at n different points $t^{(1)}, \ldots, t^{(n)}$ in the d-dimensional space. As a result of this measurement, we get n values $\widetilde{x}_i$ that are Δ-close to the actual values of the quantity x at the corresponding point $t^{(i)}$: $|\widetilde{x}_i - x(t^{(i)})| \le \Delta$.

If we are interested in the value $x(t)$ of the quantity x at a point $t \ne t^{(i)}$, then we have to use one of the measured values $\widetilde{x}_i$.

We assume that we know the rate g with which $x(t)$ changes with t. Thus, if we use the the result $\widetilde{x}_i$ of measuring $x(t^{(i)})$ to estimate $x(t)$, we can guarantee that $|x(t^{(i)}) - x(t)| \le g \cdot \rho(t, t^{(i)})$, where $\rho(a, b)$ denotes the distance between the two points in the d-dimensional space. Since $|\widetilde{x}_i - x(t^{(i)})| \le \Delta$, we can thus conclude that $|\widetilde{x}_i - x(t)| \le |\widetilde{x}_i - x(t^{(i)})| + |x(t^{(i)}) - x(t)| \le \Delta + g \cdot \rho(t, t^{(i)})$, i.e.,

$$|\widetilde{x}_i - x(t)| \le \Delta + g \cdot \rho(t, t^{(i)}). \tag{2}$$

Thus, the smaller the distance between t and $t^{(i)}$, the smaller the resulting error. So, to get the most accurate estimate for $x(t)$, we must select, for this estimate, the point $t^{(i)}$ which is the closest to t.

In general, once we fix the accuracy Δ, the sample size n, and the points $t^{(1)}, \ldots, t^{(n)}$ at which the measurement are performed, we can guarantee that for every t, the value $x(t)$ can be reconstructed with the accuracy $\Delta + g \cdot \rho_0$, where ρ_0 is the largest possible distance between a point t and the sample set $\{t^{(1)}, \ldots, t^{(n)}\}$.

Thus, once we fixed Δ and n, we should select the points $t^{(i)}$ in such a way that this "largest distance" ρ_0 attains the smallest possible value.

In the 1-D case, the corresponding allocation is easy to describe. Indeed, suppose that we want to allocate such points $t^{(i)}$ on the interval $[0, T]$. We want to minimize the distance ρ_0 corresponding to a given sample size n – or, equivalently, to minimize the sample size given a distance ρ_0. Every point t is ρ_0-close to one of the sample points $t^{(i)}$, so it belongs to the corresponding interval

$$[t^{(i)} - \rho_0, t^{(i)} + \rho_0].$$

Thus, the interval $[0, T]$ of width T is covered by the union of n intervals $[t^{(i)} - \rho_0, t^{(i)} + \rho_0]$ of widths $2\rho_0$. The width T of the covered interval cannot exceed the sum of the widths of the covering intervals, so we have $T \le n \cdot (2\rho_0)$, hence always $\rho_0 \ge T/(2n)$. Actually, we can have $\rho_0 = T/2n$ if we select the points $t^{(i)} = (i - 0.5) \cdot (T/n)$. Then:

- for the values $t \in [0, T/n]$, we take, as the estimate for $x(t)$, the result $\widetilde{x}_1$ of measuring $x(t^{(1)}) = x(T/(2n))$;
- for the values $t \in [T/n, 2T/n]$, we take, as the estimate for $x(t)$, the result $\widetilde{x}_2$ of measuring $x(t^{(2)}) = x((3/2) \cdot (T/n))$;
- ...

- for the values $t \in [(i-1)\cdot T/n, i\cdot T/n]$, we take, as the estimate for $x(t)$, the result $\widetilde{x}_i$ of measuring $x(t^{(i)}) = x((i-1/2)\cdot(T/n))$;
- ...

So, the optimal location of points is when they are on a grid $t^{(1)} = 0.5\cdot T/n$, $t^{(2)} = 1.5\cdot T/n$, $t^{(3)} = 2.5\cdot T/n$, ..., and each point $t^{(i)}$ "serves" the values t from the corresponding interval $[(i-1)\cdot T/n, i\cdot T/n]$ (the interval that contains this point $t^{(i)}$ as its center), serves in the sense that for each point t from this interval, as the measured value of $x(t)$, we take the value $x^{(i)}$. These intervals corresponding to individual points $t^{(i)}$ cover the entire interval $[0,T]$ without intersection,

In this optimal location, when we perform n measurements, we get $\rho_0 = T/(2n)$.

Similarly, in the general d-dimensional case, we can place n points on a d-dimensional grid. In this case, each point $t^{(i)}$ "serves" the corresponding cube; these cubes cover the whole domain without intersection. If we denote, by V, the d-dimensional volume of the spatial (or spatio-temporal) domain that we want to cover, then we can conclude that each point $x^{(i)}$ serves the cube of volume V/n. Since the volume of a d-dimensional cube of linear size Δt is equal to $(\Delta t)^d$, we can thus conclude that the linear size of each of the cubes serves by a measurement point is $(V/n)^{1/d}$.

Within this cube, each point $t^{(i)}$ is located at the center of the corresponding cube. Thus, for each point t within this cube and for each coordinate j, the absolute value $|t_j - t_j^{(i)}|$ between the j-th coordinate of this point t and the j-th coordinate of the cube's center $t^{(i)}$ does not exceed one half of the cube's linear size: $|t_j - t_j^{(i)}| \le (1/2)\cdot(V/n)^{1/d}$. Therefore, for

$$\rho(t, t^{(i)}) = \sqrt{\left(t_1 - t_1^{(i)}\right)^2 + \ldots + \left(t_d - t_d^{(i)}\right)^2},$$

we get

$$\rho(t, t^{(i)}) \le \rho \stackrel{\text{def}}{=} \sqrt{d\cdot\left(\frac{1}{2}\cdot\left(\frac{V}{n}\right)^{1/d}\right)^2} = \sqrt{d}\cdot\frac{1}{2}\cdot\frac{V^{1/d}}{n^{1/d}}.$$

We have already mentioned that for every point t, the accuracy with which we can reconstruct $x(t)$ is bounded by the value $\Delta + g\cdot\rho_0$. Thus, this accuracy is bounded by $\Delta + g\cdot\sqrt{d}\cdot\frac{1}{2}\cdot\frac{V^{1/d}}{n^{1/d}}$.

We are now ready to formally describe the corresponding trade-off problems.

Trade-off problems for engineering and science: formulation. In engineering applications, we know the overall accuracy Δ_0, and we want to minimize the cost of the resulting measurement. In this case, the trade-off problem takes the following form:

$$\text{Minimize } n\cdot F(\Delta) \to \min_{\Delta, n} \text{ under the constraint } \Delta + \frac{g_0}{n^{1/d}} = \Delta_0, \qquad (3)$$

where $F(\Delta)$ is a cost of a single measurement made by a measuring instrument with accuracy Δ, and we denoted

$$g_0 \stackrel{\text{def}}{=} g \cdot \sqrt{d} \cdot \frac{1}{2} \cdot V^{1/d}. \tag{4}$$

In scientific applications, when we are given the cost F_0, and the problem is to achieve the highest possible accuracy within this cost. In this case, we arrive at the following problem

$$\text{Minimize } \Delta + \frac{g_0}{n^{1/d}} \to \min_{\Delta, n} \text{ under the constraint } n \cdot F(\Delta) = F_0. \tag{5}$$

Engineering situation: solution. For the basic cost model $F(\Delta) = c/\Delta$ [11], the engineering problem (3) has the following solution:

$$\Delta_{\text{opt}} = \frac{1}{d+1} \cdot \Delta_0; \quad n_{\text{opt}} = \left(\frac{g_0}{\Delta_0} \cdot \frac{d+1}{d} \right)^d. \tag{6}$$

Similarly to the static case [11], the optimal trade-off between accuracy and the sample size is attained when both error components are of approximately the same size.

Science situation: solution. For the basic cost model $F(\Delta) = c/\Delta$, the science problem (3) has the following solution:

$$n_{\text{opt}} = \left(\frac{F_0}{c} \cdot \frac{g_0}{d} \right)^{d/(d+1)}; \quad \Delta_{\text{opt}} = \frac{n_{\text{opt}} \cdot c}{F_0}. \tag{7}$$

In this case too, in the optimal trade-off, the error bound coming from the accuracy of individual measurements is approximately equal to the error bound coming from the finiteness of the sample.

Case of non-smooth processes: how to describe them. In the above text, we considered the case the dependence of the quantity x on time and/or space t is smooth. In this case, for small changes Δt, this dependence can be approximately described by a linear function $x(t + \Delta t) = x(t) + g_1 \cdot \Delta t_1 + \ldots + g_d \cdot \Delta t_d$. So, if we know the upper bound g on the length $\|(g_1, \ldots, g_d)\|$ of the gradient of $x(t)$, we can bound the difference $x(t + \Delta t) - x(t)$ between the values of the quantity x at close points $t + \Delta t$ and t by the product $g \cdot \|\Delta t\| = g \cdot \rho(t, t + \Delta t)$.

In practice, we often encounter non-smooth processes. For example, meteorological data exhibit random change (similar to the Brownian motion); as the result of this, the dependence of the corresponding quantities x on time and spatial coordinates is not smooth.

For the particular case of a Brownian motion, the difference between the values of the quantity x at nearly points grows as the square root of the distance between these points: $|x(t + \Delta t) - x(t)| \le C \cdot \|\Delta t\|^{1/2}$ for some real number

C. In many physical processes, this dependence can be described by a more general power law, i.e., $|x(t+\Delta t)-x(t)| \le C \cdot \|\Delta t\|^{\beta}$ for some real numbers C and $\beta \in (0,1)$. Such processes are a particular case of *fractals*; see, e.g., [9] (This notion is closely related with the notion of a fractal dimension: namely, the graph of the corresponding dependence $x(t)$ has a fractal dimension $d+(1-\beta)$.)

In [10], it is explained why scale invariance naturally leads to the power law – and thus, to the fractal dependence.

Measurement errors in the case of non-smooth processes. Let us use these formulas to estimate measurement errors for the case of non-smooth processes. We have already mentioned that if we perform (appropriately located) n measurements in a d-dimensional space, then the distance from each point t of the domain of interest to one of the points $t^{(i)}$ in which the measurement was made does not exceed $\rho_0 = \sqrt{d} \cdot \dfrac{1}{2} \cdot \dfrac{V^{1/d}}{n^{1/d}}$.

In the fractal case, we can conclude that the error of approximating the desired value $x(t)$ with the measured value $x(t^{(i)})$ does not exceed $C \cdot \rho^{\beta}$. Thus, if we perform n measurements with a measuring device of accuracy Δ, the resulting accuracy in reconstructing all the values of $x(t)$ is bounded by the value

$$\Delta + C \cdot \rho_0^{\beta} = \Delta + C \cdot d^{\beta/2} \cdot \frac{1}{2^{\beta}} \cdot \frac{V^{\beta/d}}{n^{\beta/d}} = \Delta + \frac{g_{\beta}}{n^{\beta/d}},$$

where we denoted

$$g_{\beta} \stackrel{\text{def}}{=} C \cdot d^{\beta/2} \cdot \frac{1}{2^{\beta/d}} \cdot V^{\beta/d}.$$

Trade-off problems for engineering and science: formulation and solution. In the situation when we know the overall accuracy Δ_0, and we want to minimize the cost of the resulting measurement, the trade-off problem takes the following form:

$$\text{Minimize } n \cdot F(\Delta) \text{ under the constraint } \Delta + \frac{g_{\beta}}{n^{\beta/d}} = \Delta_0. \tag{8}$$

In the situation when we are given the limit F_0 on the cost, and the problem is to achieve the highest possible accuracy within this cost, we arrive at the following problem

$$\text{Minimize } \Delta + \frac{g_{\beta}}{n^{\beta/d}} \text{ under the constraint } n \cdot F(\Delta) = F_0. \tag{9}$$

From the mathematical viewpoint, these formulas are similar to the formulas corresponding to the smooth case, with the only difference that instead of raising n to the power $1/d$, we now raise n to the power $1/d'$, where $d' \stackrel{\text{def}}{=} d/\beta$.

Thus, for the basic cost model $F(\Delta) = c/\Delta$ [11], the engineering problem has the following solution:

$$\Delta_{\text{opt}} = \frac{\beta}{d+\beta} \cdot \Delta_0; \quad n_{\text{opt}} = \left(\frac{g_{\beta}}{\Delta_0} \cdot \frac{d+\beta}{d} \right)^{d}. \tag{10}$$

For the basic cost model $F(\Delta) = c/\Delta$, the science problem has the following solution:

$$n_{\text{opt}} = \left(\frac{F_0}{c} \cdot \frac{g_\beta}{d}\right)^{d/(d+\beta)}; \quad \Delta_{\text{opt}} = \frac{n_{\text{opt}} \cdot c}{F_0}. \tag{11}$$

in this case too, in the optimal trade-off, both error components are of approximately the same value.

Case of more accurate measuring instruments. In [11], we have shown that for more accurate measuring instrument, the cost $F(\Delta)$ of a measurement depends on its accuracy as $F(\Delta) = c/\Delta^\alpha$. Once we go beyond the basic cost model $\alpha = 1$, we get $\alpha = 3$, and then, as we increase accuracy, we switch to a different value α.

For such a power law, in the engineering case, the optimal accuracy is $\Delta_{\text{opt}} = \frac{\alpha}{\alpha+2} \cdot \Delta_0$. In particular, for $\alpha = 3$, we have $\Delta_{\text{opt}} = \frac{3}{5} \cdot \Delta_0$.

4 Case Study: In Brief

A real-life example in which we used similar arguments to made a selection between the accuracy and the sample size is the design of radioastronomical telescope system [1, 2, 3, 4, 5, 7, 8]. As we have mentioned, for the radiotelescope of diameter D, the measurement accuracy is proportional to λ/D, and the cost is proportional to D^4.

The design of a large system of radiotelescopes has several objectives:

- first, we would like to solve *radioastrometry* problems, i.e., determine the location of the radiosources with as much accuracy as possible;
- second, we would like to solve the *radioimaging* problems, i.e., for each of the radiosources, we would like to know not only its location, but also its *image* – i.e., how the intensity (and polarization) of the source changes from one point of this source to the other.

In the first problem, we are interested in measuring a well-defined unchanging quantity. In the second problem, we are interested in finding the actual dependence of the measured quantity on the spatial location.

In the second problem, similar to what we discussed in the general case, the more samples we take (i.e., the more telescopes we build), the more points we will get on the image. On the other hand, within a given overall cost, if we build more telescopes, then the amount of money allocated to each telescope will be smaller, so each telescope will be small ($D' \ll D$), and the resulting accuracy $\Delta \sim 1/D$ of each of the many measurements will be not so good.

In our analysis, we have found an optimal trade-off between accuracy and sample size. This analysis was used in the design of the successful Russian network of radiotelescopes.

5 Conclusions

In general, if the measurement error consists of several components, then the optimal trade-off between the accuracy Δ and the same size n occurs when these components are approximately of the same size.

In particular, if we want to achieve the overall accuracy Δ_0, as a first approximation, it is reasonable to take $\Delta = \Delta_0/2$ – and select the sample size for which the resulting overall error is Δ_0.

A more accurate description of optimal selections in different situations is as follows:

- for the case when we measure a single well-defined quantity (or the average value of varying quantity), we should take $\Delta = \dfrac{1}{3} \cdot \Delta_0$;
- for the case when we are interested in reconstructing all the values $x(t)$ of a smooth quantity x depending on d parameters $t = (t_1, \ldots, t_d)$, we should take $\Delta = \dfrac{1}{d+1} \cdot \Delta_0$;
- for the case when are interested in reconstructing all the values $x(t)$ of a non-smooth quantity x depending on d parameters $t = (t_1, \ldots, t_d)$, we should take $\Delta = \dfrac{\beta}{d+\beta} \cdot \Delta_0$, where β is the exponent of the power law that describes how the difference $x(t + \Delta t) - x(t)$ changes with $\|\Delta t\|$.

For the case of more accurate measuring instruments, when the cost $F(\Delta)$ of a single measurement starts growing as c/Δ^3, we should take $\Delta = \dfrac{3}{5} \cdot \Delta_0$. In general, if $F(\Delta) = c/\Delta^\alpha$, we should take $\Delta = \dfrac{\alpha}{\alpha+2} \cdot \Delta_0$.

Acknowledgments. This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Max Planck Institut für Mathematik, and by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08.

The authors are thankful to the anonymous referees for valuable suggestions.

References

1. Dravskikh, A., Finkelstein, A.M., Kreinovich, V.: Astrometric and geodetic applications of VLBI 'arc method'. In: Modern Astrometry, Proceedings of the IAU Colloquium, Vienna, vol. 48, pp. 143–153 (1978)
2. Dravskikh, A.F., et al.: Possibility of using a reference-object method to form a phase-stable multielement long-baseline interferometric system. In: Bulletin of the Special Astrophysical Observatory – North Caucasus, vol. 16, pp. 72–80. Allerton Press, NY (1984)
3. Dravskikh, A.F., et al.: The method of arcs and differential astrometry. Soviet Astronomy Letters 5(3), 160–162 (1979)

4. Dravskikh, A.F., et al.: Optimization of the procedure for measuring arcs by radiointerferometry. Soviet Astronomy Letters 5(4), 227–228 (1979)
5. Kreinovich, V., et al.: Interval estimates for closure phase and closure amplitude imaging in radio astronomy. Interval Computations 2(4), 51–71 (1992)
6. Kreinovich, V., et al.: Computational complexity and feasibility of data processing and interval computations. Kluwer, Dordrecht (1997)
7. Kreinovich, V., et al.: Open-ended configurations of radio telescopes: A geometrical analysis. Geombinatorics 13(2), 79–85 (2003)
8. Kreinovich, V., et al.: Open-ended configurations of radio telescopes: Towards optimal design. In: Proceedings of the 2002 World Automation Congress WAC 2002, Orlando, Florida, June 9–13, 2002, pp. 101–106 (2002)
9. Mandelbrot, B.B.: The fractal geometry of Nature. Freeman, San Francisco (1982)
10. Nguyen, H.T., Kreinovich, V.: Applications of continuous mathematics to computer science. Kluwer, Dordrecht (1997)
11. Nguyen, H.T., Kreinovich, V.: Trade-off between sample size and accuracy: Case of dynamic measurements (in these proceedings)
12. Novitskii, P.V., Zograph, I.A.: Estimating the Measurement Errors. Energoatomizdat, Leningrad (in Russian) (1991)
13. Orlov, A.I.: How often are the observations normal? Industrial Laboratory 57(7), 770–772 (1991)
14. Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice. Springer, New York (2005)
15. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall/CRC, Boca Raton, Florida (2004)
16. Stevens, D.: Analysis of biological systems. In: Proceedings of the NIH MARC Winter Institute on Undergraduate Education in Biology, January 7–11, 2005, Santa Cruz, California (2005)
17. Verschuur, G.L., Kellerman, K.I.: Galactic and Extragalactic Radio Astronomy. Springer, NY (1988)

Estimating Quality of Support Vector Machines Learning under Probabilistic and Interval Uncertainty: Algorithms and Computational Complexity

Canh Hao Nguyen[1], Tu Bao Ho[1], and Vladik Kreinovich[2]

[1] School of Knowledge Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
{canhhao,bao}@jaist.ac.jp
[2] Department of Computer Science, University of Texas, El Paso, TX 79968-0518, USA
vladik@utep.edu

Summary. Support Vector Machines (SVM) is one of the most widely used technique in machines leaning. After the SVM algorithms process the data and produce some classification, it is desirable to learn how well this classification fits the data. There exist several measures of fit, among them the most widely used is kernel target alignment. These measures, however, assume that the data are known exactly. In reality, whether the data points come from measurements or from expert estimates, they are only known with uncertainty. As a result, even if we know that the classification perfectly fits the nominal data, this same classification can be a bad fit for the actual values (which are somewhat different from the nominal ones). In this paper, we show how to take this uncertainty into account when estimating the quality of the resulting classification.

1 Formulation of the Problem

Machine learning: main problem. In many practical situations, we have examples of several types of objects, and we would like to use these examples to teach the computers to distinguish between objects of different types. Each object can be characterized by the corresponding values of several relevant quantities. If we denote the number of these quantities by d, then we can say that each object i can be represented by a d-dimensional vector $x^{(i)} = (x_1^{(i)}, \ldots, x_k^{(i)}, \ldots, x_d^{(i)})$, where $x_k^{(i)}$ denotes the value of the k-th quantity for i-th object. So, from the mathematical viewpoint, the problem is as follows: in d-dimensional space X, we have several points $x^{(1)}, \ldots, x^{(n)}$ belonging to different classes, and we need to be able, given a new point $x \in X$, to assign it to one of these classes.

In the simplest case when we have two classes, we have several points belonging to the first class, and several points which do not belong to the first class, and we must find a separating algorithm.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 57–69, 2008.
springerlink.com

Linear classification: main idea. In the past, a typical approach to data classification was to find a hyperplane $c_1 \cdot x_1 + \ldots + c_d \cdot x_d = c_0$ which separates the two classes – so that $c_1 \cdot x_1^{(i)} + \ldots + c_d \cdot x_d^{(i)} > c_0$ for all *positive* examples (i.e., examples from the first class) and $c_1 \cdot x_1^{(i)} + \ldots + c_d \cdot x_d^{(i)} < c_0$ for all *negative* examples (i.e., examples which do not belong to the first class).

Linear classification: limitations. The main limitations of linear classification approach is that in many important practical cases, there is no hyperplane separating positive and negative examples.

For example, suppose that we want to teach the computer to distinguish between the center of the city and its suburbs. To do that, we mark several places in the center as positive examples and places in the suburbs as negative examples. Here, a natural idea is to take $d = 2$, so that x_1 and x_2 are two coordinates of each point. To make it easier, we can take the central square of the city as the origin of the coordinate system, i.e., as a point $(0, 0)$.

In this example, separation is straightforward: points whose distance $\sqrt{x_1^2 + x_2^2}$ to the center is below a certain threshold t are within the city center, while points for which the distance is $> t$ are in the suburbs. However, no straight line can separate close points from the distant ones – because on each side of the straight line we have points which are far away from the center.

Support Vector Machines: main idea. What can we do when there is no linear separation? In the 2-D case, as long as there is a separation, i.e., as long as the same point $x \in X$ does not appear as both a positive and a negative example, we can draw a curve separating positive points from negative ones. Similarly, in the d-dimensional case, we can always draw a $(d-1)$-dimensional surface separating positive and negative examples. Moreover, we can always find a continuous function $f(x_1, \ldots, x_d)$ such that $f(x_1^{(i)}, \ldots, x_d^{(i)}) > 0$ for all positive examples and $f(x_1^{(i)}, \ldots, x_d^{(i)}) < 0$ for all negative examples.

A continuous function $f(x_1, \ldots, x_d)$ can be, with arbitrary accuracy, approximated by polynomials; thus, be selecting a good enough accuracy, we can have a polynomial

$$\widetilde{f}(x_1, \ldots, x_d) = c_0 + c_1 \cdot x_1 + \ldots + c_d \cdot x_d + \sum_{k=1}^{d} \sum_{l=1}^{d} c_{kl} \cdot x_k \cdot x_l + \ldots$$

which has the same separating property, i.e.,

$$c_0 + c_1 \cdot x_1^{(i)} + \ldots + c_d \cdot x_d^{(i)} + \sum_{k=1}^{d} \sum_{l=1}^{d} c_{kl} \cdot x_k^{(i)} \cdot x_l^{(i)} + \ldots > 0$$

for all positive examples and

$$c_0 + c_1 \cdot x_1^{(i)} + \ldots + c_d \cdot x_d^{(i)} + \sum_{k=1}^{d} \sum_{l=1}^{d} c_{kl} \cdot x_k^{(i)} \cdot x_l^{(i)} + \ldots < 0$$

for all negative examples. These formulas clearly show that this non-linear separation means that we linearly separate points $(x_1, \ldots, x_n, x_1^2, x_1 \cdot x_2, \ldots)$.

Instead of polynomials, we could use trigonometric polynomials or sums of Gaussian functions, or any other class of approximating functions. In all these cases, what we are doing is mapping each point x into a point $\phi(x) = (\phi_1(x), \ldots, \phi_p(x), \ldots, \phi_N(x))$ in a higher-dimensional space (of dimension $N \geq d$), and then use linear separation to separate the resulting points $\phi(x^{(1)}), \ldots, \phi(x^{(n)})$ in the N-dimensional space. This, in a nutshell, is the main idea behind the Support Vector Machines (SVM) techniques; see, e.g., [10].

Need to estimate classification quality. The fact that we have a surface separating positive examples from negative examples does not necessarily mean that this classification is good. Intuitively, if we have a new example x which is similar to one of the previously given examples $x^{(i)}$, then this new example should be classified to the same class as $x^{(i)}$. So, we want to make sure not only that all the positive examples are on the right side of the separating surface, but also that the points which are close to these examples are also on the same side of the separating surface. In other words, we want to make sure that all the examples are sufficiently far away from the separating surface. Thus, some reasonable measure of the distance from this surface can serve as the measure of the quality of the resulting classification.

Several such criteria have been proposed. These criteria are usually defined in terms of the *kernel matrix* $k_{ij} \stackrel{\text{def}}{=} \langle \phi(x^{(i)}), \phi(x^{(j)}) \rangle$, where

$$\langle \phi, \phi' \rangle \stackrel{\text{def}}{=} \sum_{p=1}^{N} \phi_p \cdot \phi'_p.$$

KTA. The most widely used criterion is the kernel target alignment (KTA) A [1], which is defined as follows (in our notations):

$$A = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} k_{ij} \cdot y_i \cdot y_j}{n \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} k_{ij}^2},$$

where $y_i = 1$ for positive examples and $y_i = -1$ for negative examples. This criterion has a very intuitive meaning. In the ideal situation, the separation should be as sharp as possible: we should have all the vectors $\phi(x^{(i)})$ corresponding to the positive examples to be equal to some unit vector e and all the vectors corresponding to the negative examples to be equal to $-e$. In this ideal situation, the kernel matrix is equal to $y_i \cdot y_j$. To estimate the quality of a classification, it is reasonable to check how close the actual kernel matrix is to this ideal one. One way to check is to consider both matrices as vectors in a $N \times N$ dimensional space, and estimate the cosine of the angle between these vectors; if the vectors

coincide, the angle is 0, and the cosine is 1; if the vectors are close, the angle is close to 0, and the cosine is close to 1, etc. This cosine is equal to

$$\frac{\langle K, \cdot y^T \rangle_F}{\sqrt{\langle K, K \rangle_F \langle y \cdot y^T, y \cdot y^T \rangle_F}},$$

where

$$\langle K, K' \rangle_F \stackrel{\text{def}}{=} \sum_{i=1}^{n} \sum_{j=1}^{n} k_{ij} \cdot k'_{ij},$$

so we get the above expression for the KTA.

Class Separability Measure (CSM). An alternative measure to KTA has been proposed in [13]. This measure is actually defined for a general case of classifying the data into several ($C \geq 2$) classes. The main idea is that in a good classification, data points within each class should be close to each other, while data points from different classes should be far away from each other. In other words, "within-class" scatter should be much smaller than the "between-classes" scatter.

Each class is naturally characterized by its average. Thus, for each data point, its contribution to the "within-class" scatter can be described as a (squared) distance from this data point to the average, and its contribution to the "between-classes" scatter can be described as a (squared) distance between the average of this class and the overall average.

In the SVM approach, each data point $x^{(i)}$ is represented by the vector $\phi(x^{(i)})$. Thus, the above idea can be reformulated as follows. For each class S_c, $c = 1, 2, \ldots, C$, let n_c denote the number of data points classified into this class. Let ϕ_c denote the average of all the vectors $\phi(x^{(i)})$ from the c-th class, and let ϕ denote the average of all n vectors $\phi(x^{(i)})$. Then, we can define the *within-class scatter* s_w as

$$s_w \stackrel{\text{def}}{=} \sum_{c=1}^{C} \sum_{i \in S_c} \|\phi(x^{(i)}) - \phi_c\|^2,$$

and the *between-classes scatter* as

$$s_b \stackrel{\text{def}}{=} \sum_{c=1}^{C} n_c \cdot \|\phi_c - \phi\|^2.$$

We can also define *total scatter* as the sum $s_t \stackrel{\text{def}}{=} s_w + s_b$. A classification is of good quality if $s_w \ll s_b$, i.e., equivalently, if $s_b \approx s_t$ and the ratio $C \stackrel{\text{def}}{=} \frac{s_b}{s_t}$ is close to 1. This ratio C is used as an alternative quality characteristic.

For the case of two classes, we will denote the number of the corresponding examples as n^+ and n^-, and the averages of the corresponding vectors $\phi(x^{(i)})$ by ϕ^+ and ϕ^-. The value of the CSM ratio C can be computed in terms of the kernel matrix $k_{ij} = \langle \phi(x^{(i)}), \phi(x^{(j)}) \rangle$ as follows:

- First, for every i, we compute

$$a_i^+ = \frac{1}{n^+} \sum_{j:y_j=1} k_{ij}; \quad a_i^- = \frac{1}{n^-} \sum_{j:y_j=-1} k_{ij}.$$

- Second, we compute

$$a^{++} = \frac{1}{n^+} \sum_{j:y_j=1} a_i^+, \quad a^{+-} = \frac{1}{n^-} \sum_{j:y_j=-1} a_i^+,$$

$$a^{-+} = \frac{1}{n^+} \sum_{j:y_j=1} a_i^-, \quad a^{--} = \frac{1}{n^+} \sum_{j:y_j=-1} a_i^-,$$

  and $s_b = a^{++} - a^{+-} - a^{-+} + a^{--}$.
- Then, we compute

$$s_w = \sum_{i=1}^{n} k_{ii} - n^+ \cdot a^{++} - n^- \cdot a^{--},$$

  and $C = \dfrac{s_b}{s_b + s_w}$.

A new alternative quality measure: FSM. In many practical examples, KTA and CSM provides a reasonable estimate for the quality of fit, in the sense that cases when we have a better fit have a larger value of KTA or CSM. However, there are examples when the values of KTA and CSM are larger for the cases when intuitively, the classification quality is worse.

One reason for the sometimes counterintuitive character of CSM is that CSM estimates a within-class scatter based on deviations in all directions. For example, if for some coordinate $\phi_p(x)$, we have $\phi_p(x^{(i)}) = 1$ for all positive examples and $\phi_p(x^{(i)}) = -1$ for all negative examples, then intuitively, we have a perfect classification. However, since the values $\phi_q(x^{(i)})$ for $q \neq p$ may be widely scattered, we can have a huge value of the within-class scatter, and thus, a very low value of the CSM measure of fit.

To avoid this problem, it is reasonable to take into account only the scatter in the direction between the centers ϕ^- and ϕ^+. The corresponding Feature-Spaced Measure (FSM) was proposed in [7].

To estimate this measure, we do the following:

- First, we compute the average ϕ^+ of the values $\phi(x^{(i)})$ for all the positive examples and the average ϕ^- of the values $\phi(x^{(i)})$ for all the negative examples. In the ideal case, as we have mentioned, we should have $\phi^+ = e$ and $\phi^- = -e$ for some unit vector e.
- Then, we estimate the vector e as the unit vector in the direction of the difference $\phi^+ - \phi^-$, i.e., as $e = \dfrac{\phi^+ - \phi^-}{\|\phi^+ - \phi^-\|}$.

- Next, for each example i, we compute the projection $p_i = \langle \phi(x^{(i)}), e \rangle$ of the vector $\phi^{(i)}$ to the direction e.
- Finally, we compute the population means

$$p^+ = \frac{1}{n^+} \cdot \sum_{i:y_i=1} p_i; \quad p^- = \frac{1}{n^-} \cdot \sum_{i:y_i=-1} p_i,$$

where n^+ and n^- $(= n - n^+)$ denote the numbers of positive and negative examples, compute population variances

$$V^+ = \frac{1}{n^+ - 1} \cdot \sum_{i:y_i=1} (p_i - p^+)^2; \quad V^- = \frac{1}{n^- - 1} \cdot \sum_{i:y_i=-1} (p_i - p^-)^2,$$

and the desired value

$$\frac{\sqrt{V^+} + \sqrt{V^-}}{\|\phi^+ - \phi^-\|}.$$

This algorithm describes how to compute these values based on the vectors $\phi(x^{(i)})$; alternatively, as shown in [7], we can compute it from the kernel matrix k_{ij} as follows:

- First, we compute the values a_i^+ and a_i^- as in the CSM case.
- Second, we compute the values a^{++}, a^{+-}, a^{-+}, and a^{--} as in the CSM case, and compute $\|\phi^+ - \phi^-\|^2 = a^{++} - a^{+-} - a^{-+} + a^{--}$.
- Then, we compute

$$V^+ = \frac{\sum_{i:y_i=1} ((a_i^- - a^{-+}) - (a_i^+ - a^{++}))^2}{(n^+ - 1) \cdot \|\phi^+ - \phi^-\|^2},$$

$$V^- = \frac{\sum_{i:y_i=-1} ((a_i^+ - a^{+-}) - (a_i^- - a^{--}))^2}{(n^- - 1) \cdot \|\phi^+ - \phi^-\|^2},$$

and the desired value

$$\frac{\sqrt{V^+} + \sqrt{V^-}}{\|\phi^+ - \phi^-\|}.$$

2 How to Take into Account Probabilistic and Interval Uncertainty: Formulation of the Problem and Linearized Algorithms for Solving This Problem

Need to take into account probabilistic and interval uncertainty. In presenting algorithms for computing the SVM quality measures, we (implicitly) assumed that we know the exact values of the data points $x^{(i)} = (x_1^{(i)}, \ldots, x_d^{(i)})$. In reality, the values $x_k^{(i)}$ come from measurements or from expert estimates, and both measurements and expert estimates are never 100% accurate. As a result, the

measured (estimated) values $\widetilde{x}_k^{(i)}$ of the corresponding quantities are, in general, different from the (unknown) actual values $x_k^{(i)}$.

It is desirable to take into account this measurement (estimation) uncertainty when estimating the quality measures.

Need to describe measurement and/or estimation uncertainty. In order to gauge the effect of the measurement (estimation) uncertainty on the values of the quality measures, we must have the information about the measurement (estimation) uncertainty, i.e., the information about the measurement (estimation) errors $\Delta x_k^{(i)} \stackrel{\text{def}}{=} \widetilde{x}_k^{(i)} - x_k^{(i)}$.

For simplicity, in the following text, we will mainly talk about measurement errors; for estimation errors the situation is very similar.

Upper bound on the measurement error. How can the measurement error be described? First, the manufacturer of a measuring instrument must provide us with an upper bound Δ on the absolute value $|\Delta x|$ of the measurement error Δx. If no such bound was guaranteed, this would mean that the difference Δx can be arbitrarily large; in this situation, after getting a measurement result say $\widetilde{x} = 1$, we cannot be sure whether the actual value x of the measured quantity is 1, 0, 10, 100, or 1,000,000. In this situation, $\widetilde{x} = 1$ is a wild guess, not a measurement result.

When we know this upper bound Δ, this means that the actual value Δx of the measurement error must be inside the interval $[-\Delta, \Delta]$.

Probabilistic information. In addition to the upper bound Δ, we often also know the probabilities of different values Δx from the interval $[-\Delta, \Delta]$.

This situation of probabilistic uncertainty is traditionally used in engineering and scientific practice. Most frequently, scientists and engineers consider the situation when the measurement error is normally distributed, with 0 mans and known standard deviation σ; see, e.g,., [9].

Case of interval uncertainty. In many important practical situations, we do not have the information about the probabilities of different values of Δx, we only know the upper bound Δ.

The reason is that the probabilistic information usually comes from comparing the results of measuring the same quantity with two different measuring instruments: the one used for actual measurements and the *standard* (much more accurate) one – whose results are so much closer to the actual values that we can ignore the corresponding measurement errors and consider these results actual values.

There are two situations when this comparison is not done. The first such situation is the situation of cutting-edge measurements, when we are actually using the best possible measuring instrument. For example, if we perform some protein measurements by using a state-of-the-art electronic microscope, it would be nice to be able to compare the results with a much more accurate microscope – but ours is already the best.

Another case when the probabilities are not determined is when we have limited resources. For example, in geophysics, in every seismic experiment, we use a large number of sensors to measure the corresponding time delays. It would be nice to be able to compare all these sensors with more accurate ones. However, the detailed comparison of each sensor requires the use of costly standard sensors and, as a result, costs several orders of magnitude more than the cost of buying a new sensor – so we often cannot do this detailed probabilistic "calibration" within our limited resources.

In both cases, the only information we have about the measurement error $\Delta x = \widetilde{x} - x$ is the upper bound Δ: $|\Delta x| \le \Delta$. In such situations, once we have the measurement result $\widetilde{x}$, the only conclusion that we can make about the (unknown) actual value x is that x belongs to the interval $\mathbf{x} = [\underline{x}, \overline{x}]$, where $\underline{x} \stackrel{\text{def}}{=} \widetilde{x} - \Delta$ and $\overline{x} \stackrel{\text{def}}{=} \widetilde{x} + \Delta$. This situation is called the situation of *interval uncertainty.*

Estimating the measures of fit under measurement uncertainty: formulation of the problem. In general, we have an algorithm

$$Q(x_1^{(1)}, \ldots, x_d^{(1)}, x_1^{(2)}, \ldots, x_d^{(2)}, \ldots, x_1^{(n)}, \ldots, x_d^{(n)})$$

which transforms the values

$$x_1^{(1)}, \ldots, x_d^{(1)}, x_1^{(2)}, \ldots, x_d^{(2)}, \ldots, x_1^{(n)}, \ldots, x_d^{(n)}$$

of the corresponding quantities into the value

$$y = Q(x_1^{(1)}, \ldots, x_d^{(1)}, x_1^{(2)}, \ldots, x_d^{(2)}, \ldots, x_1^{(n)}, \ldots, x_d^{(n)})$$

of the corresponding quality characteristic. Due to measurement errors, we do not know the actual values $x_k^{(i)}$. Instead, we only know the intervals $[\underline{x}_k^{(i)}, \overline{x}_k^{(i)}]$ of possible values of $x_k^{(i)}$ – and possible also the probabilities of different values from these intervals.

Different values $x_k^{(i)} \in [\underline{x}_k^{(i)}, \overline{x}_k^{(i)}]$ lead, in general, to different values of the measure of fit $y = Q(x_1^{(1)}, \ldots, x_d^{(n)})$. It is therefore desirable to find the range $\mathbf{y}$ of possible values of y:

$$\mathbf{y} = \{Q(x_1^{(1)}, \ldots, x_d^{(n)}) \,|\, x_1^{(1)} \in [\underline{x}_1^{(1)}, \overline{x}_1^{(1)}], \ldots, x_d^{(n)} \in [\underline{x}_d^{(n)}, \overline{x}_d^{(n)}]\},$$

and, if possible, the probability of different values of y within this interval.

Case of relatively small measurement error: possibility of linearization. When the measurement errors Δx_i are relatively small, we can use linearization.

By definition of the measurement error $\Delta x_k^{(i)} = \widetilde{x}_k^{(i)} - x_k^{(i)}$, hence $x_k^{(i)} = \widetilde{x}_k^{(i)} - \Delta x_k^{(i)}$. When the measurement errors $\Delta x_k^{(i)}$ of direct measurements are relatively small, we can expand the expression

$$\Delta y = \widetilde{y} - y = Q(\widetilde{x}_1^{(1)}, \ldots, \widetilde{x}_d^{(n)}) - Q(x_1^{(1)}, \ldots, x_d^{(n)}) =$$

$$Q(\widetilde{x}_1^{(1)}, \ldots, \widetilde{x}_d^{(n)}) - Q(\widetilde{x}_1^{(1)} - \Delta x_1^{(1)}, \ldots, \widetilde{x}_d^{(n)} - \Delta x_d^{(n)})$$

in Taylor series and only keep linear terms in the resulting expansion. Since

$$y = Q(\widetilde{x}_1^{(1)} - \Delta x_1^{(1)}, \ldots, \widetilde{x}_d^{(n)} - \Delta x_d^{(n)}) \approx Q(\widetilde{x}_1^{(1)}, \ldots, \widetilde{x}_d^{(n)}) - \sum_{i=1}^{n} \sum_{k=1}^{d} \frac{\partial Q}{\partial x_k^{(i)}} \cdot \Delta x_k^{(i)},$$

we conclude that $\Delta y = \widetilde{y} - y = \sum\limits_{i=1}^{n} \sum\limits_{k=1}^{d} c_k^{(i)} \cdot \Delta x_k^{(i)}$, where $c_k^{(i)} \stackrel{\text{def}}{=} \dfrac{\partial Q}{\partial x_k^{(i)}}$.

Linearization: probabilistic case. When $\Delta x_k^{(i)}$ are independent normally distributed random variables with 0 means and known standard deviations $\sigma_k^{(i)}$, the linear combination $\Delta y = \sum\limits_{i=1}^{n} \sum\limits_{k=1}^{d} c_k^{(i)} \cdot \Delta x_k^{(i)}$ is also normally distributed, with 0 mean and standard deviation

$$\sigma = \sqrt{\sum_{i=1}^{n} \sum_{k=1}^{d} (c_k^{(i)} \cdot \sigma_k^{(i)})^2}.$$

So, in this case, to find the uncertainty in the value of the measure of fit, it is sufficient to be able to compute the values of the corresponding partial derivatives $c_k^{(i)}$.

Linearization: interval case. The dependence of Δy on $\Delta x_k^{(i)}$ is linear: it is increasing relative to $x_k^{(i)}$ if $c_k^{(i)} \geq 0$ and decreasing if $c_k^{(i)} < 0$. So, to find the largest possible value Δ of Δy, we must take:

- the largest possible value $\Delta x_k^{(i)} = \Delta_k^{(i)}$ when $c_k^{(i)} \geq 0$, and
- the smallest possible value $\Delta x_k^{(i)} = -\Delta_k^{(i)}$ when $c_k^{(i)} < 0$.

In both cases, the corresponding term in the sum has the form $|c_k^{(i)}| \cdot \Delta_k^{(i)}$, so we can conclude that

$$\Delta = \sum_{i=1}^{n} \sum_{k=1}^{d} |c_k^{(i)}| \cdot \Delta_k^{(i)}.$$

Similarly, the smallest possible value of Δy is equal to $-\Delta$. Thus, the range of possible values of y is equal to $[\underline{y}, \overline{y}] = [\widetilde{y} - \Delta, \widetilde{y} + \Delta]$. So, to compute Δ, it is also sufficient to know the partial derivatives $c_k^{(i)}$.

How to compute the derivatives. For all the above characteristics y, we have an explicit expression in terms of the values k_{ij}. Thus, we can find the explicit analytic formulas in terms of the corresponding derivatives as

$$c_k^{(i)} = \sum_{a=1}^{n} \sum_{b=1}^{n} \frac{\partial y}{\partial k_{ab}} \cdot \frac{\partial k_{ab}}{\partial x_k^{(i)}}.$$

Here, the first partial derivative can be explicitly computed: e.g., for KTA $Q = A$, we have

$$\frac{\partial y}{\partial k_{ab}} = \frac{y_a \cdot y_b}{n \cdot \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} k_{ij}^2} - 2k_{ab} \cdot \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} k_{ij} \cdot y_i \cdot y_j}{n \cdot \left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} k_{ij}^2\right)^2}.$$

For $k_{ab} = \sum\limits_{p=1}^{N} \phi_p(x^{(a)}) \cdot \phi_p(x^{(b)})$, the derivative $\dfrac{\partial k_{ab}}{\partial x_k^{(i)}}$ is only different from 0 if $a = i$ or $b = i$:

$$\frac{\partial k_{ib}}{\partial x_k^{(i)}} = \sum_{p=1}^{N} \frac{\partial \phi_p}{\partial x_k}(x^{(i)}) \cdot \phi_p(x^{(b)}) \quad \text{for} \quad a = i \quad \text{and} \quad b \neq i;$$

$$\frac{\partial k_{ai}}{\partial x_k^{(i)}} = \sum_{p=1}^{N} \phi_p(x^{(a)}) \cdot \frac{\partial \phi_p}{\partial x_k}(x^{(i)}) \quad \text{for} \quad a \neq i \quad \text{and} \quad b = i;$$

$$\frac{\partial k_{ii}}{\partial x_k^{(i)}} = 2\sum_{p=1}^{N} \frac{\partial \phi_p}{\partial x_k}(x^{(i)}) \cdot \phi_p(x^{(i)}) \quad \text{for} \quad a = b = i.$$

3 In General, Estimating Quality of SVM Learning under Interval Uncertainty Is NP-Hard

Motivations. In the previous section, we considered the case when measurement errors are small, e.g., no more than 10%, so that we can ignore terms which are quadratic in terms of these errors. For example, for 10% = 0.1, the quadratic terms are proportional to $0.1^2 = 1\% \ll 10\%$ and thus, indeed, much smaller than the original errors. In this case, we can linearize the formulas for the quality of SVM learning and get efficient algorithms for computing the range of the corresponding quality characteristics.

In practice, however, the measurement errors are often not very small. For example, for a realistic measurement error of 30%, the square is $\approx 10\%$ and is no longer negligible in comparison with the original measurement errors. In such situations, we can no longer use linearized techniques, we must consider the original problem of computing the range $[\underline{y}, \overline{y}]$ of a given characteristic $Q(x_1^{(1)}, \ldots, x_d^{(n)})$ under interval uncertainty:

$$[\underline{y}, \overline{y}] = \{Q(x_1^{(1)}, \ldots, x_d^{(n)}) \,|\, x_1^{(1)} \in [\underline{x}_1^{(1)}, \overline{x}_1^{(1)}], \ldots, x_d^{(n)} \in [\underline{x}_d^{(n)}, \overline{x}_d^{(n)}]\}.$$

It turns out that in general, this problem is NP-hard – at least it is NP-hard for the most widely used measures of fit KTA and CSM.

Crudely speaking, NP-hard means that there is practically no hope of designing an efficient algorithm which would always correct compute this range; for precise definitions, see, e.g., [3, 4, 8].

Theorem 1. *Computing the range of KTA under interval uncertainty is NP-hard.*

Proof. To prove NP-hardness of our problem, we will reduce a known NP-hard problem to our problem of computing the range $\mathbf{A}$ of KTA A under interval uncertainty. Specifically, we will reduce, to our problem, the following *partition* problem [3] that is known to be NP-hard:

- Given k positive integers $s_1, \ldots, s_k$,
- check whether it is possible to find the values $\varepsilon_i \in \{-1, 1\}$ for which $\sum_{i=1}^{k} \varepsilon_i \cdot s_i = 0$.

To each instance $s_1, \ldots, s_k$ of this problem, we assign the following instance of the problem of computing $\mathbf{A}$: we take $d = 1$, $n = k + 1$, $y_1 = \ldots = y_k = 1$, $y_{k+1} = -1$, $\mathbf{x}^{(i)} = [-s_i, s_i]$ for $i \le k$, and $\mathbf{x}^{(n)} = \{2S\}$, where $S \stackrel{\text{def}}{=} \max_{i=1,\ldots,k} s_i$. As ϕ, we take a 2-dimensional mapping $\phi = (\phi_1, \phi_2)$ consisting of the following two piece-wise linear functions:

$$\phi_1(x) = \begin{cases} x & \text{if } x \le S \\ 2S - x & \text{if } S \le x \le 2S \\ 0 & \text{if } x \ge 2S \end{cases}; \quad \phi_2(x) = \begin{cases} 0 & \text{if } x \le S \\ x/S - 1 & \text{if } S \le x \le 2S \\ 1 & \text{if } x \ge 2S \end{cases}.$$

In this case,

$$k_{ij} = \langle \phi(x^{(i)}), \phi(x^{(j)}) \rangle = \begin{cases} x^{(i)} \cdot x^{(i)} & \text{if } i, j < n, \\ 1 & \text{if } i = j = n, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} k_{ij} \cdot y_i \cdot y_j = \sum_{i=1}^{k} \sum_{j=1}^{k} x^{(i)} \cdot x^{(i)} + 1 = \left(\sum_{i=1}^{k} x^{(i)} \right)^2 + 1;$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} k_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \left(x^{(i)} \right)^2 \cdot \left(x^{(j)} \right)^2 + 1 = \left(\sum_{i=1}^{k} \left(x^{(i)} \right)^2 \right)^2 + 1; \text{ and}$$

$$A = \frac{\left(\sum_{i=1}^{k} x^{(i)} \right)^2 + 1}{n \cdot \sqrt{\left(\sum_{i=1}^{k} \left(x^{(i)} \right)^2 \right)^2 + 1}}.$$

The numerator is always greater than or equal to 1. Since $|x^{(i)}| \le s_i$, we have $\left(x^{(i)} \right)^2 \le s_i^2$ and hence, the denominator is always $\le n \cdot \sqrt{\left(\sum_{i=1}^{k} s_i^2 \right)^2 + 1}$. Thus,

we always have $A \geq A_0 \stackrel{\text{def}}{=} \dfrac{1}{n \cdot \sqrt{\left(\sum\limits_{i=1}^{k} s_i^2\right)^2 + 1}}$. The only possibility for $A = A_0$ is when the numerator of the fraction A is equal to 1, and its denominator is equal to $n \cdot \sqrt{\left(\sum\limits_{i=1}^{k} s_i^2\right)^2 + 1}$. This is only possible when $|x^{(i)}| = s_i$ for all i, i.e., when $x^{(i)} = \varepsilon_i \cdot s_i$ for some $\varepsilon_i \in \{-1, 1\}$, and $\sum\limits_{i=1}^{k} x^{(i)} = 0$ – i.e., exactly when the original instance of the partition problem has a solution. So, $\underline{A} = A_0$ if and only if the original instance has a solution. This reduction proves that our problem is indeed NP-hard.

Theorem 2. *Computing the range of CSM under interval uncertainty is NP-hard.*

Proof. Under the same reduction as in Theorem 1, we get $n^+ = k$, $n^- = 1$, $a_n^+ = 0$ and for $i < n$, we have $a_i^+ = \dfrac{1}{k} \cdot \sum\limits_{j=1}^{k} x^{(i)} \cdot x^{(j)} = x^{(i)} \cdot E$, where $E \stackrel{\text{def}}{=} \dfrac{1}{k} \cdot \sum\limits_{i=1}^{k} x^{(i)}$. Similarly, $a_n^- = 1$ and $a_i^- = 0$ for all $i < n$. Thus, $a^{++} = \dfrac{1}{k} \cdot \sum\limits_{i=1}^{k} a_i^+ = E \cdot \dfrac{1}{k} \cdot \sum\limits_{i=1}^{k} x^{(i)} = E^2$, $a^{+-} = a^{-+} = 0$, and $a^{--} = 1$. Hence, $s_b = E^2 + 1$, $s_w = \sum\limits_{i=1}^{k} \left(x^{(i)}\right)^2 - k \cdot E^2 - 1$ and thus, $C = \dfrac{E^2 + 1}{\sum\limits_{i=1}^{k} \left(x^{(i)}\right)^2 - (k-1) \cdot E^2}$. The numerator is ≥ 1, the denominator is $\leq \sum\limits_{i=1}^{k} s_i^2$, hence $C \geq C_0 \stackrel{\text{def}}{=} \dfrac{1}{\sum\limits_{i=1}^{k} s_i^2}$. The only possibility to have $C = C_0$ is when $E = 0$ and $|x^{(i)}| = s_i$ for all i, i.e., when the original instance of the partition problem has a solution. The theorem is proven.

4 Conclusion

For classification produced by machine learning techniques, it is desirable to learn how well this classification fits the data. There exist several measures of fit, among them the most widely used is kernel target alignment.

The existing formulas for these measures assume that the data are known exactly. In reality, whether the data points come from measurements or from expert estimates, they are only known with uncertainty. As a result, even if we know that the classification perfectly fits the nominal data, this same classification can be a bad fit for the actual values (which are somewhat different from

the nominal ones). In this paper, we show how, when the measurement errors are relatively small, we can take this uncertainty into account when estimating the quality of the resulting classification. We also show that in the general case of large uncertainty, the problem of estimating the range of these measures of fit is NP-hard.

Acknowledgments. Vladik Kreinovich supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, and by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08.

The authors are thankful to the anonymous referees for valuable comments.

References

1. Cristianini, N., et al.: On kernel-target alignment. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14, MIT Press, Cambridge, Massachusetts (2002)
2. Ferson, S., et al.: Computing variance for interval data is NP-hard. SIGACT News 33(2), 108–118 (2002)
3. Garey, M.R., Johnson, D.S.: Computers and Intractability, a Guide to the Theory of NP-Completeness. WH Freeman and Company, San Francisco, California (1979)
4. Kreinovich, V., et al.: Computational Complexity and Feasibility of Data Processing and Interval Computations. Kluwer, Dordrecht (1997)
5. Kreinovich, V., et al.: Interval versions of statistical techniques, with applications to environmental analysis, bioinformatics, and privacy in statistical databases. Journal of Computational and Applied Mathematics 199, 418–423 (2007)
6. Jaulin, L., et al.: Applied Interval Analysis: With Examples in Parameter and State Estimation, Robust Control and Robotics. Springer, London (2001)
7. Nguyen, C.H., Ho, T.B.: Kernel matrix evaluation. In: Manuela, M., Veloso, M.M. (eds.) Proceedings of the 20th International Joint Conference on Artificial Intelligence IJCAI 2007, Hyderabad, India, January 6–12, 2007, pp. 987–992 (2007)
8. Papadimitriou, C.H., Steiglitz, K.: Combinatorial Optimization: Algorithms and Complexity. Dover Publications, Inc., Mineola, New York (1998)
9. Rabinovich, S.G.: Measurement Errors and Uncertainty. Theory and Practice. Springer, Berlin (2005)
10. Schölkopf, B., Smola, A.J.: Learning with Kernels. MIT Press, Cambridge, Massachusetts (2002)
11. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, New York (2004)
12. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
13. Wang, L., Chan, K.L.: Learning kernel parameters by using class separability measure. In: NIPS's Sixth Kernel Machines Workshop, Whistler, Canada (2002)

Imprecise Probability as an Approach to Improved Dependability in High-Level Information Fusion

Alexander Karlsson, Ronnie Johansson, and Sten F. Andler

School of Humanities and Informatics, University of Skövde, Sweden
{alexander.karlsson, ronnie.johansson, sten.f.andler}@his.se

Summary. The main goal of information fusion can be seen as improving human or automatic decision-making by exploiting diversities in information from multiple sources. High-level information fusion aims specifically at decision support regarding situations, often expressed as "achieving situation awareness". A crucial issue for decision making based on such support is trust that can be defined as "accepted dependence", where dependence or dependability is an overall term for many other concepts, e.g., reliability. This position paper reports on ongoing and planned research concerning imprecise probability as an approach to improved dependability in high-level information fusion. We elaborate on high-level information fusion from a generic perspective and a partial mapping from a taxonomy of dependability to high-level information fusion is presented. Three application domains: defense, manufacturing, and precision agriculture, where experiments are planned to be implemented are depicted. We conclude that high-level information fusion as an application-oriented research area, where precise probability (Bayesian theory) is commonly adopted, provides an excellent evaluation ground for imprecise probability.

1 Introduction

Information fusion (IF) is a research field that has been tightly coupled with defense applications (e.g., [27]) for many years. However, recently researchers in other domains such as manufacturing (e.g., [38]) and precision agriculture (e.g., [34]) have started to recognize the benefits of IF. IF, sometimes also referred to as data fusion, can be depicted as done by Dasarathy [13].

> *"Information fusion encompasses the theory, techniques, and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources (sensor, databases, information gathered by human etc.) such that the resulting decision or action is in some sense better (qualitatively and quantitatively, in terms of accuracy, robustness and etc.) than would be possible, if these sources were used individually without such synergy exploitation."*

From Dasarathy's description, it is seen that the overall goal is to improve decision making, and since there most often exist uncertainty regarding decisions, it

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 70–84, 2008.
springerlink.com

can be concluded that *uncertainty management* is crucial to IF. In fact, it has even been argued that the main goal of an IF system is to reduce uncertainty [9]. Many methods for handling uncertainty in the IF domain are based on *Bayesian theory*, e.g., *Kalman filtering* [16] and *Bayesian networks* (BNs) [6].

Most research in the IF domain so far has addressed problems in low-level IF, e.g., *target tracking* by *multi-sensor fusion*, while the higher abstraction levels of reasoning, referred to as *high-level information fusion* (HLIF), have been a relatively uncharted research field. Furthermore, those attempts that do exist for HLIF rarely address *dependability* issues (cf. [36]).

In this position paper, we elaborate on HLIF from a generic perspective and a partial mapping from a taxonomy of dependability to HLIF is presented. We argue for imprecise probability as an interesting approach for improved dependability in HLIF, and that more research on this topic is needed. We also argue for the importance of more research on deployment of methods based on imprecise probabilities in "real-world" applications.

The paper is organized as follows: in Sect. 2, we depict the foundations of IF. A thorough description of HLIF is presented in Sect. 3. In Sect. 4, we describe a partial mapping from a taxonomy of dependability to HLIF. Imprecise probability as an approach to improved dependability in HLIF is described in Sect. 5. Three application domains, *defense*, *manufacturing*, and *precision agriculture*, for evaluation of imprecise probability, are presented in Sect. 6. Lastly, in Sect. 7, we argue for the importance of evaluation of imprecise probability in comparison with precise probability (Bayesian theory) through experiments in "real-world" applications.

2 Information Fusion

In order to allow for easy communication among researchers of IF, the *Joint Directors of Laboratories* (JDL) data fusion group has established a model that comprises the IF domain [43]. The model, referred to as the *JDL model*, has been revised many times (e.g., [35, 26]), either due to the lack of some important aspect of IF, or for the purpose of making it more general. Steinberg et al. [35] have presented the following variant of the JDL model, hereafter referred to as the *revised JDL model*, with five functions or levels:

- **Level 0 – Sub-Object Data Assessment:** estimation and prediction of signal observable states
- **Level 1 – Object Assessment:** estimation and prediction of entity states, based on observation-to-track association
- **Level 2 – Situation Assessment:** estimation and prediction of relations among entities
- **Level 3 – Impact Assessment:** estimation and prediction of effects of actions on situations
- **Level 4 – Process Refinement:** continuous improvement of the information fusion process

The claim that uncertainty management is crucial to IF is here reinforced since words such as *estimation* and *prediction* appear in all of the levels except Level 4. With HLIF, we refer to Level 2, 3, and with low-level IF to Level 0, 1, in the revised JDL model. The reason to not include Level 4 in HLIF, is that it may be regarded as a *meta-process* that is a part of all levels, i.e., refinement of processes at each level.

3 High-Level Information Fusion

The aim of high-level information fusion (HLIF) is to establish the current *situation*, and possible *impacts* of that situation conditional on a *set of actions*. Since HLIF mainly has addressed issues in the defense domain, we here elaborate on it from a generic perspective. A terminology for HLIF that captures concepts such as *situations* and *impacts* is presented. It should be noted that there exists a framework, referred to as *situation theory* [14] for which there are some similarities to the terminology that we introduce here, e.g., that situations are about relations (this can also be seen from the revised JDL model). Kokar et al. [24] have developed an *ontology* for situation awareness that is based on situation theory and which is referred to as *situation awareness ontology* (STO). However, uncertainty is not the main focus of the above framework; thus, the concepts introduced here aim at providing a generic and clear understanding of HLIF from the perspective of uncertainty.

3.1 Level 2 – Situation Assessment

As can be seen from the revised JDL model, the main concern in Level 2 – Situation Assessment – is *relations among entities*. As noted by Kokar et al. [23, 24], a binary relation in mathematics, denoted by $\mathcal{R}$, has the following structure:

$$\mathcal{R} \subseteq X \times Y \tag{1}$$

$$X \times Y \stackrel{def.}{=} \{(x, y) : x \in X, y \in Y\} \tag{2}$$

However, in order to allow other relations than binary, it is necessary to consider *n-ary* relations that can be formally depicted as, cf. [24]:

$$\mathcal{R} \subseteq X_1 \times \ldots \times X_n \tag{3}$$

$$X_1 \times \ldots \times X_n \stackrel{def.}{=} \{(x_1, \ldots, x_n) : x_i \in X_i\} \tag{4}$$

A relation can be defined *intensionally* by a predicate P that decides which n-tuples that actually belong to the relation [24]:

$$\mathcal{R} \stackrel{def.}{=} \{(x_1, \ldots, x_n) : P(x_1, \ldots, x_n),\ x_i \in X_i\} \tag{5}$$

The relations that are of interest in HLIF, are usually not observable in a direct way; thus, uncertainty regarding the predicate, and hence the relation, is most

often evident. As Kokar et al. [24] have also noted, and which can be seen from the revised JDL model [35], a situation can consist of more than one relation. Consequently, it is necessary to interpret a situation as a *set of relations* denoted by $\mathcal{S}$ and more formally stated as:

$$\mathcal{S} \stackrel{def.}{=} \{\mathcal{R}_1, \ldots, \mathcal{R}_k\} \tag{6}$$

One may think of the relations as representing different *concepts* that for some reason are needed by a decision maker in order to make a decision about situations in a particular application domain.

Since there most often exists uncertainty regarding which n-tuples that satisfy the predicate for a given relation in HLIF; it is necessary to be able to consider the elements of the relation as *hypotheses*, which we here denote by $(x_1, \ldots, x_n)_j^h$ to indicate that it is a hypothesis with respect to a specific relation $\mathcal{R}_j$, $1 \leq j \leq k$. Since the elements of a given relation $\mathcal{R}_j$ now can be considered as hypotheses, it is also necessary to consider the relation itself as such, a *relational hypothesis*, denoted by $\mathcal{R}_j^h$. Lastly, since a situation $\mathcal{S}$ is defined using relations that may be hypotheses, a situation can also be a hypothesis, denoted by $\mathcal{S}_i^h$, $i \in \mathcal{J}_\mathcal{S}$, where $\mathcal{J}_\mathcal{S}$ is an index set.

Let the set of available information (sensor readings, domain knowledge and stored information) be denoted by ξ. Note that ξ may contain information that is uncertain, e.g., information from an unreliable source, imprecise, i.e., information which in some sense refers to more than one possibility, and inconsistent, e.g., information sources are in conflict [7] (for more detail, see [22, Sect. II-A1]). We will here use *belief* as a generic term for quantifying a rational agent's belief, thus, belief in the above sense is not associated with any particular *Uncertainty Management Method* (UMM). The following belief measure needs to be assessed in Level 2 – Situation Assessment:

$$\mu_\mathcal{S}(\mathcal{S}_i^h = \mathcal{S}|\xi), \tag{7}$$

i.e., the degree of belief for a specific situation hypothesis $\mathcal{S}_i^h$ being the "true" current situation $\mathcal{S}$ conditional on ξ. Depending on the application domain and the type of relations involved in $\mathcal{S}$, it may also be necessary to define belief measures that capture some specific part of a situation in more detail. Examples of such measures are:

$$\mu_{\mathcal{T}_j}((x_1, \ldots, x_n)_j^h \in \mathcal{R}_j|\xi) \tag{8}$$

$$\mu_{\mathcal{R}_j}(\mathcal{R}_j^h = \mathcal{R}_j|\xi), \tag{9}$$

where $\mu_{\mathcal{T}_j}$ denotes the degree of belief for a single tuple $\mathcal{R}_j$, and $\mu_{\mathcal{R}_j}$ depicts the degree of belief for the "true" set that constitutes the relation. In particular scenarios it could be sufficient to define some of these belief measures in terms of the others by for example using the *mean*. As an example $\mu_{\mathcal{R}_j}$ can be defined in terms of $\mu_{\mathcal{T}_j}$ by using the mean, in the following way:

$$\mu_{\mathcal{R}_j}(\mathcal{R}_j^h = \mathcal{R}_j|\xi) = \frac{1}{|X_1 \times \ldots \times X_n|} \left[\sum_{(x_1,\ldots,x_n)_j^h \in \mathcal{R}_j^h} \mu_{\mathcal{T}_j}((x_1,\ldots,x_n)_j^h \in \mathcal{R}_j|\xi) + \sum_{(x_1,\ldots,x_n)_j^h \in \overline{\mathcal{R}}_j^h} \mu_{\mathcal{T}_j}((x_1,\ldots,x_n)_j^h \notin \mathcal{R}_j|\xi) \right], \quad \text{where } \overline{\mathcal{R}}_j^h = (X_1 \times \ldots \times X_n) \backslash \mathcal{R}_j^h. \tag{10}$$

The last part of the equation can be simplified if one assumes that the following holds:

$$\mu_{\mathcal{T}_j}((x_1,\ldots,x_n)_j^h \notin \mathcal{R}_j|\xi) = 1 - \mu_{\mathcal{T}_j}((x_1,\ldots,x_n)_j^h \in \mathcal{R}_j|\xi) \tag{11}$$

However, depending on the UMM, this is not always the case (e.g., Dempster-Shafer theory). In the general case one might want to assess the measures $\mu_{\mathcal{T}_i}$, $\mu_{\mathcal{R}_i}$, and $\mu_{\mathcal{S}}$ more specifically by applying some other method than just using the mean over an existing belief measure.

3.2 Level 3 – Impact Assessment

Consider Level 3 – Impact Assessment – where the aim is to estimate effects on situations given certain actions [35]. The representation of situations still applies since *"estimation and prediction of effects on situations"* as stated in the revised JDL model can be interpreted as estimation and prediction of "future situations", *impacts*, which we here will denote by $\mathcal{I}$, and $\mathcal{I}_i^h$, $i \in \mathcal{J}_\mathcal{I}$, when considered as a hypothesis. From a decision maker's point of view, a certain set of planned actions is expected to lead to a desirable impact. Now, since there most often exists uncertainty regarding the *current situation*, something that is reflected in the belief measure $\mu_\mathcal{S}$, it is also necessary to incorporate this uncertainty when estimating future situations, impacts $\mathcal{I}_i^h$. Consequently, a belief measure for Impact Assessment, $\mu_\mathcal{I}$, has the following appearance:

$$\mu_\mathcal{I}(\mathcal{I}_i^h = \mathcal{I}|\mathcal{A}, \mu_\mathcal{S}, \xi), \tag{12}$$

i.e., the degree of belief regarding a possible impact $\mathcal{I}_i^h$ is conditional on: a set of actions $\mathcal{A}$, the belief measure for the current situation $\mu_\mathcal{S}$, and the set of available information ξ. Additional belief measures that capture some specific part of an impact in more detail, similarly to expressions (8) and (9), may also be defined for Impact Assessment.

4 Dependable High-Level Information Fusion

One of the main issues in HLIF is to assess the belief measures $\mu_\mathcal{S}$ and $\mu_\mathcal{I}$ over possible current situations, $\{\mathcal{S}_i^h\}_{i\in\mathcal{J}_\mathcal{S}}$ and possible impacts $\{\mathcal{I}_i^h\}_{i\in\mathcal{J}_\mathcal{I}}$. Since

these quantifications constitute a basis for *HLIF-based decision making*, human or automatic, a question one can pose is how *trustworthy* such quantifications are? However, a clarification on what is actually meant by "trustworthiness" or "trust" in HLIF is necessary since there is a lack of research on that topic. Avižienis et al. [5] define trust as *"accepted dependence"* and have presented a taxonomy of *dependability* that is well-accepted within the dependable computing domain. In HLIF, however, some of these concepts, e.g., *reliability* and *robustness*, are also utilized but with no consistent meaning; thus, researchers have adapted their particular interpretation in a specific application domain and problem. We will here present a partial mapping from this taxonomy to HLIF that preserves as much consistency as possible with respect to how concepts have been utilized in HLIF. Since specific characteristics of dependability is application dependent, this partial mapping should be seen as a guideline for interpreting dependability in HLIF. We will later use this terminology when we discuss why imprecise probability seems to be an interesting approach to improved dependability in HLIF (Sect. 5).

4.1 High-Level Information Fusion as a Service

The basis for the concepts in the dependability taxonomy is a *service*; thus, we need to clearly state what type of service the involved functions in HLIF provide. By the description of HLIF in Sect. 3, it can be argued that a *HLIF service* provides the artifacts listed in the introduction of this section, i.e., $\{\mathcal{S}_i^h\}_{i\in\mathcal{J}_\mathcal{S}}$, $\{\mathcal{I}_i^h\}_{i\in\mathcal{J}_\mathcal{I}}$, $\mu_\mathcal{S}$, and $\mu_\mathcal{I}$. These artifacts are utilized by a decision maker (human or automatic) in order to make a decision concerning situations, in other words, a HLIF service aims at providing decision support regarding situations.

4.2 Reliability

We start this partial mapping with the attribute of dependability referred to as *reliability*; an attribute that has many different interpretations [30, 26, 42]. Svensson [36] has proposed the following interpretation of reliability for HLIF-based decision-support systems:

> *"Technical system property of delivering quantitative results which are reasonably close to best possible, subject to known statistical error distributions".*

However, *"...results which are reasonably close to best possible..."* could be hard to interpret since *"best possible"* needs to be more clearly defined, and *"...statistical error distributions"* is too specific in many circumstances, e.g., when subjective opinions (from domain experts) are utilized.

In the dependability taxonomy, it is seen that reliability is defined as *"continuity of correct service"*. Llinas [25] has listed *"Correctness in reasoning"* as an important criteria for evaluation of fusion performance in a context of HLIF. Thus, one can argue that the key to think about reliability is *correctness*. The question then becomes what a correct HLIF service is, and a natural answer is

that correctness refers to what such service actually delivers, i.e., correctness of the artifacts: $\{\mathcal{S}_i^h\}_{i\in\mathcal{J}_\mathcal{S}}$, $\{\mathcal{I}_i^h\}_{i\in\mathcal{J}_\mathcal{I}}$, $\mu_\mathcal{S}$, and $\mu_\mathcal{I}$.

For the sets of hypotheses, correctness refers to the extent of all plausible hypotheses actually being in the sets (cf. [31]), i.e., *exhaustivity*. For the belief measures, it can be argued that correctness is achieved when the measures reflect the character of the available information (cf. [39, Section 1.1.4]). As an example, if information is scarce, uncertain, imprecise or conflicting (see further [22]), this should be reflected in the belief measure.

Lastly, if we assume that ξ continuously gets updated, e.g., via a sensor stream, then it is necessary for the belief measures to continuously reflect ξ, thus in agreement with *"continuity..."* in the definition of reliability found in the taxonomy.

4.3 Fault

A *fault* in the dependability taxonomy is defined as the cause of an *error* that in its turn is something that may cause a *failure*, i.e., a deviation from correct service [5]. Since we have already argued that a correct HLIF service delivers an exhaustive set of plausible hypotheses and belief measures that reflect the character of ξ, the negation of this statement, i.e., a service that provides a non-exhaustive set of plausible hypothesis or belief measures that do not reflect ξ, is considered to be a service that is *not* correct.

Based on these arguments, faults can be defined as: uncertain, imprecise, inconsistent, and lack of information (for more detail, see [22, Sect. II-A1]), since if inadequately handled, all of these may lead to service incorrectness. For a non-exhaustive set of hypotheses, *insufficient or inaccurate domain knowledge* about the process can also be considered as a fault since design of hypotheses most often rely on domain knowledge. Another important fault in HLIF concerns *unreliable sources*. If we consider a source as providing a service, reliability of this service would be correctness of the source output. It is possible to account for unreliable sources by introducing *reliability coefficients* that quantify the degree of reliability for the sources [30]. Thus, one can say that a service, based on unreliable sources, is still reliable as long as one know the *quality of sources*, e.g., reliability coefficients, and compensate for this.

4.4 Safety

The next concept in the taxonomy that we will consider is *safety*, which is defined as *"Absence of catastrophic consequences on the user(s) and the environment"* [5]. Seen from a decision maker's point of view, one can aim at a minimized number of *possible catastrophic consequences*. In essence, when a decision is taken by utilizing a HLIF service, a possible catastrophic consequence may be interpreted as an *unexpected impact* of such decision. There are two important so called *secondary attributes* (attributes that refine primary attributes [5]) that we consider to be a part of safety: *robustness* and *stability*.

Robustness

Svensson [36] has proposed the following definition of robustness for HLIF-based decision-support systems:

> *"Property of a system to react appropriately to exceptional conditions, including to avoid making large changes in recommendations as a consequence of small changes in input data."*

Bladon et al. [6] have proposed the following description of robust in conjunction with a Situation-Assessment system:

> *"Robust: able to handle inconsistent, uncertain, and incomplete data."*

Llinas [25] has listed the following criteria for evaluation of HLIF performance:

> *"Adaptability in reasoning (robustness)"*

Antony [3] claims that:

> *"Robustness measures the fragility of a problem-solving approach to changes in the input space."*

Avižienis et al. [5] define robustness as a secondary attribute in the following way:

> *"...dependability with respect to external faults, which characterizes a system reaction to a specific class of faults."*

When looking at the statements above, it can be argued that most of them relate to a *reaction*. The description of robustness by Avižienis et al. [5] suggests that we need to find a class of faults that the reaction refers to. In order to be able to distinguish "robust" from "reliable", we partially adopt Svensson's interpretation that robustness is about *"...exceptional conditions..."*. Consequently, we define the class of faults as *exceptional* which in HLIF may be *exceptional degrees* of:

- Uncertain, imprecise, and inconsistent information
- Lack of information
- Insufficient or inaccurate domain knowledge
- Unreliable sources

Exceptional degrees is something that is dependent on the application domain at hand, and must therefore be defined accordingly. The desired reaction to exceptional faults, from a decision maker's point of view, would be to expect that the service still fulfill reliability, i.e., correctness. Altogether, robustness in HLIF is about being able to provide a reliable service even though exceptional faults are present.

Stability

Stability is included in the definition of robustness by Svensson [36]:

> *"...avoid making large changes in recommendations as a consequence of small changes in input data."*

In order to allow for as precise and clear meaning as possible for both robustness and stability, we prefer to view stability as a separate secondary attribute to safety. It is somewhat unclear what Svensson exactly means by *"...large changes in recommendations..."*. When are recommendations of actions considered to be different from each other? One may even think of two very different recommendations that are expected to lead to essentially the same impact. Thus, we rephrase stability as: small variations in input should not cause changes in actions that are expected to lead to different impacts. Here, "expected" and "different impacts", needs to be more clearly defined, something that must be done with respect to a specific problem and application domain.

As an example, assume that one has two different sets of actions, $\mathcal{A}_1$ and $\mathcal{A}_2$, which for some reason are expected to lead to different impacts. Assume that the following holds for an impact $\mathcal{I}_i^h$:

$$\begin{aligned}\mu_{\mathcal{I}}(\mathcal{I}_i^h = \mathcal{I}|\mathcal{A}_1, \mu_{\mathcal{S}}, \xi) - \\ \mu_{\mathcal{I}}(\mathcal{I}_i^h = \mathcal{I}|\mathcal{A}_2, \mu_{\mathcal{S}}, \xi) = \kappa\end{aligned} \tag{13}$$

Let the input in our interpretation of stability refer to ξ, more concretely, let ξ constitute a sensor stream of information. Now, assume that the stream becomes noisy, i.e., small random variations in the information are present, denoted by ξ'. Such variation may cause, at a given time instant, the following equation to hold:

$$\begin{aligned}\mu_{\mathcal{I}}(\mathcal{I}_i^h = \mathcal{I}|\mathcal{A}_1, \mu_{\mathcal{S}}, \xi') - \\ \mu_{\mathcal{I}}(\mathcal{I}_i^h = \mathcal{I}|\mathcal{A}_2, \mu_{\mathcal{S}}, \xi') = \kappa'\end{aligned} \tag{14}$$

By fulfilling stability the following is prevented:

$$|\kappa - \kappa'| > \epsilon \tag{15}$$

In other words, the difference in belief is not allowed to deviate "too much" due to random variations in ξ'. Here ϵ is a parameter that quantifies an *acceptable deviation* with respect to the variation in ξ'. In the worst case, a decision maker may choose to implement $\mathcal{A}_1$ when Eq. (13) holds and $\mathcal{A}_2$ when Eq. (14) holds. Since the action sets were expected to lead to impacts that are in some sense different from each other, such behavior is considered to be "unsound". In this example, the input was the set of available information ξ, but one may equally well consider the belief measure $\mu_{\mathcal{S}}$ as the input, i.e., small variations in $\mu_{\mathcal{S}}$ should not cause the behavior defined by Eq. (15).

5 Imprecise Probability - Dependable High-Level Information Fusion

Imprecise probability [41] refers to a family of theories that allow imprecision in the belief measures, e.g., a probability interval. Walley [39, 40, 41] has argued

extensively for the importance of imprecision in probabilities and describes several different sources of it such as: *lack of information*, *conflicting information* (inconsistent information), and *conflicting beliefs* (e.g., conflict amongst a group of domain experts), to name a few of them. Lack of information is related to a specific type of uncertainty referred to as *epistemic* or *reducible* [33], since gathering more information may reduce this type of uncertainty; closely related to one of the goals of IF, i.e., reducing uncertainty. The two latter sources of imprecision in probabilities: conflicting information and conflicting beliefs, are also obvious in an IF context, since utilizing multiple sources of information typically lead to conflict.

As pointed out by Walley [39, Section 5.1.5], a significant difference between Bayesian theory and imprecise probability, is the way the amount of information is reflected in the belief measure. If little or no prior information concerning some process is available, Bayesian theory propose a *non-informative prior*, e.g., *maximum entropy* [20], while imprecise probability utilize the degree of imprecision to reflect the amount of information; substantial information implies a small interval of possible probabilities, and scarce information a large interval of possible probabilities. Thus, when utilizing imprecise probability, the information can actually be seen in the belief measure itself, while in Bayesian theory the same belief measure can be adopted before any information is available, as well as later when a large amount of information is available. Subsequently, if we consider reliability in HLIF as a correctness criterion, where the belief measure should reflect the character of the available information, even in exceptional cases when there is a severe lack of it (related to robustness), Bayesian theory cannot adequately fulfill this. From a decision maker's point of view, reliability can be thought of as a sort of "soundness" criterion, i.e., the decision maker will be aware of the quality of the belief measure.

Imprecise probabilities also allow a direct way of handling the problem of stability. Consider *Bayesian networks* (BNs) [17], a method that is commonly proposed for HLIF [6, 12, 21], where precise probabilities, usually referred to as *conditional probability tables* (CPTs) needs to be assessed from data, or elicited from a domain expert. A problem with such networks, besides assessment or elicitation of precise CPTs, is that it is necessary to perform *sensitivity analysis*, i.e., examine for chaotic behavior [37] by perturbation of the CPTs. By utilizing imprecise probability (e.g., [10, 11]) it is possible to account for "possible" values of the CPTs in a direct way. Thus, instead of assessing or eliciting precise probabilities followed by sensitivity analysis, where the CPTs are perturbed; imprecise probability account for this from the start, i.e., the imprecision constitutes "possible" probabilities that potentially could have resulted from a perturbation of precise probability.

6 Application Domains

Information fusion (IF) has its roots in the defense domain and is still tightly coupled to it. In this section, we first depict the current state of IF techniques

and applications within defense. Then, we describe two "civilian" application domains, *manufacturing* and *precision agriculture*, where researchers have started to recognize the benefits of IF, mainly through initial studies in low-level IF.

6.1 Defense

There exist a number of different IF applications within the defense domain such as: *ocean surveillance*, *air-to-air and surface-to-air defense*, *battlefield intelligence*, *surveillance*, *target acquisition*, and *strategic warning and defense* [16]. In a defense context, Level 1 – Object Assessment – concerns the problem of detecting objects and their corresponding physical attributes, e.g., vehicle type (e.g, tank or jeep), position, velocity, and heading.

The goal of Level 2 – Situation Assessment – is to establish relationships among the identified objects and events, which belong to the environment under consideration [16]. A common relation applied at this level is clustering, e.g., clustering of vehicles into *platoons*, *companies*, and *battalions* [32]. Lastly, in Level 3 – Impact Assessment – predictions are made about future situations, e.g., different threats of enemies [16].

So far in the defense domain, most of the research has concerned Level 1 – Object Assessment –, e.g., *target tracking* with multi-sensor fusion. Most attempts to HLIF in defense involves Bayesian theory and especially BNs [6, 12, 21]. Other approaches to HLIF in the defense domain are: *Dempster-Shafer theory* [32], *genetic algorithms*, *fuzzy logic*, *neural networks* [18], *case-based reasoning*, and *fuzzy belief networks* [27].

6.2 Manufacturing

A well-known problem in manufacturing is planning of *maintenance* such that the cost and risk of failure are minimized. According to Jardine et al. [19], maintenance can be divided into: *unplanned maintenance* (breakdown maintenance) and *planned maintenance*. In unplanned maintenance, utilization of a physical asset occurs until breakdown, an approach that enables for maximum amount of utilization while there are no serious failures, but on the other hand, a breakdown can potentially cause a halt in production or even more serious failures, leading to severe economic loss. In planned maintenance, a schedule is utilized for each physical asset. The advantage of this approach is that it reduces the number of breakdowns, but for a cost of decreased utilization, since maintenance is performed independently of the actual condition of the physical asset. Due to an increased complexity in machines, planned maintenance has become a costly activity [19].

Recently, there has been an increased interest in multi-sensor fusion as a means to achieve more reliable prognosis and diagnosis in maintenance [19], and researchers have started to notice the commonalities between the IF domain and manufacturing [38, 19].

6.3 Precision Agriculture

The aim of precision agriculture is to account for large *within-field spatial and temporal variations* of different *crop* and *soil* factors [34]. By measuring different *soil properties* such as texture, moisture content, nitrogen (N) content and pH [1], the field can be divided into zones that have different needs, e.g., of fertilization or pesticides. When combining a *geographical information system* with a global positioning system, each zone can be targeted, through model simulation, with its corresponding need of for instance N-fertilization, pesticides, or watering. In the case of fertilization, it is also common to utilize so called *on-the-go sensors* (e.g., radiometric sensors) where sensor readings are used as further input for fertilization calculations. Since these calculations are performed during the actual fertilization process, they need to meet certain time constraints. Precision agriculture is both economical and environmentally friendly since the purpose is to estimate the exact need for optimal yield and minimum influence on the environment [28].

7 Discussion and Future Work

While there are many articles that describe theoretical aspects of imprecise probabilities (e.g., [41]), and comparative studies of uncertainty management methods addressing "artificial" (toy) problems (e.g., [15]), the more practical aspects where such methods are implemented and evaluated in "real-world" applications seem to be to a large extent overlooked (there are exceptions, e.g., [4, 8, 29]). We believe that the only way for imprecise probability to gain recognition by researchers in HLIF in particular and other research communities in general, is to conduct experiments that actually demonstrate benefits of such approach in comparison with precise probability. We have here described three application domains which will be utilized for this purpose. The exact theories to be evaluated in the family of imprecise probability will be selected in accordance to characteristics of the specific problem in each of these application domains. A common feature of all these domains, and most IF applications, is that decisions must be made within a certain period of time, i.e., certain time constraints need to be met. Such constraints may be challenging to meet when using imprecise probability, due to the additional complexity introduced by imprecision (sets of probability measures instead of a single probability measure).

Since many attempts to address HLIF rely on Bayesian theory such as BNs; imprecise probability will naturally be evaluated against existing precise solutions. Consequently, HLIF provides an excellent evaluation ground for imprecise probability. There is also genuine need for addressing dependability issues in HLIF, an area that has been more or less neglected, or as Svensson [36] puts it:

> *"Indeed, unless concepts and methodologies are found and generally applied which enable researchers and developers to achieve and demonstrate reliability of high-level information fusion methods and algorithms, operational decision makers are unlikely to be willing to trust or use decision support systems based on such techniques."*

In a recent publication, Antonucci et al. [2] have recognized the benefits of utilizing imprecise probability – *credal networks* [10, 11] – in IF. The application of credal networks to HLIF is definitely something that should be further investigated and contrasted against BNs.

8 Conclusions

In this position paper we have reported on ongoing and planned future work on deployment and evaluation of imprecise probability in high-level information fusion (HLIF) applications. A detailed description of HLIF and a partial mapping from a dependability taxonomy to HLIF were presented. There is a need to find more dependable methods within HLIF, and imprecise probability seems to be an interesting approach to improve dependability. We have also argued that it is important to implement and evaluate imprecise probability in "real-world" applications, if such methods are going to be acknowledged. Since HLIF is an application oriented research area, where most methods are based on Bayesian theory, we have also concluded that HLIF is an excellent evaluation ground for this purpose. Three application domains: defense, manufacturing, and precision agriculture, for evaluation of imprecise probability as an approach to improved dependability in HLIF, were described. Design of experiments in these domains, which contrasts the benefits and drawbacks of imprecise probability to precise probability, is our next step, something that in itself is challenging and valuable to the dissemination of research on imprecise probability.

Acknowledgement

This work was supported by the Information Fusion Research Program (University of Skövde, Sweden) in partnership with the Swedish Knowledge Foundation under grant 2003/0104 (URL: http://www.infofusion.se).

References

1. Adamchuk, V.I., et al.: On-the-go soil sensors for precision agriculture. Computers and electronics in agriculture 44, 71–91 (2004)
2. Antonucci, A., et al.: Credal networs for military identification problems. In: 5th International Symposium on Imprecise Probability: Theories and application (2007)
3. Antony, R.: Data fusion automation: A top-down perspective. In: Hall, D.L., Llinas, J. (eds.) Handbook of Multisensor Data Fusion, pp. 6-1–6-25. CRC Press, Boca Raton (2001)
4. Aughenbaugh, J.M.: Managing uncertainty in engineering design using imprecise probabilities and principles of information economics. PhD thesis, Georgia Institute of Technology (2006)
5. Avižienis, A., et al.: Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions on dependable and secure computing 1, 11–33 (2004)
6. Bladon, P., Hall, R.J., Wright, W.A.: Situation assessment using graphical models. In: Proceedings of the 5th International Conference on Information fusion (2002)

7. Bloch, I., et al.: Fusion: General concepts and characteristics. International Journal of Intelligent Systems 16, 1107–1134 (2001)
8. Borotschnig, H., et al.: A comparison of probabilistic, possibilistic and evidence theoretic fusion schemes for active object recognition. Computing 62, 293–319 (1999)
9. Bossé, É., Guitouni, A., Valin, P.: An essay to characterise information fusion systems. In: Proceedings of the 9th International Conference on Information fusion (2006)
10. Cozman, F.G.: Credal networks. Artificial intelligence 120, 199–233 (2000)
11. Cozman, F.G.: Graphical models for imprecise probabilities. International Journal of Approximate Reasoning 39, 167–184 (2005)
12. Das, S., Lawless, D.: Trustworthy situation assessment via belief networks. In: Proceedings of the 5th International Conference on Information fusion (2002)
13. Dasarathy, B.V.: Information fusion - what, where, why, when, and how? Information Fusion 2, 75–76 (2001)
14. Devlin, K.: Logic and Information. Cambridge university press, Cambridge (1995)
15. Ferson, S., et al.: Summary from the epistemic uncertainty workshop: consensus amid diversity. Reliability Engineering & System Safety 85, 355–369 (2004)
16. Hall, D.L., Llinas, J.: Multisensor data fusion. In: Hall, D.L., Llinas, J. (eds.) Handbook of Multisensor Data Fusion, pp. 1-1–1-10. CRC Press, Boca Raton (2001)
17. Heckerman, D.: Bayesian networks for data mining. Data Mining and Knowledge Discovery 1, 79–119 (1997)
18. Hinman, M.L.: Some computational approaches for situation assessment and impact assessment. In: Proceedings of the 5th International Conference on Information fusion (2002)
19. Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing 20, 1483–1510 (2006)
20. Jaynes, E.T.: Information theory and statistical mechanics. The Physical Review 106(4), 620–630 (1957)
21. Johansson, F., Falkman, G.: Implementation and integration of a Bayesian network for prediction of tactical intention into a ground target simulator. In: Proceedings of the 9th International Conference on Information fusion (2006)
22. Karlsson, A.: Dependable and generic high-level information fusion - methods and algorithms for uncertainty management. Tech. Rep. HS-IKI-TR-07-003, School of Humanities and Informatics, University of Skövde, Sweden (2007)
23. Kokar, M.M.: Situation awareness: Issues and challenges, in panel discussion: Higher level fusion: Unsolved, difficult, and misunderstood problems/approaches in levels 2-4 fusion research. In: Proceedings of the 7th International Conference on Information fusion (2004)
24. Kokar, M.M., Matheus, C.J., Baclawski, K.: Ontology-based situation awareness. Information fusion (2007), doi:10.1016/j.inffus.2007.01.004
25. Llinas, J.: Assessing the performance of multisensor fusion processes. In: Hall, D.L., Llinas, J. (eds.) Handbook of Multisensor Data Fusion, pp. 20-1–20-18. CRC Press, Boca Raton (2001)
26. Llinas, J., et al.: Revisiting the JDL data fusion model II. In: Proceedings of the Seventh International Conference on Information Fusion (2004)
27. Looney, C.G., Liang, L.R.: Cognitive situation and threat assessment of ground battlesspaces. Information Fusion 4, 297–308 (2003)
28. Mcbratney, A., Whelan, B., Ancev, T.: Future directions of precision agriculture. Precision Agriculture 6, 7–23 (2005)

29. Roemer, M.J., Kacprzynski, G.J., Orsagh, R.F.: Assessment of data and knowledge fusion strategies for prognostics and health management. In: IEEE Aerospace Conference Proceedings (2001)
30. Rogova, G.L., Nimier, V.: Reliability in information fusion: Literature survey. In: Proceedings of the 7th International Conference on Information fusion (2004)
31. Rogova, G.L., et al.: Reasoning about situations in the early post-disaster response environment. In: Proceedings of the 9th International Conference on Information fusion (2006)
32. Schubert, J.: Evidential force aggregation. In: Proceedings of the 6th International Conference on Information Fusion (2003)
33. Soundappan, P., et al.: Comparison of evidence theory and Bayesian theory for uncertainty modeling. Reliability engineering & System safety 85, 295–311 (2004)
34. Stafford, J.V.: Implementing precision agriculture in the 21st century. Journal of Agricultural Engineering Research 76, 267–275 (2000)
35. Steinberg, A.N., Bowman, C.L., White, F.E.: Revisions to the JDL data fusion model. In: Dasarathy, B.V. (ed.) Proceedings of SPIE - Sensor fusion: Architectures, Algorithms, and Applications III, vol. 3719 (1999)
36. Svensson, P.: On reliability and trustworthiness of high-level fusion-based decision support systems: Basic concepts and possible formal methodologies. In: Proceedings of the 9th International Conference on Information fusion (2006)
37. Thöne, H., Güntzer, U., Kießling, W.: Increased robustness of Bayesian networks through probability intervals. International Journal of Approximate Reasoning 17, 37–76 (1997)
38. Vin, L.J.D., et al.: Information fusion for simulation based decision support in manufacturing. Robotics and Computer-Integrated Manufacturing 22, 429–436 (2006)
39. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, Boca Raton (1991)
40. Walley, P.: Measures of uncertainty in expert systems. Artificial Intelligence 83, 1–58 (1996)
41. Walley, P.: Towards a unified theory of imprecise probability. International Journal of Approximate Reasoning 24, 125–148 (2000)
42. Wang, P.: Confidence as higher-order uncertainty. In: 2nd International Symposium on Imprecise Probabilities and Their Applications (2001)
43. White Jr., F.E.: Data fusion Lexicon. Joint Directors of Laboratories, Technical panel for C^3, Data Fusion Sub-Panel, Naval Ovean Systems Center, San Diego (1987)

Part III

Uncertainty Modelling and Reasoning in Knowledge-Based Systems

Label Semantics as a Framework for Granular Modelling

Jonathan Lawry

Artificial Intelligence Group, Department of Engineering Mathematics
University of Bristol, BS8 1TR, United Kingdom
J.Lawry@bris.ac.uk

Summary. An alternative perspective on granular modelling is introduced where an information granule characterises the relationship between a label expression and elements in an underlying perceptual space. Label semantics is proposed as a framework for representing information granules of this kind. Mass relations and linguistic decision trees are then introduced as two types of granular models in label semantics. Finally, its shown how linguistic decision trees can be combined within an attribute hierarchy to model complex multi-level composite mappings.

1 Introduction to Granular Modelling

Fundamental to human communication is the ability to effectively describe the continuous domain of sensory perception in terms of a finite set of description labels. It is this process of *granular modelling* which permits us to process and transmit information efficiently at a suitable level of detail, to express similarity and difference between perceptual experiences and to generalize from current knowledge to new situations. Furthermore, it allows us to express information and knowledge in a way that is robust to small variations, noise and sensory aggregations in a complex multi-dimensional and evolving perceptual environment. Given these advantages, the formalization of granular models within a mathematical theory can allow for the effective modelling of complex multi-dimensional systems in such a way as to be understandable to practitioners who are not necessarily experts in formal mathematics.

The use of labels as a means of discretizing information plays a central role in granular modelling. Indeed one possible definition of an *information granule* could be in terms of the mapping between labels and domain values as follows:

> An information granule is a characterisation of the relationship between a discrete label or expression and elements of the underlying (often continuous) perceptual domain which it describes.

From this perspective crisp sets, fuzzy sets [19], random sets [10] and rough sets [11] can all correspond to information granules in that they can be used to characterise just such a relationship between a label and the elements of the

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 87–102, 2008.
springerlink.com

underlying domain. A typical form of information granule is as the extension of the concept symbolically represented by a given label. For a label L the extension of L identifies the set of domain elements to which L can be truthfully or appropriately applied. Fuzzy sets, random sets and rough sets are then mechanisms according to which gradedness, uncertainty and imprecision can respectively be introduced into the definition of concept extensions.

The above definition of information granule should be contrasted with that of Zadeh [20] who explains granularity in terms of (possibly fuzzy) clusters of points as follows:

> A granule is a clump of objects (points) which are drawn together by indistinguishability, similarity, proximity and functionality.

However, while different there are a number of clear connections between the two notions of information granule. Gärdenfors [3] introduces *conceptual spaces* as metric spaces of sensory inputs in which the extensions of concepts correspond to convex regions. Certainly from this perspective elements within the extension of a concept are indeed likely to be linked in terms of their similarity and proximity to one another. Also the functionality of an object can directly inference the way that it is labelled or classified. For example. the labelling of parts of the face as nose, mouth, ear etc is, as noted by Zadeh [20], significantly dependant on their respective functions.

Label semantics [5], [6] is a representation framework to encode the conventions for allocating labels and compound expressions generated from labels, as descriptions of elements from the underlying domain. As such it provides a useful tool for granular modelling when formulated as above with an emphasis on the association of points and labels. The notion of vagueness is also closely related to that of information granularity in that for most examples of information processing in natural language the information granules are not precisely defined. Indeed this semantic imprecision can often result in more flexible and robust granular models. Label semantics is based on an epistemic theory of vagueness [18] according to which the individual agents involved in communication believe in the existence of language conventions shared across the population of communicators but are (typically) uncertain as to which of the available labels can be appropriately used to describe any given instance.

2 Underlying Philosophy of Vagueness

In contrast to fuzzy set theory [19], *label semantics* encodes the meaning of linguistic labels according to how they are used by a population of communicating agents to convey information. From this perspective, the focus is on the decision making process an intelligent agent must go through in order to identify which labels or expressions can actually be used to describe an object or value. In other words, in order to make an assertion describing an object in terms of some set of linguistic labels, an agent must first identify which of these labels are appropriate or assertible in this context. Given the way that individuals learn language

through an ongoing process of interaction with the other communicating agents and with the environment, then we can expect there to be considerable uncertainty associated with any decisions of this kind. Furthermore, there is a subtle assumption central to the label semantic model, that such decisions regarding appropriateness or assertibility are meaningful. For instance, the fuzzy logic view is that vague descriptions like 'John is *tall*' are generally only partially true and hence it is not meaningful to consider which of a set of given labels can truthfully be used to described John's height. However, we contest that the efficacy of natural language as a means of conveying information between members of a population lies in shared conventions governing the appropriate use of words which are, at least loosely, adhere to by individuals within the population.

In our everyday use of language we are continually faced with decisions about the best way to describe objects and instances in order to convey the information we intend. For example, suppose you are witness to a robbery, how should you describe the robber so that police on patrol in the streets will have the best chance of spotting him? You will have certain labels that can be applied, for example *tall, short, medium, fat, thin, blonde, etc*, some of which you may view as inappropriate for the robber, others perhaps you think are definitely appropriate while for some labels you are uncertain whether they are appropriate or not. On the other hand, perhaps you have some ordered preferences between labels so that *tall* is more appropriate than *medium* which is in turn more appropriate than *short*. Your choice of words to describe the robber should surely then be based on these judgements about the appropriateness of labels. Yet where does this knowledge come from and more fundamentally what does it actually mean to say that a label is or is not appropriate? Label semantics proposes an interpretation of vague description labels based on a particular notion of appropriateness and suggests a measure of subjective uncertainty resulting from an agent's partial knowledge about what labels are appropriate to assert. Furthermore, it is suggested that the vagueness of these description labels lies fundamentally in the uncertainty about if and when they are appropriate as governed by the rules and conventions of language use. The underlying assumption here is that some things can be correctly asserted while others cannot. Exactly where the dividing line lies between those labels that are and those that are not appropriate to use may be uncertain, but the assumption that such a division exists would be a natural precursor to any decision making process of the kind just described.

The above argument is very close to the epistemic view of vagueness as expounded by Timothy Williamson [18]. Williamson assumes that for the extensions of a vague concept there is a precise but unknown dividing boundary between it and the extension of the negation of that concept. However, while there are marked similarities between the epistemic theory and the label semantics view, there are also some subtle differences. For instance, the epistemic view would seem to assume the existence of some objectively correct, but unknown, definition of a vague concept. Instead of this we argue that individuals when faced with decision problems regarding assertions find it useful as part of a decision making strategy to assume that there is a clear dividing line

between those labels which are and those which are not appropriate to describe a given instance. We refer to this strategic assumption across a population of communicating agents as the *epistemic stance*, a concise statement of which is as follows:

> *Each individual agent in the population assumes the existence of a set of labelling conventions, valid across the whole population, governing what linguistic labels and expressions can be appropriately used to describe particular instances.*

In practice these rules and conventions underlying the appropriate use of labels would not be imposed by some outside authority. In fact, they may not exist at all in a formal sense. Rather they are represented as a distributed body of knowledge concerning the assertability of predicates in various cases, shared across a population of agents, and emerging as the result of interactions and communications between individual agents all adopting the epistemic stance. The idea is that the learning processes of individual agents, all sharing the fundamental aim of understanding how words can be appropriately used to communicate information, will eventually converge to some degree on a set of shared conventions. The very process of convergence then to some extent vindicates the epistemic stance from the perspective of individual agents. Of course, this is not to suggest complete or even extensive agreement between individuals as to these appropriateness conventions. However, the overlap between agents should be sufficient to ensure the effective transfer of useful information.

A further distinction between our view of appropriateness and the epistemic view of Williamson can be found in the local, or instance-based, nature of appropriateness judgements. Arguments in favour of the epistemic view concern the existence of a precise boundary between the extension of a concept and that of its negation. The appropriateness of labels on the other hand is judged with reference to a particular instance. From this perspective it is unlikely that agents would generate an explicit representation of the extension of a vague concept. Instead their knowledge would be based on previous experience of assertions about similar instances from a range of other agents and a subsequent process of interpolation between these examples. In most cases decision problems about assertions would then typically concern a particular instance, so that the problem of identifying concept boundaries would not be directly considered.

The epistemic stance allows agents to meaningfully apply epistemic models of uncertainty to quantify their subjective belief in whether certain labels are appropriate. In the sequel we will introduce two related probabilistic measures of an agent's uncertainty concerning the appropriateness of vague expressions and explore the resulting calculus.

3 Label Semantics

Label semantics proposes two fundamental and inter-related measures of the appropriateness of labels as descriptions of an object or value. We begin by

assuming that for all agents there is a fixed shared vocabulary in the form of a finite set of basic labels LA for describing elements from the underlying universe Ω. These are building blocks for more complex compound expressions which can then also be used as descriptors as follows. A countably infinite set of expressions LE can be generated through recursive applications of logical connectives to the basic labels in LA. So for example, if Ω is the set of all possible rgb values and LA is the set of basic colour labels such as *red*, *yellow* ,*green*, *orange* etc then LE contains those compound expressions such as *red* & *yellow*, *not blue nor orange* etc. The measure of appropriateness of an expression $\theta \in LE$ as a description of instance x is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective probability that θ can be appropriately used to describe x. From an alternative perspective, when faced with describing instance x, an agent may consider each label in LA and attempt to identify the subset of labels that are appropriate to use. This is a totally meaningful endeavour for agents who adopt the epistemic stance. Let this complete set of appropriate labels for x be denote by $\mathcal{D}_x$. In the face of their uncertainty regarding labelling conventions agents will also be uncertain as to the composition of $\mathcal{D}_x$, and we represent this uncertainty with a probability mass function $m_x : 2^{LA} \to [0, 1]$ defined on subsets of labels. Hence, for the subset of labels $\{red, orange, yellow\}$ and rgb value x, $m_x(\{red, orange, yellow\})$ denotes the subjective probability that $\mathcal{D}_x = \{red, orange, yellow\}$, or in other words that $\{red, orange, yellow\}$ is the complete set of basic colour labels with which it is appropriate to describe x. We now provide formal definitions for the set of expressions LE and for mass functions m_x, following which we will propose a link between the two measures $\mu_\theta(x)$ and m_x for expression $\theta \in LE$.

Definition 1. *Label Expressions*
The set of label expressions LE generated from LA, is defined recursively as follows:

- *If $L \in LA$ then $L \in LE$*
- *If $\theta, \varphi \in LE$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi \in LE$.*

Definition 2. *Mass Function on Labels*
$\forall x \in \Omega$ a mass function on labels is a function $m_x : 2^{LA} \to [0, 1]$ such that $\sum_{S \subseteq LA} m_x(S) = 1$.

Note that there is no requirement for the mass associated with the empty set to be zero. Instead, $m_x(\emptyset)$ quantifies the agent's belief that none of the labels are appropriate to describe x. We might observe that this phenomena occurs frequently in natural language, especially when labelling perceptions generated along some continuum. For example, we occasionally encounter colours for which none of our available colour descriptors seem appropriate. Hence, the value $m_x(\emptyset)$ is an indicator of the describability of x in terms of the labels LA.

Now depending on labelling conventions there may be certain combinations of labels which cannot all be appropriate to describe any object. For example, *small* and *large* cannot both be appropriate. This restricts the possible values of $\mathcal{D}_x$ to the following set of focal elements:

Definition 3. *Set of Focal Elements*
Given labels LA together with associated mass assignment $m_x : \forall x \in \Omega$, the set of focal elements for LA is given by $\mathcal{F} = \{S \subseteq LA : \exists x \in \Omega,\ m_x(S) > 0\}$.

The link between the mass function m_x and the appropriateness measures $\mu_\theta(x)$ is motivated by the intuition that the assertion 'x is θ' directly provides information dependent on θ, as to what are the possible values for $\mathcal{D}_x$. For example, the assertion 'x is *blue*' would mean that *blue* is an appropriate label for x, from which we can infer that $blue \in \mathcal{D}_x$. Similarly, the assertion 'x is *green and not blue*' would mean that *green* is an appropriate label for x while *blue* is not, so that we can infer $green \in \mathcal{D}_x$ and $blue \notin \mathcal{D}_x$. Another way of expressing this information is to say that $\mathcal{D}_x$ must be a member of the set of sets of labels which contain *green* but do not contain *blue* i.e. $\mathcal{D}_x \in \{S \subseteq LA : green \in S,\ blue \notin S\}$. More generally, we can define a functional mapping λ from LE into $2^{2^{LA}}$ (i.e. the set containing all possible sets of label sets) for which the assertion 'x is θ' enables us to infer that $\mathcal{D}_x \in \lambda(\theta)$. This mapping is defined recursively as follows:

Definition 4. *λ-mapping*
$\lambda : LE \rightarrow 2^{\mathcal{F}}$ is defined recursively as follows: $\forall \theta,\ \varphi \in LE$

- $\forall L \in LA\ \lambda(L) = \{S \in \mathcal{F} : L \in S\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$
- $\lambda(\neg\theta) = \lambda(\theta)^c$.

The λ-mapping then provides us with a means of evaluating the appropriateness measure of an expression θ directly from m_x, as corresponding to the subjective probability that $\mathcal{D}_x \in \lambda(\theta)$ so that:

Definition 5. *Appropriateness Measures*
For any expression $\theta \in LE$ and $x \in \Omega$, the appropriateness measure $\mu_\theta(x)$ can be determined from the mass function m_x according to:

$$\forall \theta \in LE\ \mu_\theta(x) = \sum_{S \in \lambda(\theta)} m_x(S).$$

From this relationship the following list of general properties hold for expressions θ and φ in LE [5]:

Theorem 1. *Lawry [5],[6]*

- *If $\theta \models \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) \leq \mu_\varphi(x)$*
- *If $\theta \equiv \varphi$ then $\forall x \in \Omega\ \mu_\theta(x) = \mu_\varphi(x)$*
- *If θ is a tautology then $\forall x \in \Omega\ \mu_\theta(x) = 1$*
- *If θ is a contradiction then $\forall x \in \Omega\ \mu_\theta(x) = 0$*
- $\forall x \in \Omega\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$.

Notice, here that the laws of excluded middle and non-contradiction are preserved since for any expression θ, $\lambda(\theta \vee \neg\theta) = \lambda(\theta) \cup \lambda(\theta)^c = 2^{2^{LA}}$ and

$\lambda(\theta \wedge \neg\theta) = \lambda(\theta) \cap \lambda(\theta)^c = \emptyset$. Also, the idempotent condition holds since $\lambda(\theta \wedge \theta) = \lambda(\theta) \cap \lambda(\theta) = \lambda(\theta)$.

The λ-mapping provides us with a clear formal representation for linguistic constraints, where the imprecise constraint 'x is θ' on x is interpreted as the precise constraint $\mathcal{D}_x \in \lambda(\theta)$ on $\mathcal{D}_x$.

3.1 Ordering Labels

As discussed above an agent's estimation of both m_x and $\mu_\theta(x)$ should depend on their experience of language use involving examples similar to x. Clearly the form of this knowledge is likely to be both varied and complex. However, one natural type of assessment for an agent to make would be to order or rank label in terms of their estimated appropriateness for x. This order information could then be combined with estimates of appropriateness measure values for the basic labels (i.e. elements of LA) in order to provide estimates of values for compound expressions (i.e. elements of LE). Hence we assume that:

> *An agent's knowledge of label appropriateness for an instance x, can be represented by an ordering on the basic labels LA and an allocation of uncertainty values to the labels consistent with this ordering.*

Effectively we are assuming that through a process of extrapolation from experience agents are, for a given instance, able to (at least partially) rank labels in terms of their appropriateness and then, consistent with this ranking, to estimate a subjective probability that each label is appropriate. On the basis of both the ordering and probability assignment to basic labels the agent should then be able to evaluate the appropriateness measure of more complex compound expressions. The ranking of available labels would seem to be an intuitive first step for an agent to take when faced with the decision problem about what to assert. Also, the direct allocation of probabilities to a range of complex compound expressions so that the values are internally consistent is a fundamentally difficult task. Hence, restricting such evaluations to only the basic labels would have significant practical advantages in terms of computational complexity.

Definition 6. *(Ordering on Labels)*
For $x \in \Omega$ let $\preceq_x$ be an ordering on LA such that for $L, L' \in LA$, $L' \preceq_x L$ means that L is at least as appropriate as a label for x as L'.

The identification by an agent of an ordering on labels $\preceq_x$ for a particular $x \in \Omega$ (as in definition 6), restricts the possible label sets which they can then consistently allocate to $\mathcal{D}_x$. For instance, $L' \preceq_x L$ then this implies that if $L' \in \mathcal{D}_x$ then so is $L \in \mathcal{D}_x$, since L is as least as appropriate a description for x as L'. Hence, given $\preceq_x$ for which $L' \preceq_x L$ it must hold that $m_x(S) = 0$ for all $S \subseteq LA$ where $L' \in S$ and $L \notin S$. Trivially, from definition 5 this also means that $\mu_{L'}(x) \leq \mu_L(x)$. Given these observations an important question is whether the information provided by ordering $\preceq_x$ together with a set of appropriateness values $\mu_L(x) : L \in LA$ for the basic labels, consistent with $\preceq_x$, is sufficiently to specify a unique mass function m_x? Notice that in the label

semantics framework the identification of a unique mass function m_x in this way immediately enables the agent to apply definition 5 in order to evaluate the appropriateness $\mu_\theta(x)$ of any compound expression θ from the appropriateness measure values for the basis labels. In fact, in the case that $\preceq_x$ is a total (linear) ordering it is not difficult to see that such a unique mapping does indeed exist between the mass function and the appropriateness measures of basic labels. To see this suppose that we index the labels in LA so that $L_n \preceq_x L_{n-1} \preceq_x \ldots \preceq_x L_1$ with corresponding appropriateness measures $\mu_{L_n}(x) = a_n \leq \mu_{L_{n-1}}(x) = a_{n-1} \leq \ldots \leq \mu_{L_1}(x) = a_1$. Now from the above discussion we have that in this case the only possible values for $\mathcal{D}_x$ are from the nested sequence of sets $\emptyset, \{L_1\}, \{L_1, L_2\}, \ldots, \{L_1, \ldots, L_i\}, \ldots, \{L_1, \ldots, L_n\}$. This together with the constraints imposed by definition 5 that for each label $a_i = \mu_{L_i}(x) = \sum_{S:L_i \in S} m_x(S)$ results in the following unique mass function:

$$m_x := \{L_1, \ldots, L_n\} : a_n, \ldots, \{L_1, \ldots, L_i\} : a_i - a_{i+1},$$
$$\ldots, \{L_1\} : a_1 - a_2, \emptyset : 1 - a_1$$

Hence, for $\preceq_x$ a total ordering we see that $\mu_\theta(x)$ can be determined as a function of the appropriateness measure values $\mu_L(x) : L \in LA$ on the basic labels. For an expression $\theta \in LE$, this function is a composition of the above mapping, in order to determine a unique mass function, and the consequent summing of mass function values across $\lambda(\theta)$, as given in definition 5, to evaluate $\mu_\theta(x)$. Although functional in this case, the calculus for appropriateness measures cannot be truth-functional in the sense of fuzzy logic since appropriateness measures satisfy all the classical Boolean laws and a well known result due to Dubois and Prade [2] shows that no truth-functional calculus can in general preserve all such laws. For a more detailed discussion of the difference between functionality and truth-functionality see Lawry [6]. The following theorem shows that in the case where $\preceq_x$ is a total ordering the max and min combination rules can be applied in certain restricted cases:

Theorem 2. *[5, 16]*
Let $LE^{\wedge,\vee} \subseteq LE$ denote those expressions generated recursively from LA using only the connectives $\wedge$ and $\vee$. If the appropriateness measures on basic labels are consistent with a total ordering $\preceq_x$ on LA then $\forall \theta, \varphi \in LE^{\wedge,\vee}$ it holds that:

$$\mu_{\theta\wedge\varphi}(x) = \min\left(\mu_\theta(x), \mu_\varphi(x)\right), \ \ \mu_{\theta\vee\varphi}(x) = \max\left(\mu_\theta(x), \mu_\varphi(x)\right).$$

In the case that $\preceq_x$ is only a partial ordering on LA then in general this does not provide the agent with sufficient information to determine a unique mass function from the appropriateness measure values on the basic labels. Instead, further information is required for the agent to evaluate a mass function and consequently the appropriateness of compound label expressions. In Lawry [7] it is proposed that this additional information takes the form of conditional independence constraints imposed by a Bayesian network generated by $\preceq_x$. These additional assumptions are then sufficient to determine m_x uniquely. Details of

this approach, however, are beyond the scope of this current paper. Instead in the examples presented in the sequel we will assume that the ordering $\preceq_x$ is total.

4 Granular Models in Label Semantics

In label semantics information granules correspond to appropriateness measure for a fixed expression i.e. an information granule is a function $\mu_\theta : \Omega \to [0,1]$ for some label expression $\theta \in LE$. In fact, within the scope of this definition we can also use mass functions to represent information granules. Specifically, for a fixed focal set $F \subseteq LA$ an information granule may also be represented by the function corresponding to the values of $m_x(F)$ as x varies across Ω. To see this notice that the value $m_x(F)$ can also be represented by the appropriateness measure $\mu_{\alpha_F}(x)$ where $\alpha_F = (\bigwedge_{L \in F} L) \wedge (\bigwedge_{L \notin F} \neg L)$ is the label expression stating that all and only the labels in F are appropriate. Hence, for a focal set $F \subseteq LA$ the corresponding information granule is the function $\mu_{\alpha_F} : \Omega \to [0,1]$. For example, in figure 1, information granules are defined in terms of the appropriateness measures for labels *low*, *medium* and *high*, represented by trapezoidal functions of $x \in \Omega = [0, 30]$. Assuming a total ordering $\preceq_x$ on labels for all $x \in \Omega$ results in mass functions m_x for the focal sets $\mathcal{F} = \{\{l\}, \{l, m\}, \{m\}, \{m, h\}, \{h\}\}$, shown as triangular functions in figure 2. These triangular functions then correspond to the information granules generated by the focal sets in $\mathcal{F}$. The direct use of focal sets as information granules in granular models can in some cases allow for more straightforward information processing. In particular, note that the mass function m_x defines a probability distribution on $\mathcal{D}_x$ which can in turn make it relatively straightforward to evaluate probability values from a granular model based on such functions.

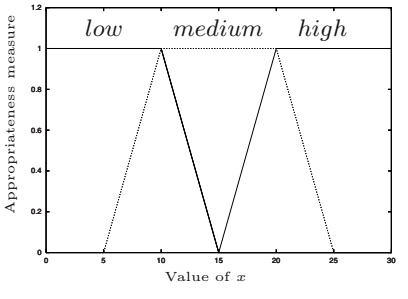


Fig. 1. Appropriateness measure values for labels *low*, *medium*, *high* viewed as a function of x as x varies across $\Omega = [0, 30]$

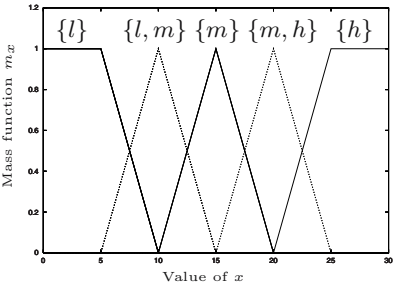


Fig. 2. Mass function values for the sets $\{low\}$, $\{low, medium\}$, $\{medium\}$, $\{medium, high\}$, $\{high\}$ viewed as a function of x as x varies across $\Omega = [0, 30]$

Consider the following formalization of a simple modelling problem: Given attributes $x_1, \ldots, x_{k+1}$ with universes $\Omega_1, \ldots, \Omega_{k+1}$ suppose that x_{k+1} is dependent on $x_1, \ldots, x_k$ according to some functional mapping $g : \Omega_1 \times \ldots \times \Omega_k \rightarrow \Omega_{k+1}$ (i.e. $x_{k+1} = g(x_1, \ldots, x_k)$). In the case that Ω_{k+1} is finite then this is referred to as a classification problem whereas if Ω_{k+1} is an infinite subset of $\mathbb{R}$ (typically a closed interval) then it is referred to as a prediction or regression problem. For a learning problem, information regarding this function is then provided by a training database containing vectors of input values together with their associated output. Let this database be denoted by $DB = \{\langle x_1(i), \ldots, x_k(i), x_{k+1}(i)\rangle : i = 1, \ldots, N\}$. For a more general modelling problem information on g may take a variety of forms including qualitative information elicited from domain experts.

Label semantics can be used to represent linguistic rule based models which provide an approximation $\hat{g}$ of the underlying function mapping g. Here we consider two such models; mass relations and linguistic decision trees. For both these approaches we use appropriateness measures to define a set of labels describing each attribute $LA_j : j = 1, \ldots, k+1$ with associated label expressions $LE_j : j = 1, \ldots, k+1$ and focal sets $\mathcal{F}_j : j = 1, \ldots, k+1$. We will also describe how these models can be used within a hierarchical structure to provide a decomposed model for high-dimensional mappings.

4.1 Mass Relational Models

If we consider the problem of describing an object or instance on the basis of k attributes $x_1, \ldots, x_k$ then we need to jointly quantify the appropriateness of labels in each of the associated labels sets $LA_j : j = 1, \ldots, k$ to describe each attribute. In other words, we need to define a joint mass function on $\mathcal{D}_{x_1} \times \ldots \times \mathcal{D}_{x_k}$ mapping from $\mathcal{F}_1 \times \ldots \times \mathcal{F}_k$ into $[0, 1]$. We refer to such joint mass functions on label sets as a mass relations. Mass relations can be used to represent a granular model of the function g. Typically, this is achieved by defining a mass relation between input focal sets conditional on each of the output focal sets in $\mathcal{F}_{k+1}$. Together these can then be used to infer a mass functions on output focal sets given a vector of input attribute values.

Definition 7. *Mass Relations*
A mass relation is a conditional function $m : \mathcal{F}_1 \times \ldots \times \mathcal{F}_k \rightarrow [0, 1]$ such that for $F_i \in \mathcal{F}_i : i = 1, \ldots, k+1$, $m(F_1, \ldots, F_k|F_{k+1})$ is the conditional joint mass function value of the input focal sets $F_1, \ldots, F_k$ given output focal set F_{k+1}. This can be evaluated from a database DB according to:

$$m(F_1, \ldots, F_k|F_{k+1}) = \frac{\sum_{i \in DB} \prod_{j=1}^{k+1} m_{x_j(i)}(F_j)}{\sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1})}$$

A set of mass relations conditional on each of the output focal sets in $\mathcal{F}_{k+1}$ generates a set of weighted rules of the form:

$$(\mathcal{D}_{x_1} = F_1) \wedge \ldots \wedge (\mathcal{D}_{x_k} = F_k) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : w \text{ where}$$
$$w = m(F_{k+1}|F_1, \ldots, F_k) = \frac{m(F_1, \ldots, F_k|F_{k+1})m(F_{k+1})}{\sum_{F_{k+1}} m(F_1, \ldots, F_k|F_{k+1})m(F_{k+1})}$$
$$\text{and } m(F_{k+1}) = \frac{1}{N} \sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1}).$$

Given a vector of input values $\mathbf{x} = \langle x_1, \ldots, x_k \rangle$ we can use Jeffrey's rule [4] to determine a mass function on the output focal sets $\mathcal{F}_{k+1}$ from a mass relation between $\mathcal{F}_1 \times \ldots \times \mathcal{F}_k$ and $\mathcal{F}_{k+1}$, as follows:

$$\forall F_{k+1} \in \mathcal{F}_{k+1}\ m(F_{k+1}|\mathbf{x}) =$$
$$\sum_{F_1 \in \mathcal{F}_1} \ldots \sum_{F_k \in \mathcal{F}_k} m(F_{k+1}|F_1, \ldots, F_k) m_{\mathbf{x}}(F_1, \ldots, F_k) \text{ where}$$
$$m_{\mathbf{x}}(F_1, \ldots, F_k) = \prod_{i=1}^{k} m_{x_i}(F_i)$$

In practice it can be computationally expensive to calculate the mass relation exactly and typically we need to use some form of approximation. One of the simplest is to assume conditional independence between $\mathcal{D}_{x_1}, \ldots \mathcal{D}_{x_k}$ given the values of $\mathcal{D}_{x_{k+1}}$. In this case:

$$m(F_1, \ldots, F_k|F_{k+1}) = \prod_{i=1}^{k} m(F_i|F_{k+1})$$

An extension to this approach involves searching for dependency groupings amongst the attributes and assume conditional independence (given F_{k+1}) between these groups (see [14] for details).

Recent applications of mass relations have focussed on the area of flood prediction where they have been used to model river flow [13] and also tidal surges up to the Thames barrier in London [15].

4.2 Linguistic Decision Trees

A linguistic decision tree is a decision tree with attributes as nodes and linguistic descriptions of attributes as branches. Also associated with each branch, there is a mass function over the output focal sets.

Definition 8. *Linguistic Decision Trees (LDT)*
A linguistic decision tree is a decision tree where the nodes are attributes from $x_1, \ldots, x_k$ and the edges are label expressions describing each attribute. More formally, supposing that the j'th node at depth d is the attribute x_{j_d} then there is a set of label expressions $\mathcal{L}_{j,d} \subseteq LE_i$ forming the edges from x_{j_d} such that $\lambda(\bigvee_{\theta \in \mathcal{L}_{j,d}} \theta) \supseteq \mathcal{F}_{j_d}$ and $\forall \theta, \varphi \in \mathcal{L}_{j,d}\ \lambda(\theta \wedge \varphi) \cap \mathcal{F}_{j_d} = \emptyset$. Also a branch B from a LDT consists of a sequence of expressions $\varphi_1, \ldots, \varphi_m$ where $\varphi_d \in \mathcal{L}_{j,d}$ for some

$j \in \mathbb{N}$ for $d = 1, \ldots, m$, augmented by a conditional mass value $m(F_{k+1}|B)$ for every output focal set $F_{k+1} \in \mathcal{F}_{k+1}$. Hence, every branch B encodes a set of weighted linguistic rules of the form:

$$(x_{j_1} \text{ is } \varphi_1) \wedge \ldots \wedge (x_{j_m} \text{ is } \varphi_m) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B)$$
where x_{j_d} is a depth d attribute node.

Also the mass assignment value $m(F_{k+1}|B)$ can be determined from DB according to:

$$m(F_{k+1}|B) = \frac{\sum_{i \in DB} m_{x_{k+1}(i)}(F_{k+1}) \prod_{d=1}^{m} \mu_{\varphi_d}(x_{j_d}(i))}{\sum_{i \in DB} \prod_{d=1}^{m} \mu_{\varphi_d}(x_{j_d}(i))}$$

Notice that a branch of a linguistic decision tree can be rewritten using the λ-function so that it refers only to constraints on $\mathcal{D}_{x_i} : i = 1, \ldots, k$. This means that the rules generated by LDT branches are a more general form of the rules generated by mass relations. For example, the branch rule

$$(x_{j_1} \text{ is } \varphi_1) \wedge \ldots \wedge (x_{j_m} \text{ is } \varphi_m) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B)$$
can be rewritten as
$$(\mathcal{D}_{x_{j_1}} \in \lambda(\varphi_1)) \wedge \ldots \wedge (\mathcal{D}_{x_{j_m}} \in \lambda(\varphi_m)) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B)$$

Given a vector of input attribute values $\mathbf{x} = \langle x_1, \ldots, x_k \rangle$ we can use a LDT to determine a mass function on output focal sets as follows: Suppose the LDT has branches $B_1, \ldots, B_t$ each with an associated mass function $m(\bullet|B_j) : j = 1, \ldots, t$

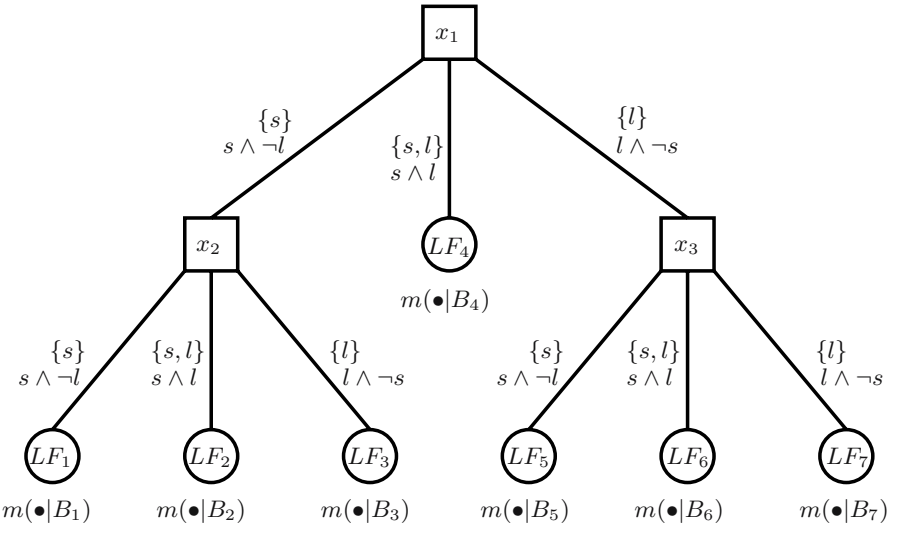


Fig. 3. An example of a linguistic decision tree

on $\mathcal{F}_{k+1}$. By applying Jeffrey's rule we obtain an aggregated mass function on $\mathcal{F}_{k+1}$ for a given input vector $\mathbf{x}$ according to:

$$m(F_{k+1}|\mathbf{x}) = \sum_{j=1}^{t} m(F_{k+1}|B_j)P(B_j|\mathbf{x}) \text{ where if}$$

$$B_j = (x_{j_1} \text{ is } \varphi_1) \wedge \ldots \wedge (x_{j_m} \text{ is } \varphi_m) \text{ then } P(B_j|\mathbf{x}) = \prod_{d=1}^{m} \mu_{\varphi_d}(x_{j_d})$$

Notice that from a computational viewpoint the output mass function above is determined from the LDT as a function of the input masses $m_{x_1}, \ldots, m_{x_k}$ only. From this perspective a LDT can be viewed as a function mapping from mass functions on the input attribute labels to a mass function on the output labels.

$$m(\bullet|\mathbf{x}) = LDT(m_{x_1}, \ldots, m_{x_k})$$

Figure 3 shows an example of a simple linguistic decision tree involving only three input attributes x_1, x_2 and x_3. Each of these is described by the same set of two overlapping labels $LA = \{small, large\}$, so that we have focal sets $\{\{s\}, \{s, l\}, \{l\}\}$. Each branch in figure 3 is labelled with its associated linguistic expressions together with their corresponding representation in terms of focal sets. In this case all the linguistic expressions involved are atomic in nature so that their λ-mappings contain only one focal set. For example, the branch B_2 ending in the leaf node LF_2 encodes the follow rules: For all $F_{k+1} \in \mathcal{F}_{k+1}$,

$$(x_1 \text{ is } s \wedge \neg l) \wedge (x_2 \text{ is } s \wedge l) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B_2)$$

Now $\lambda(s \wedge \neg l) = \{\{s\}\}$ and $\lambda(s \wedge l) = \{\{s, l\}\}$ and hence the above rule has the following focal set representation:

$$(\mathcal{D}_{x_1} = \{s\}) \wedge (\mathcal{D}_{x_2} = \{s, l\}) \rightarrow (\mathcal{D}_{x_{k+1}} = F_{k+1}) : m(F_{k+1}|B_2)$$

Linguistic decision trees can be learnt from data using the LID3 algorithm [12]. This is an extension of ID3 to allow for the type of calculations on mass functions required for a LDT. Recent application of LID3 include classification of radar images [9] and online path planning [17].

4.3 Linguistic Attribute Hierarchies

In many cases the function g is complex and it is difficult to define $\hat{g}$ as a direct mapping between $x_1, \ldots, x_k$ and x_{k+1}. Attribute hierarchies [1] are a well known approach to this problem and involve breaking down the function g into a hierarchy of sub-functions each representing a new intermediate attribute. A bottom-up description of this process is as follows: The set of original attributes $\{x_1, \ldots, x_k\}$ are partitioned into attribute subsets $S_1, \ldots, S_m$ and new attributes $z_1, \ldots z_m$ are defined as functions of each partition set respectively, so that $z_i = G_i(S_i)$ for $i = 1, \ldots, m$. The function g is then defined as a new function F of

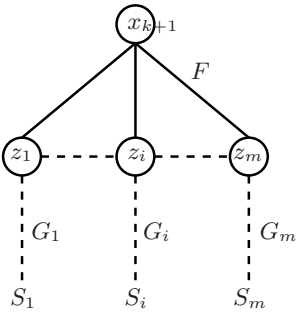


Fig. 4. Attribute hierarchy showing partition of attributes

the new attributes $z_1, \ldots z_m$, so that $x_{k+1} = g(x_1, \ldots, x_k) = F(z_1, \ldots, z_m) = F(G_1(S_1), \ldots, F(G_m(S_m))$. The same process can then be repeated recursively for each partition set S_i, to generate a new layer of new variables as required.

The identification of attribute hierarchies and their associated functional mappings is often a highly subjective process involving significant uncertainty and imprecision. Hence, the relationship between certain levels in the hierarchy, can best be described in terms of linguistic rules and relations. This can allow for judgements and rankings to be made at a level of granularity appropriate to the level of precision at which the functional mappings can be realistically defined. In linguistic attribute hierarchies the functional mappings between parent and child attribute nodes in the attribute hierarchy are defined in terms of weighted linguistic rules (typically linguistic decision trees) which explicitly model both the uncertainty and vagueness which often characterises our knowledge of such aggregation functions.

In linguistic attribute hierarchies the functional relationship between child and parent nodes are not defined precisely. Instead the labels for a parent attribute are defined in terms of the labels describing the attributes corresponding to its child nodes, by means of a linguistic decision tree. To illustrate this idea consider the following simple linguistic attribute hierarchy as shown in figure 5. Here we have 4 input attributes $x_1, \ldots, x_4$ and output attribute x_5, these being described by label sets $LA_1, \ldots, LA_5$ with focal sets $\mathcal{F}_1, \ldots, \mathcal{F}_5$ respectively. The labels for x_5 are defined in terms of the labels for two intermediate level attributes z_1 and z_2 by a linguistic decision tree LDT_1. Let LA_{z_1}, LA_{z_2} and $\mathcal{F}_{z_1}$, $\mathcal{F}_{z_2}$ be the labels and focal sets for z_1 and z_2 respectively. Furthermore, the labels for z_1 are defined in terms of those for x_1 and x_2 according to linguistic decision tree LDT_2, and the labels for z_2 are defined in terms of those for x_3 and x_4 according to linguistic decision tree LDT_3. Information is then propagated up through the hierarchy as mass functions on the relevant focal sets. Specifically, LDT_2 combines mass functions on $\mathcal{F}_1$ and $\mathcal{F}_2$ in order to generate a mass function on $\mathcal{F}_{z_1}$. Similarly mass functions on $\mathcal{F}_3$ and $\mathcal{F}_4$ are combined using LDT_3 to generate a mass function on $\mathcal{F}_{z_2}$. These two mass functions on $\mathcal{F}_{z_1}$ and $\mathcal{F}_{z_2}$ respectively are then combined according to LDT_1 in order to obtain a mass assignment on the output

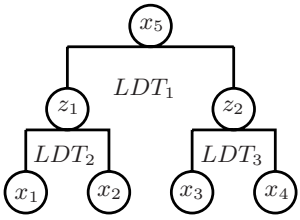


Fig. 5. Example of a simple linguistic attribute hierarchy

focal sets $\mathcal{F}_5$ conditional on the inputs. At the mass function level the complete mapping is:

$$m(\bullet|x_1,\ldots,x_4) = LDT_1(LDT_2(m_{x_1},m_{x_2}),LDT_3(m_{x_3},m_{x_4}))$$

5 Conclusions

An alternative perspective on granular modelling has been introduced where information granules encode the relationship between description labels and the underlying perceptual domain. Label semantics has been introduced as a framework for modelling linguistic vagueness and granularity. In this framework information granules are appropriateness measures and mass functions which quantify the appropriateness of label expressions to describe elements from the underlying universe. Two types of granular models have been described; mass relations and linguistic decision trees. These encode the relationship between labels on input values and those on the output values in an imprecisely defined functional mapping. In addition, we have shown how linguistic decision trees can be used as part of attribute hierarchies to combine information in complex multi-level composite mappings.

References

1. Bohanec, M., Zupan, B.: A Function-Decomposition Method for Development of Hierarchical Multi-Attribute Decision Models. Decision Support Systems 36, 215–223 (2004)
2. Dubois, D., Prade, H.: An Introduction to Possibility and Fuzzy Logics. In: Smets, P., et al. (eds.) Non-Standard Logics for Automated Reasoning, pp. 742–755. Academic Press, London (1988)
3. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge (2000)
4. Jeffrey, R.C.: The Logic of Decision. Gordon and Breach, New York (1965)
5. Lawry, J.: A framework for linguistic modelling. Artificial Intelligence 155, 1–39 (2004)
6. Lawry, J.: Modelling and Reasoning with Vague Concepts. Springer, Heidelberg (2006)

7. Lawry, J.: Appropriateness measures: An uncertainty model for vague concepts, Synthese (to appear, 2007)
8. Lawry, J., He, H.: Linguistic Attribute Hierarchies for Multiple-Attribute Decision Making. In: Proceedings 2007 IEEE International Conference on Fuzzy Systems FUZZ-IEEE 2007 (2007)
9. McCulloch, D.R., et al.: Classification of Weather Radar Images using Linguistic Decision Trees with Conditional Labelling. In: Proceedings 2007 IEEE International Conference on Fuzzy Systems FUZZ-IEEE 2007 (2007)
10. Nguyen, H.T.: On Modeling of linguistic Information using Random Sets. Information Science 34, 265–274 (1984)
11. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing, Dordrecht (1991)
12. Qin, Z., Lawry, J.: Decision Tree Learning with Fuzzy Labels. Information Sciences 172, 91–129 (2005)
13. Randon, N., et al.: River Flow Modelling Based on Fuzzy Labels. In: Information Processing and Management of Uncertainty IPMU 2004 (July 2004)
14. Randon, N., Lawry, J.: Classification and Query Evaluation using Modelling with Words. Information Sciences 176, 438–464 (2006)
15. Randon, N., et al.: Fuzzy Bayesian Modelling of Sea-Level along the East Coast of Britain. IEEE Transaction on Fuzzy Systems (to appear, 2007)
16. Tang, Y., Zheng, J.: Linguistic Modelling based on Semantic Similarity Relation amongst Linguistic Labels. Fuzzy Sets and Systems 157, 1662–1673 (2006)
17. Turnbull, O., et al.: Fuzzy Decision Tree Cloning of Flight Trajectory Optimisation for Rapid Path Planning. In: Proceedings 45 IEEE Conference on Decision and Control (2006)
18. Williamson, T.: Vagueness, Routledge (1994)
19. Zadeh, L.A.: Fuzzy Sets. Information and Control 8(3), 338–353 (1965)
20. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90, 111–127 (1997)

Approximating Reasoning for Fuzzy-Based Information Retrieval

Tho T. Quan and Tru H. Cao

Faculty of Computer Science and Engineering
Hochiminh City University of Technology, Vietnam
{qttho,tru}@cse.hcmut.edu.vn

Summary. Among modern applications of document retrieval, a practical system for retrieving scientific publications has recently been attracting much attention from research community. In a scientific document, there are many types of uncertainty information occurring, such as research areas of the documents or authors. Thus, a method for efficiently handling uncertainty information when retrieving scientific information, as well as other kinds of uncertainty information, is currently desirable. In our paper, we propose a novel fuzzy retrieval framework based on approximating reasoning for document retrieval. We also discuss using approximating reasoning to discover additional relations in the database to support more advanced search functions in an intelligent information retrieval system. This paper also introduces an experimental system implementing our proposed technique. The performance of the system is then evaluated and analyzed.

1 Introduction

To help researchers retrieve documents effectively, there are many intelligent information systems or digital libraries [1] that have been developed for document retrieval purpose. In typical document retrieval systems, some Boolean-based query-to-document matching functions [2] are adopted to calculate the keywords relevances. Since there is generally always some uncertainty information concerning keywords or documents, fuzzy logic [3] is proposed to be incorporated in document retrieval systems to better reflect queries and knowledge captured in documents. There are various introduced approaches [4, 5] which have made use of fuzzy query to enhance and improve document retrieval. Basically, a fuzzy query is a query in which the keywords are associated with some membership values [6]. Then, the membership values will be taken into account when the systems calculate the relevances between documents and query keywords. As such, the documents related to a query keyword with greater membership value will be reasonably ranked higher when retrieved.

However, most of the existing techniques on fuzzy query only focus on fuzzifying queries with membership values. That means, uncertainty information of the stored documents has not been investigated and analyzed deeply to support

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 103–114, 2008.
springerlink.com

enhancing document retrieval. Thus, the accuracy and efficiency of these techniques are still limited. In [7], there is an attempt to mine uncertainty information from the Web documents to form a so-called fuzzy concept matrix, which makes use of fuzzy logic to imply the similarities of concepts. However, such fuzzy concept matrix is constructed based heavily on the rates from users. It makes the fuzzy concept matrix construction time-consuming. Also, since most users are not experts in the domains of the rated documents, the concept similarities suggested by them may not be highly precise.

One of practical domains that highly involves uncertainty information is that of scholarly publications. Currently, research and prototyping systems such as the Institute for Scientific Information [8], Open Citation (OpCit) project [9], and CiteSeer [10] have been developed to explore and retrieve scientific publications. However, these existing systems can only support some basic search functions like search for documents or authors based on inputted keywords. Meanwhile, uncertainty information stored in scientific documents should be useful for user to further explore and retrieve scholarly knowledge to fulfill their desired search requirement. For example, information on research areas of documents or researchers can help scholars effectively to find documents related to their research or identify experts in certain domains. However, such uncertainty information is not always explicitly available in a pre-defined database. It is also quite difficult for the database constructor to infer such information manually from raw data.

In this paper, we propose a fuzzy retrieval framework that applies approximating reasoning to overcome the discussed problems. The rest of paper is organized as follows. Section 2 introduces our proposed fuzzy retrieval framework. Section 3 gives formal definitions of fuzzy representations of uncertainty objects and queries. Section 4 presents fuzzy-based retrieval process, which consists of two steps: proposition extraction and approximating reasoning. In Section 5, we discuss using approximating reasoning for fuzzy-based additional relation discovery in database. Section 6 presents an experimental system for fuzzy retrieval of scientific documents. Finally, Section 7 concludes the paper.

2 Fuzzy Retrieval Framework

Figure 1 presents our proposed fuzzy retrieval framework. In the framework, the knowledge to be inferred and retrieved (e.g. documents) is firstly stored in a database called *Fundamental Knowledge* database. Like other typical information retrieval systems, in order to support retrieving data effectively, some Knowledge Discovery in Databases (KDD) techniques, or *data mining* techniques, may be applied to mine additional hidden knowledge from the Fundamental Knowledge database. The mined knowledge is then stored in *Mined Knowledge* database. Traditionally, the KDD technique often used in document retrieval systems is clustering; and the mined knowledge is a set of document clusters. However, in the scope of this paper, we do not focus on this KDD aspect of the framework.

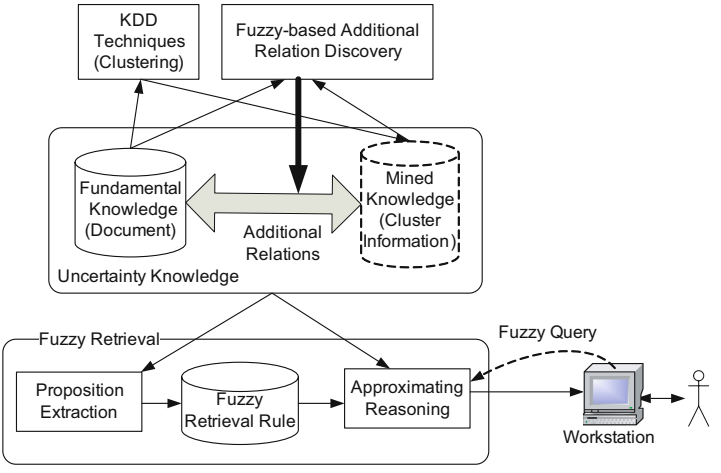


Fig. 1. Fuzzy retrieval framework

In our context, we assume that uncertainty information occurs in both Fundamental Knowledge and Mined Knowledge databases. Thus, the *Fuzzy Retrieval* process will be able to retrieve knowledge from fuzzy queries submitted by users. Fuzzy Retrieval consists of two sub-processes: *Proposition Extraction* and *Approximating Reasoning*. Firstly, Proposition Extraction represents knowledge in the databases as *fuzzy retrieval rules*. Subsequently, when a user submits a fuzzy query, *Approximating Reasoning* is performed to identify the fuzzy retrieval rule that best matches with the submitted query. Then the knowledge corresponding to the identified rule will then be retrieved and returned to the user accordingly.

In addition, the framework also uses approximating reasoning in the *Fuzzy-based Additional Relation Discovery* process to discover additional relations in the databases. The discovered relations can be used to support more advanced search functions using the same fuzzy retrieval mechanism discussed above.

3 Fuzzy Representations of Uncertainty Objects and Queries

We recall the definition of a fuzzy set as follows.

Definition 1 ***(Fuzzy Set).*** *A* fuzzy set *A on a domain U, is defined by a membership function from U to* $[0.1]$*, i.e. each item in A has a membership value given by μ. We denote $\Phi(S)$ as a fuzzy set generated from a traditional set of items S.*

Based on the above basic definition of fuzzy set, we then construct definition on fuzzy representation of an uncertainty object as follows.

Definition 2 ***(Fuzzy Representation of Object).*** *Each object O can be represented by a fuzzy set $\Phi(O)$ on a set of attributes $A = \{A_1, A_2, \ldots, A_m\}$ as $\Phi(O) = \{A_1(\mu_1), A_2(\mu_2), \ldots, A_m(\mu_n)\}$ where μ_i is the membership of O with attribute A_i. $\Phi(O)$ is called the* fuzzy representation *of O on A.*

Example 1. Given a document D and a set of keywords $K = \{$*Data Mining*, *Clustering*, *Fuzzy Logic*$\}$. The relevance degrees between D and each keyword in K are respectively defined as fuzzy membership value of 0.8, 0.12 and 0.61. Thus, the fuzzy representation of D on K is $\Phi(O)=\{$*Data Mining*(0.8), *Clustering*(0.12), *Fuzzy Logic*$(0.61)\}$.

Similarly, we define a fuzzy representation of a set of objects as follows.

Definition 3 ***(Fuzzy Representation of a Set).*** *Given a set of objects $S = \{O_1, O_2, \ldots, O_n\}$, the fuzzy representation of S on a set of attributes A is a fuzzy set $\Phi(C)$ as $\Phi(C) = \Phi(O_1) \cap \Phi(O_2) \cap \ldots \Phi(O_n)$ where $\Phi(O_n)$ is a fuzzy representation of O_i on A.*

In our research, when performing clustering on the database, we have also stored uncertainty information in the generated clusters. Each cluster can then be represented as a fuzzy representation on a set of keywords as stated in Definition 3. Thus, such stored information could be used effectively for fuzzy retrieval in which the submitted queries also contain uncertain requirements. This leads us to the following definition of fuzzy query as follows.

Definition 4 ***(Fuzzy Query).*** *A* fuzzy query *on a set of keywords K is the fuzzy set $Q_f = \Phi(K)$.*

4 Fuzzy Retrieval

Typically, in order to retrieve appropriate documents from inputted queries, document retrieval systems will first model the documents and queries as arithmetic vectors, mostly $tf \cdot idf$ ones. Then, the retrieval process is performed based on vector similarity calculation. However since this method does not employ fundamental principle of fuzzy logic, it has hardly fulfilled effectively jobs involving uncertainty such as dealing with fuzzy query.

In this paper, we propose a new approach of using approximating reasoning to perform retrieval from fuzzy queries. Generally, the retrieval process featured in our approach consists of 2 processes: *Proposition Extraction* and *Approximating Reasoning*.

4.1 Proposition Extraction

The proposition extraction step will represent each cluster of data as a fuzzy proposition. The generated propositions will then be used for approximating reasoning in order to retrieve information in the next step. Thus, each extracted proposition is referred to as *fuzzy retrieval rule*.

Theoretically, a proposition can be represented as a statement "*x is A*" where x is a variable and A is a value. In fuzzy logic, a proposition can be represented as a fuzzy set U, which implies "*x is U*". For reasoning, the proposition that has widely been used in fuzzy logic is the "IF-THEN" proposition, which can be represented as follows: IF <proposition> THEN <proposition>. Assume that we have a cluster C_i that contains a set of objects $\{O_1, O_2, \ldots, O_n\}$. Logically, we can construct an "IF-THEN" proposition as follows: "IF x is O_1 or x is O_2 or ... or x is O_n THEN x belongs to C_i".

However, for retrieval purpose, we do need to represent proposition in a higher uncertainty level. Since a cluster is a group of similar items, we should retrieve a certain cluster as a result returned to a query, if the query relates to at least an item in the cluster. Besides, there may be more than one clusters relating to a certain query in various degrees of relevance, thus the fact "a query relates to a cluster" is an uncertain statement. In that case, we need to represent the following proposition: "IF query q relates to O_1 or q relates to O_2 or ...or q relates to O_n THEN C may relate to q". Obviously, classical logic cannot represent the proposition "C may relate to q".

Theoretically, a proposition "q relates O_i", where O_i is an object, can be represented as $P(O_i) = \{A_i(\mu_{i1}), A_i(\mu_{i2}), \ldots, A_i(\mu_{in})\}$, where $\{A_1, A_2, , A_m\}$ is the set of attributes and μ_{ij} is the membership of O_i with attribute A_j. Thus, the IF proposition "IF query q relates to O_1 or q relates to O_2 or ... or q relates to O_n" can be represented as a fuzzy set $\Phi(O_1) \cup \Phi(O_2) \cup \ldots \Phi(O_n)$.

Since the aim of the IF-THEN proposition is to conclude if a query relates to clusters of objects, the THEN proposition should be a fuzzy set F_S on the domain D_C, where D_C is the set of clusters. The problem is how to calculate the membership for each cluster C_j in D_C when we construct the IF-THEN proposition from a certain cluster C_i.

Since the membership value implies that "If a query q relates to C_i, then how much does q relate to C_j", the membership value of C_j in the THEN part of the proposition should be the subsethood of $\Phi(C_i)$ of $\Phi(C_j)$.

Example 2. Given a dataset of three documents $D_1 = \{$*Data Mining* (0.8), *Clustering* (0.12), *Fuzzy Logic* (0.61)$\}$, $D_2 = \{$*Data Mining* (0.9), *Clustering*(0.85), *Fuzzy Logic*(0.13)$\}$ and $D_3 = \{$*Data Mining* (0.1), *Clustering* (0.14), *Fuzzy Logic* (0.87)$\}$, assume that the dataset can be clustered into two clusters $CK_1 = \{D_1, D_2\}$ and $CK_2 = \{D_1, D_3\}$. (It is noted that some soft-clustering techniques allow an object belonging to more than one clusters). Thus, two IF-THEN propositions for P_1 and P_2, are obtained respectively. Since CK_1 has two objects D_1 and D_2, the IF part of P_1 can be represented as $P(D_1) \cup P(D_2) =$ $\{$*Data Mining* (0.8), *Fuzzy Logic* (0.61)$\}\cup\{$*Data Mining* (0.9), *Clustering* (0.85)$\} = \{$*Data Mining* (0.9), *Clustering* (0.85), *Fuzzy Logic* (0.61)$\}$. Similarly, P_2 has the IF part as $\{$*Data Mining* (0.8), *Fuzzy Logic* (0.87)$\}$. We have $subsethood(\Phi(CK_1), \Phi(CK_1)) = 1$, $subsethood(\Phi(CK_1), \Phi(CK_2)) = 0.41$, $subsethood(\Phi(CK_2), (CK_1)) = 0.42$ and $subsethood(\Phi(CK2), \Phi(CK2)) = 1$. Therefore, P_1 is represented as a fuzzy rule: IF $\{$*Data Mining* (0.9), *Clustering*

(0.85), *Fuzzy Logic* (0.61)} THEN $\{CK_1(1), CK_2(0.41)\}$. Similarly, P_2 is represented as IF{*Data Mining* (0.8), *Fuzzy Logic* (0.87)} THEN$\{CK_1(0.42), CK_2(1)\}$.

4.2 Approximating Reasoning

After the proposition extraction step, we have obtained a set of propositions as fuzzy rules. The next step is to use the generated rules for retrieval using approximating reasoning. For example, assume that we have a fuzzy rule "IF x is A THEN y is B" where A and B are fuzzy sets. Then, if we have a new proposition "x is A", we need to find what conclusion we can get about y.

Theoretically, a proposition IF $< FP1 >$ THEN $< FP2 >$ where $FP1$ and $FP2$ are two fuzzy propositions that can be interpreted as a relation connecting $FP1$ and $FP2$. In classical propositional logic, the rule "IF x THEN y" means "x implies y" and is written as $x \rightarrow y$. When x and y are fuzzy propositions then $x \rightarrow y$ is a fuzzy relation. There are several definitions for the fuzzy implication relation ($\rightarrow$) listed as follows:

- Dienes-Rescher Implication: $\mu_{x \rightarrow y} = \max[1 - \mu_x(u), \mu_y(v)]$
- Zadeh Implication: $\mu_{x \rightarrow y} = \max[min[\mu_x(u), \mu_y(v)], 1 - \mu_x(u)]$
- Godel implication: $\mu_{x \rightarrow y} = \mu_x(u) \leq \mu_y(v) ? 1 : \mu_y(v)$
- Mamdami Implication: $\mu_{x \rightarrow y} = \min[\mu_x(u), \mu_y(v)]$
- Larsen Implication: $\mu_{x \rightarrow y} = \mu_x(u) \cdot \mu_y(v)$

where x and y are two propositions represented as two fuzzy sets on two domains U and V; u and v are two objects in U and V; $\mu_x(u)$ and $\mu_y(v)$ are membership values of u and v in x and y respectively.

Using such implications, we can construct the fuzzy relations R between propositions in the IF part and the THEN part in an "IF-THEN" proposition. Then, we can perform reasoning for new data object O. The conclusion is C such that $C = O \cdot R$. For example, using the Mamdami implication, we can construct fuzzy relations for topic keywords and clusters from proposition $P1$ from 1 as

$$R_1 = \begin{matrix} \\ \textit{Data Mining} \\ \textit{Clustering} \\ \textit{Fuzzy Logic} \end{matrix} \begin{bmatrix} 0.9 & 0.41 \\ 0.85 & 0.41 \\ 0.61 & 0.41 \end{bmatrix}$$

Assume that we have a query Q which can be represented as a fuzzy set of keywords as $Q = \{$*Data Mining* (0.45), *Clustering* (0.86), *Fuzzy Logic* (0.14)$\}$. The conclusion C_{Q1} of Q using proposition P_1 with the Mamdami implication is

$$C_{Q1} = QR_1 = [0.45\ 0.86\ 0.14] \bullet \begin{bmatrix} 0.9 & 0.41 \\ 0.85 & 0.41 \\ 0.61 & 0.41 \end{bmatrix}$$

$$= [0.85\ 0.41] = \{CK_1(0.85), CK_2(0.41)\}$$

Hence, the conclusion obtained is that the membership grade of Q on CK_1 is 0.85 and on CK_2 is 0.41. Since we can obtain a distinct proposition from each

cluster, the final conclusion should be the union of conclusions obtained from the propositions. From the above example, the fuzzy relation R_2 can be constructed from proposition P_2 as

$$R_2 = \begin{matrix} \\ Data\ Mining \\ Clustering \\ Fuzzy\ Logic \end{matrix} \begin{bmatrix} 0.42 & 0.8 \\ 0.00 & 0.00 \\ 0.42 & 0.97 \end{bmatrix}$$

Therefore, the conclusion C_{Q2} given by P_2 is

$$C_{Q2} = QR_2 = [0.45\ 0.86\ 0.14] \bullet \begin{bmatrix} 0.42 & 0.8 \\ 0.00 & 0.00 \\ 0.42 & 0.87 \end{bmatrix}$$

$$= [0.42\ 0.45] = \{CK_1(0.42), CK_2(0.45)\}$$

The final conclusion C_Q is the union of CQ_1 and CQ_2 such that $C_Q = \{CK_1(0.85), CK_2(0.41)\} \cup \{CK_1(0.42), CK_2(0.45)\} = \{CK_1(0.85), CK_2(0.45)\}$. Hence, the final decision is that Q should relate to CK_1, since CK_1 is the semantically closest cluster of Q. Thus, document in CK_1 should be retrieved as the answer set for query Q.

5 Fuzzy-Based Additional Relation Discovery

In the previous section, we have presented a fuzzy-based retrieval technique using approximating reasoning. The technique relies on the explicit information stored in the database, for example the relations between documents and their corresponding clusters. In this section, we discuss another application of approximating reasoning for discovering additional useful relations in database. For example, the keywords extracted from document clusters can provide hints to determine the areas of the documents. Thus, if we can discover the relations between the document authors with the document clusters, we can identify research areas of the authors. This kind of discovered relation can provide significant hints to recognize experts in research areas. To achieve this goal, we first adopt the concept of fuzzy context [11] to represent data relations. First, we define the cross relation between fuzzy contexts as follows.

Definition 5 ***(Fuzzy Context).*** *A* fuzzy context *is a triple* $K = (G, M, I = \varphi(G \times M))$ *where* G *is a set of objects,* M *is a set of attributes, and* I *is a fuzzy set on domain* $G \times M$*. Each relation* $(g, m) \in I$ *has a membership value* $\mu(g, m)$ *in [0,1].*

Example 3. Table 1 presents a fuzzy context representing relations between documents and clusters. For instance, document D_1 belongs to cluster CK_1 with them membership value of 0.9.

Definition 6 ***(Cross Relation).*** *The fuzzy formal context* $K_s = (G_s, M_s, I_s)$ *is said to have a* cross relation *of to a fuzzy formal context* $Kt = (G_t, M_t, I_t)$ *if* $M_s = G_t$*.*

Table 1. A fuzzy context

	CK_1	CK_2	CK_3
D_1	0.9	0.61	0.61
D_2	0.85	-	-
D_3	-	0.87	0.87

Table 2. A fuzzy context having cross relation with the one in Table 1

	D_1	D_2	D_3
$Author_1$	1.0	-	1.0
$Author_2$	0.5	1.0	-

The cross relation represents an inter-context relation that probably occurs between the fuzzy contexts when the set of objects of a context is regarded as the set of attributes of an other contexts. For example, the context represented by the cross table shown in Table 2 has cross relation with the context in Table 1, while the documents are used as attributes of the authors. The membership value of 1.0 implies that the author is the first author of the document, while 0.5 implies that the author is the second author.

Definition 7 ***(Composed Context).*** *Given a fuzzy formal context* $K_s = (G_s, M_s, I_s)$ *that has cross relation with a fuzzy formal context* $K_t = (M_s, M_t, I_t)$*, the* composed context *of* K_s *and* K_t *is the fuzzy formal context* $K_{st} = (G_s, M_t, I_{st} = I_s \cdot I_t)$.

For example, since the fuzzy formal context $K_C = (G_C, M_C, I_C)$ given in Table 4 has a cross relation with the fuzzy formal context $K_A = (G_A, M_A, I_A)$ given in Table 2, we can generate the composed context $K_{CA} = (G_{CA}, M_{CA}, I_{CA})$ where

$$I_{CA} = I_C \bullet I_A = \begin{bmatrix} 1.0 & 0 & 1.0 \\ 0.5 & 1.0 & 0.87 \end{bmatrix} \bullet \begin{bmatrix} 0.9 & 0.61 & 0.61 \\ 0.85 & 0.00 & 0.00 \\ 0.00 & 0.87 & 0.87 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.87 & 0.87 \\ 0.85 & 0.5 & 0.5 \end{bmatrix}$$

Table 3 represents the composed context between the contexts represented in Table 2 and Table 1, respectively. In the composed context, relations between authors and research areas have the appropriate membership values that are generated using fuzzy composition as discussed above. Based on the composed context constructed, one can further perform fuzzy retrieval to support some advanced search function like expert search, similar to what has been done in Section 3 for document search. That is, users still input fuzzy queries indicating the areas they want to find information. However, instead of returning information on relevant documents, the system will answer the authors that have much publications in the areas. These authors should be regarded as experts in searched areas. For example, one can infer that $Author_1$ should be an expert

Table 3. The composed context of the fuzzy formal contexts given by Table 1 and Table 2

	CK_1	CK_2	CK_3
$Author_1$	0.9	0.87	0.87
$Author_2$	0.85	0.5	0.5

in the research area represented by cluster CK_2 due to its corresponding high membership value in the composed context.

6 Performance Evaluation

We have developed an experimental system verify our proposed technique. We have also evaluated the performance of our a system on a citation database. Our system is enhanced from the PubSearch document retrieval system [12], which has been developed previously at Nanyang Technological University, Singapore.

To construct the citation database, we have collected a set of 1400 scientific documents on the research area "Information Retrieval" published in 1987-1997 from the Institute for Scientific Information's (ISI) website . The downloaded documents are preprocessed to extract related information such as the title, authors, citation keywords, and other citation information. The extracted information is then stored as a citation database. The membership value of a document D on a citation keyword C_K in K_f is computed as $\mu(D, C_K) = \frac{n_1}{n_2}$; where n_1 is the number of documents that cite D and contain C_K, and n_2 is the number of documents that cite D. This formula is based on the premise that the more frequent a keyword occurs in the citing paper, the more important the keyword is in the cited paper.

From the constructed database enriched with fuzzy membership values, we have implemented the features of fuzzy document retrieval and expert finding as discussed in Section 3 and Section 4. Fuzzy queries are form based on an interface developed to allow user to specify the fuzzy membership of each query term. To evaluate the retrieval accuracy, we have manually classified documents into groups based on their topics. Then, the significant keywords extracted from each topic have been used as fuzzy queries. The membership values of each keywords have been inferred based on their occurrences. We have used a confidence threshold T_C to filter the most significant keywords.

Figures 2(a–c) give the performance results based on precision, recall and F-measure on both the crisp queries and fuzzy queries according to different values of confidence thresholds T_C varying from 0.0 (i.e. no confidence threshold used) to 0.9. When T_C is higher, the more keywords are filtered. From the figures, the measured recall, precision and F-measure are significantly decreased when T_C is greater than 0.2, as many important keywords are filtered out after this threshold. Figure 2(a) has shown that fuzzy queries have achieved slightly better recall performance than the crisp queries. However, in Figure 2(b), fuzzy queries have achieved significant improvement on precision performance and therefore

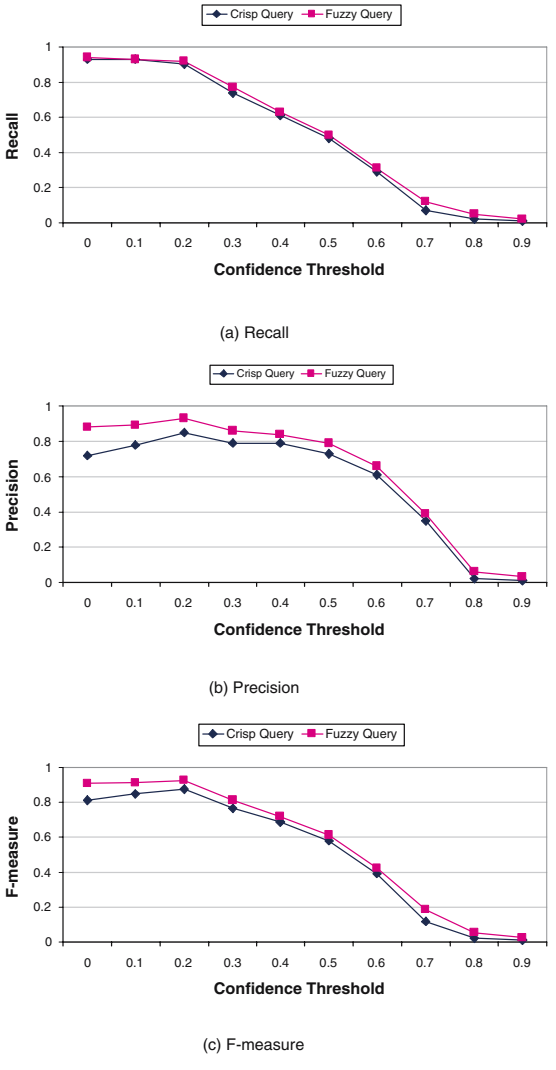


Fig. 2. Performance results based on IR measures

the F-measure performance has significantly been improved as shown in Figure 2(c). It implies that the membership values used in fuzzy queries are useful for determining the relevance of the retrieved documents.

Since precision and recall measure the *accuracy* and *efficiency*, it is also necessary to compare the retrieval accuracy of our proposed approach with other techniques that are commonly used in retrieval systems [13, 14, 15]. Four retrieval techniques are used for comparison. We have used two variations of the k-nearest neighbor (kNN) technique. The first variation, denoted as kNN1, is based on the normalized Euclidean distance for retrieval. The second, denoted as kNN2, uses the fuzzy-trigram technique. In addition, we have also used two

Table 4. Retrieval accuracy

Retrieval Technique	Retrieval Accuracy
kNN1	81.4%
kNN2	77.6%
LVQ3	93.2%
KSOM	90.3%
Reasoning with Crisp Query	84.6%
Reasoning with Fuzzy Query	92.4%

artificial neural networks (ANNs). They are the supervised LVQ3 neural network and the unsupervised KSOM.

Table 4 shows the retrieval accuracy performance. It has shown that when fuzzy queries are used, the accuracy of the proposed approach has outperformed the typical kNN technique and matched with the neural network techniques. Thus, the proposed approach is not only able to preserve the accuracy but also to improve the efficiency of the retrieval. Among the four techniques, LVQ3 has achieved the best performance in terms of accuracy. However, as LVQ3 is a supervised technique, it requires prior expert knowledge for training the network. As such, it will pose a problem when dealing with large database containing uncertain information. Although the proposed approach has achieved slightly less accurate results compared with LVQ3, it can avoid the problem faced by LVQ3.

7 Conclusion

In this paper, we have presented a technique for fuzzy-based information retrieval using approximating reasoning. In practical application of document retrieval, our technique can be either directly applied to retrieve relevant documents for fuzzy queries, or to discover additional relations on the database to support higher level of retrieval, like search for experts in research areas. An experimental system has also been implemented to evaluate the performance of our approach. The initial results are mostly encouraging. We also use the feedback gathered from users to refine our system. Further significant developments/findings will be reported in due course. The work is presented in the context of scientific document retrieval, but it could also be adapted to other types of uncertainty documents, such as medical records, machine-fault records and legal documents. This is because the basic principles are the same.

References

1. Saracevic, T., Dalbello, M.: A survey of digital library education. In: Proceedings of the American Society for Information Science and Technology, vol. 38 (2001)
2. Buell, D.A., Kraft, D.H.: Threshold values and boolean retrieval systems. Information Processing and Management 17(3), 127–136 (1981)

3. Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese 30, 407–428 (1975)
4. Kim, K.-J., Cho, S.-B.: Applied Soft Computing. In: Personalized mining of web documents using link structures and fuzzy concept networks (2006)
5. Choi, D.-Y.: Enhancing the power of web search engines by means of fuzzy query. Decision Support Systems 35(1) (2003)
6. Girill, T.R., Luk, C.H.: Fuzzy Matching as a Retrieval Enabling Technique for Digital Libraries. In: The Digital Revolution: Proceedings of the American Society for Information Science Mid-Year Meeting, pp. 139–145. Information Today, Inc. (1996)
7. Chen, S.-M., Hsiao, W.-H., Horng, Y.-J.: A knowledge-based method for fuzzy query processing for document retrieval. Cybernetics and Systems 28(1), 99–119 (1997)
8. Institute for Scientific Information, http://www.isinet.com
9. Harnad, S., Carr, L.: Integrating, navigating and analyzing eprint archives through open citation linking (the opcit project). Current Science 79, 629–638 (2000)
10. Bollacker, K., Lawrence, S., Giles, C.: Citeseer: An autonomous web agent for automatic retrieval and identification of interesting publications. In: The Third ACM Conference on Digital Libraries, pp. 116–123 (1998)
11. Quan, T.T., et al.: Automatic generation of ontology for scholarly semantic web. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 726–740. Springer, Heidelberg (2004)
12. He, Y., Hui, S.C.: Mining a web citation database for author co-citation analysis. Information Processing and Management 38(4), 491–508 (2002)
13. Watson, I.D.: Applying Case-based Reasoning: Techniques for Enterprise Systems. Morgan Kaufman Publishers, San Francisco (1997)
14. Papagni, M., Cirillo, V., Micarelli, A.: A hybrid architecture for a user-adapted training system. In: Proceedings of the 5th German Workshop on Case-based Reasoning, pp. 181–188 (1997)
15. Richter, M., Wess, S.: Similarity, uncertainty and case-based reasoning in patdex, http://citeseer.csail.mit.edu/49639.html

Probabilistic Constraints for Inverse Problems

Elsa Carvalho, Jorge Cruz, and Pedro Barahona

Centro de Inteligência Artificial, Universidade Nova de Lisboa, Portugal
elsac@uma.pt, {jc,pb}@di.fct.unl.pt

Summary. The authors previous work on probabilistic constraint reasoning assumes the uncertainty of numerical variables within given bounds, characterized by *a priori* probability distributions. It propagates such knowledge through a network of constraints, reducing the uncertainty and providing *a posteriori* probability distributions. An inverse problem aims at estimating parameters from observed data, based on some underlying theory about a system behavior. This paper describes how nonlinear inverse problems can be cast into the probabilistic constraint framework, highlighting its ability to deal with all the uncertainty aspects of such problems.

1 Introduction

Many problems of practical interest can be formulated as nonlinear inverse problems. Such problems aim at finding the parameters of a model, given by systems of equations, from noisy data. Classical approaches for these problems are based on nonlinear regression methods, which search for the model parameter values that best-fit a given criterion [9]. Best-fit approaches, often based on local search methods, provide a non reliable single solution which may not be enough to the adequate characterization of the parameters.

Other stochastic approaches [16] associate a probabilistic model to the problem, from which is possible to obtain any sort of statistical information on the model parameters. These approaches typically rely on extensive random sampling to characterize the parameter space. However, even after intensive computations, no definitive conclusion can be drawn with these approaches, because a significant subset of the parameter space may have been missed.

In contrast, bounded-error approaches [5, 4] aim at characterizing the set of all solutions consistent with the uncertainty on the parameters, the model and the data. This is achieved through constraint reasoning, where initial intervals, representing the uncertainty on parameter values, are safely narrowed by reliable interval methods. Nevertheless, this approach has a major pitfall as it considers the same likelihood for all values in the intervals. In the interval computations context a combination of intervals with probabilistic uncertainty was proposed [6]. However, its dependence on a forward evaluation algorithm makes it unsuitable for inverse problems.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 115–128, 2008.
springerlink.com

The authors previous proposal of a probabilistic constraint framework [2] aims at computing an *a posteriori* distribution from an *a priori* distribution accordingly to safe reasoning on a continuous constraint model. In this paper we argue that such framework may constitute an attractive alternative approach to inverse problems, bridging the gap between pure probabilistic reasoning and pure safe reasoning.

The paper is organized as follows. In section 2 inverse problems are introduced. In sections 3 and 4 the constraint programming framework and the basic ideas of probabilistic reasoning are presented, together with their existing approaches to inverse problems. Next probabilistic interval computations are briefly presented. Section 6 describes the authors probabilistic constraint framework and highlights its ability to deal with all the uncertainty aspects of inverse problems. Finally, conclusions and future work are discussed.

2 Inverse Problems

A mathematical model describes a system by a set of variables and equations that establish relationships between them. In the context of inverse problems, the variables are divided into model parameters, whose values completely characterize the system, and observable parameters, which can be measured. The model is typically a forward model, defining a mapping from the model parameters to the observable parameters. It allows predicting the results of measurements based on the model parameters. An inverse problem is the task of obtaining values for the model parameters from the observed data.

The forward mapping, resulting from some theory about the system behavior, is commonly represented as a vector function $\boldsymbol{f}$ from the parameter space $\mathbf{m}$ (model parameters) to the data space $\mathbf{d}$ (observable parameters):

$$\mathbf{d} = \boldsymbol{f}(\mathbf{m}) \tag{1}$$

Such relation may be represented explicitly by an analytical formula or implicitly by a complex system of equations or some special purpose algorithm.

Nonlinearity and uncertainty play a major role in modeling the behavior of most real systems. In inverse problems the main sources of uncertainty are model approximations and measurement errors. Given uncertainty, an inverse problem may have no exact solutions, since usually there are no model parameter values capable of predicting exactly all the observed data. However, since the model equations are often highly nonlinear, uncertainty may be dramatically magnified, and an arbitrarily small change in the data may induce an arbitrarily large change in the values of the model parameters.

For example, consider the data summarized in Table 1 based on the USA census over the years 1790 (normalized to 0) to 1910 with a 10 year period.

Assuming that an exponential growth is an acceptable model for the population growth, the forward mapping would be defined by the following set of equations (one for each pair $\langle t_i, d_i \rangle$):

Table 1. US Population (in millions) over the years 1790 (0) to 1910 (120)

t_i	0	10	20	30	40	50	60	70	80	90	100	110	120
d_i	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4	39.8	50.2	62.9	76.0	92.0

$$d_i = m_0 e^{m_1 t_i} \tag{2}$$

where m_0 and m_1 are the model parameters whose values must be estimated from the observed data. This is an example of an inverse problem with a simple nonlinear forward model but with no possible value combination for the model parameters satisfying the observed data set.

This problem is classically handled as a curve fitting problem [9]. Such approaches for nonlinear inverse problems are based on nonlinear regression methods which search for the model parameter values that best-fit a given criterion. For instance, the (weighted) least squares criterion minimizes a (weighted) quadratic norm of the difference between the vector of observed data and the vector of model predictions.

In the above example the weighted least squares criterion would be the minimization of the expression:

$$\sum_i \left(\frac{d_i - m_0 e^{m_1 t_i}}{\sigma_i} \right)^2 \tag{3}$$

where σ_i is the weight associated with the error of the i^{th} observation.

The minimization criteria (such as the weighted least squares criterion) are justified by the hypothesis that all problem uncertainties may be modeled using some well behaved distributions (such as Gaussians) eventually with specific parameter values for the different observations (the σ_i values). This is the case for some linear or weakly nonlinear inverse problems, where efficient computational techniques exist to solve them as curve fitting problems [9].

In generic nonlinear inverse problems, where no explicit formula can be provided for obtaining the best-fit values, minimization is often performed through local search algorithms. However, the search method may stop at a local minimum with no guarantees on the complete search space. Moreover, in most problems, a single best-fit solution may not be enough. Since other solutions could also be quite satisfactory with respect to the adopted criterion, the uncertainty around possible solutions should also be characterized. Analytic techniques can only be used for this purpose relying on some special assumptions about the model parameter distributions (for instance, assuming a single maximum). However, if the problem is highly nonlinear such assumptions do not provide realistic approximations for the uncertainty.

3 Continuous Constraint Satisfaction Problems

A Constraint Satisfaction Problem (CSP) [10] is defined by a triple (X;D;C) where X is a set of variables, each with an associated domain of possible

values in D, and C is a set of constraints on subsets of the variables. A constraint specifies which values from the domains of its variables are compatible. A solution to the CSP is an assignment of values to all its variables, which satisfies all the constraints. In continuous CSPs (CCSPs) [8, 1, 15] variable domains are continuous real intervals and constraints are equalities and inequalities. The space of possibilities is represented by boxes, i.e., the Cartesian product of real intervals. The CCSP framework is powerful enough to model a wide range of problems. In particular, engineering systems with components described as sets of continuous valued variables and relations defined by numerical equalities or inequalities, eventually with uncertain parameters. Continuous constraint reasoning eliminates value combinations from the initial search space (the Cartesian product of the initial domains), without loosing solutions. It combines pruning and branching steps until a stopping criterion is satisfied.

The pruning of the variable domains is based on constraint propagation. The main idea is to use the partial information expressed by a constraint to eliminate some incompatible values from the domain of its variables. Once the domain of a variable is reduced, this information is propagated to all constraints with that variable in their scopes. The process terminates when the domains cannot be further reduced by any constraint. Safe narrowing functions (mappings between boxes) are associated with constraints, to eliminate incompatible value combinations. Efficient methods from interval analysis (e.g. the interval Newton [11]) are often used to implement efficient narrowing functions which are correct (do not eliminate solutions) and contracting (the box obtained is smaller or equal than the original one).

Constraint propagation is a local consistency algorithm for pruning the variable domains, which is often insufficient to support safe decisions. To obtain better pruning, it is necessary to split the boxes and reapply constraint propagation to each sub-box. Such branch and prune process enforces a stronger, non local, consistency criterion. Several consistency criteria have been proposed [8, 1, 3], with distinct trade-offs between efficiency and pruning.

3.1 Constraint Approach to Inverse Problems

In the classical CCSP framework, the uncertainty associated with the problem is modeled by using intervals to represent the domains of the variables. Constraint reasoning reduces uncertainty, by reshaping the search space to become a safe approximation of the solution space. Such framework, with its efficient techniques to deal with nonlinear constraints and its safe uncertainty narrowing capabilities, is naturally appealing to handle nonlinear inverse problems beyond the classical best fitting methods.

The application of CCSPs in the context of inverse problems is known as bounded-error estimation or set membership estimation [5, 4]. The idea is to replace the search for a single best-fit solution (a parameter value combination) with the characterization of the set of all solutions consistent with the forward model, the uncertainty on the parameters and on the observations.

In its simplest form, bounded-error estimation assumes initial intervals to each problem variable, either a model or an observable parameter, and solve the CCSP with the set of constraints representing the forward model. Such strategy assumes prior knowledge on the acceptable parameter ranges as well as on the uncertainty of (difference between) predicted and observed data.

From the safe approximation of the solution space, a projection on the set of model parameters (or any subset of it) provides insight on the remaining uncertainty about their possible value combinations. In practice, since the narrowed ranges of the observable parameters are not essential, the variables that represent them may be replaced by their respective initial intervals.

Applying bounded-error estimation to the inverse problem described in section 2, it can be reformulated as a CCSP with the following set of constraints (one for each pair $\langle t_i, d_i \rangle$):

$$[d_i - \delta_i, d_i + \delta_i] \ni m_0 e^{m_1 t_i} \tag{4}$$

where δ_i is an acceptable difference between the i^{th} observation and the respective predicted value, and m_0 and m_1 are the only variables of the constraint model. The initial ranges for these variables should be provided within reasonable bounds (I_0 and I_1) and represent the parameter uncertainty that will be reduced through constraint reasoning. Figure 1 shows the approximation of the solution space that is computed with $I_0 = [0, 100]$, $I_1 = [0.01, 0.1]$ and $\delta_i = 3$ for all observations presented in Table 1. From the figure, it is clear which combinations of the model parameter values are consistent with the initial uncertainty assumptions, the forward model and the observations.

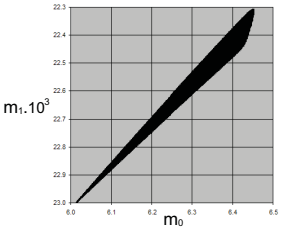


Fig. 1. Approximation of the CCSP solution space

The formulation of an inverse problem as a CCSP may easily accommodate additional requirements, in the form of constraints, which are more difficult to enforce in classical approaches. Moreover, the generality of this approach allows its application to inverse problems whose forward model is not defined by an explicit analytical formula but rather by a complex set of relations.

However, in many cases, safe reasoning is useless, intervals are often very wide, and subsequent constraint propagation is not able to narrow them. In fact, an uncertain value may range over a wide interval but a much narrower interval may include the most likely values. In some problems, the plausibility distribution of values within the bounds of an uncertain parameter is known. For

instance, uncertainty due to measuring errors may be naturally associated with an error distribution. However, the traditional CCSP framework cannot accommodate such information and thus, for each variable, all values in its domain are considered equally plausible.

4 Probabilistic Reasoning

Probability provides a classical model for dealing with uncertainty. The basic element of probability theory is the random variable, which plays a similar role to that of the CSP variables. Each random variable has a domain where it can assume values. In particular, continuous random variables assume real values. A possible world, or atomic event, is an assignment of values to all the variables of the model. An event is a set of possible worlds. The complete set of all possible worlds in the model is the sample space. If all the random variables are continuous, the sample space is the hyperspace obtained by the Cartesian product of the variable domains, and the possible worlds and events are, respectively, points and regions from such hyperspace.

Probability measures may be associated with events. In the continuous case, an assignment of a probability to a point, is representative of the likelihood in its neighborhood. A probabilistic model is an encoding of probabilistic information, allowing to compute the probability of any event, in accordance with the axioms of probability. The usual method for specifying a probabilistic model assumes, either explicitly or implicitly, a full joint probability distribution, which assigns a probability measure to each possible world.

Probabilistic reasoning aims at incorporating new information, known as evidence, by updating an *a priori* probability into an *a posteriori* probability given the evidence. The *a priori* probability is a description of what is known in the absence of the evidence. For incorporating this evidence, conditioning is used. Conditional probability $P(A|B)$ is the probability of some event A, given the occurrence of some other event B. The *a posteriori* probability is the conditional probability when the relevant evidence is taken into account.

Probabilistic graphical models [7] (Markov networks and Bayesian networks [13]) provide a powerful framework for efficient probabilistic reasoning. The idea is to use a probabilistic network that captures the structural properties of the probabilistic model (such as conditional independence) and defines an implicit full joint probability distribution. Given new evidence (information about some nodes), belief propagation [13] is one of the most efficient inference algorithms to compute *a posteriori* probabilities for all the non-evidence nodes in the network. However, such approaches, which require the full specification of a conditional probability at each node of the network, are often inadequate for continuous nonlinear problems.

4.1 Probabilistic Approach to Inverse Problems

Inverse problems are often handled by probabilistic approaches that associate an explicit probabilistic model to the problem [16]. Prior information on the model

parameters is represented by a probability distribution, which is transformed into an *a posteriori* probability distribution, by incorporating a forward theory (relating the model parameters to the observable parameters) and the actual result of the observations (with their uncertainties).

On these approaches all the information related to inverse problems is described using probability densities. The *a posteriori* probability density of the model parameters $\sigma_M(\mathbf{m})$ can be computed from these distributions. For example, assuming linear data and model spaces this probability density is:

$$\sigma_M(\mathbf{m}) = k\rho_M(\mathbf{m}) \int_D \rho_D(\mathbf{d})\Theta(\mathbf{d},\mathbf{m})d\mathbf{d} \tag{5}$$

where k is a normalization constant and $\rho_D(\mathbf{d})$, $\rho_M(\mathbf{m})$ and $\Theta(\mathbf{d},\mathbf{m})$ are the probability densities for the observable parameters, the model parameters and the underlying theory.

$\sigma_M(\mathbf{m})$ provides an explicit definition of a full joint probability distribution on the model parameter space, from which is possible to obtain any sort of statistical information on the model parameters compatible with the *a priori* uncertainty, the theoretical information and the experimental results. In the particular case where $\sigma_M(\mathbf{m})$ is identically null, some incompatible assumptions were surely made, indicating that uncertainty has been underestimated.

Only in very simple cases analytic techniques can be used to characterize $\sigma_M(\mathbf{m})$. In general, it is necessary to perform an extensive exploration of the model space. When it is small, a systematic exploration may be achieved, computing $\sigma_M(\mathbf{m})$ at every point of a grid defined over the complete space. Usually, such exploration cannot be systematic since too many points would have to be evaluated, so it is replaced by a random (Monte Carlo) exploration. For a more detailed description of these methods see [16, 12].

Nevertheless, only a number of discrete points of the continuous model space is analyzed and the results must be extrapolated to characterize the overall uncertainty. Such approach is highly dependent on the exploration length which, to provide better uncertainty characterizations, need to be reinforced in highly nonlinear problems. Moreover, contrary to constraint reasoning approaches, these probabilistic techniques cannot prune the search space based on model information. Consequently the entire space is considered for exploration, independently of its *a posteriori* probability distribution, which can have null values for inconsistent subregions.

5 Probabilistic Interval Computations

A combination of probabilistic and interval representations of uncertainty appears in [6]. This approach uses interval domains to represent the ranges of possible values and allows the incorporation of extra information about their probabilities. Such framework uses interval computations instead of the broader paradigm of continuous constraint reasoning. It is specially suited for data

processing problems, where an estimate for a quantity y must be computed, applying a known deterministic algorithm f to the available estimates for other quantities $x_1, \ldots, x_n$.

$$y = f(x_1, \ldots, x_n) \tag{6}$$

Data processing problems arise naturally when the estimates for the quantities $x_1, \ldots, x_n$ come from direct measurements obtained by physical instruments which provide upper bounds for the measurement errors, together with information about the error probability distributions. Probabilistic interval computations provide an alternative approach when only partial information about error distributions is available and standard statistical approaches [14] cannot be applied. It makes use of such partial information in the estimation of y. The idea is to maintain intervals to represent possible values of variables as well as possible values of parametric descriptors of their distributions (e.g., expected values). During interval computations such intervals are maintained consistent by a step by step evaluation process that extends basic interval arithmetic operations (see [6] for details).

Contrary to constraint approaches, that are based on undirected relations, this approach is highly dependent on the availability of a directed algorithm f to compute the pretended information for y from the $x_1, \ldots, x_n$ estimates. Clearly this is not the case for inverse problems where the model parameters are not outputs but rather inputs of the forward model. So, to apply this approach, it would be necessary to find a suitable analytical solution with respect to each model parameter.

6 Probabilistic Constraint Reasoning

In [2] the authors proposed the Probabilistic Continuous Constraint Satisfaction Problem (PCCSP) as an extension of a CCSP. A PCCSP is defined by (X;D;F;C), where X is a set of continuous random variables, each with an associated interval domain of possible values in D, distributed accordingly to the corresponding probability density function (p.d.f.) in F, and C is a set of constraints on subsets of the variables. Given a point in the domain of a random variable, its p.d.f. is representative of the *a priori* probability in its neighborhood, without considering the relations between the variables. It is assumed that all relevant relations between variables are expressed by the constraints of the model. Thus, when the constraints are not accounted for, the variables are independent.

The initial search space represents a probability space, characterized by a joint p.d.f. which, due to the independence assumption, is implicitly defined by the product of the individual p.d.f.s of the random variables. In the process of reducing uncertainty, there is a combination of continuous constraint reasoning and probabilistic reasoning. While the first reduces uncertainty by *reshaping* the search space, the second *redefines* the search space *a priori* probability distribution by computing an *a posteriori* distribution, based on the constraint reasoning outcome.

The constraints are the new information that is incorporated in the probabilistic model. The solution space is the event containing all possible worlds that satisfy the constraints. Through constraint reasoning an approximation (enclosure) of the solution space is obtained. Therefore the *a posteriori* probability is computed as a conditional probability, given the evidence represented by the approximation of the solution space. This probability is calculated by the conditional probability rule $P(A|B) = P(A \cap B)/P(B)$. The probability of region A given the evidence, is the probability of the subregion of A contained in the approximation of the solution space, divided by a normalizing factor.

The quality of the solution space approximation depends on the consistency and stopping criteria used in the constraint reasoning process. Regions of the search space that were not pruned during constraint reasoning may contain solutions, although there is no guaranty that they do. In fact, there is no knowledge why such regions are maintained. Was it due to lack of further exploration of this regions or did they contain solutions? Normally, the process of constraint reasoning, leads to non uniform sizes of the boxes that represent the solution space approximation. Nevertheless, for reasoning with probabilistic information, some kind of fairness in the exploration of the search space must be guaranteed, so that the obtained *a posteriori* distribution is not biased by heterogeneous search.

In the PCCSP, the stopping criterion is based on the maximum width ε_1 allowed for the intervals that constitute a box. A box is no further explored when all its intervals are smaller or equal to ε_1. When all the boxes meet this criterion the search stops. The stopping criterion assures some uniformity of the solution space approximation. However, due to the consistency enforcement narrowing capability, which differs between distinct regions of the search space, some heterogeneity is still present.

To maintain a generic non parametric representation of the *a posteriori* marginal p.d.f.s, some kind of discretization must be assumed. This is achieved by considering an ε_2-hypergrid, i.e., a grid where the dimension is the number n of variables in the PCCSP, and each grid unit (hypercube) has width ε_2 in all dimensions. The hypergrid allows to transform the non uniform solution space approximation, resulting from constraint reasoning, in a uniform one, providing a fair computation of the marginal p.d.f.s. The transformation is achieved by overlaying the hypergrid upon the solution space approximation, enforcing a *snap to grid* to this region. The new approximation is the set of grid hypercubes that intersect with the original approximation, producing a set of uniform boxes. Figure 2 illustrates the described reasoning process.

Once obtained the solution space approximation as a set SS of ε_2-hypergrid boxes, algorithm 1 calculates and returns the marginal *a posteriori* p.d.f. of a set of m variables (where $m \leq n$), discretized accordingly to the ε_2-hypergrid. For that purpose the algorithm maintains a m-dimensional matrix M, where each dimension corresponds to a variable. In the algorithm, $H_{box}[i]$ and $G_{box}[i]$ are the i^{th} intervals of the boxes (Cartesian product of n intervals). Given two intervals $I_1 = [l_1, r_1]$ and $I_2 = [l_2, r_2]$, the union hull $I_1 \uplus I_2$ is the interval

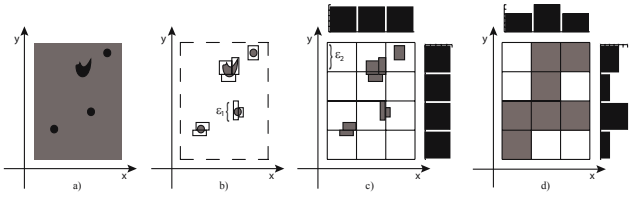


Fig. 2. Process of probabilistic constraint reasoning. (a) Initial search space and solution space; (b) Solution space approximation; (c) Hypergrid and *a priori* marginal p.d.f.s; (d) Snap to grid and *a posteriori* marginal p.d.f.s.

$[min(l_1, l_2), max(r_1, r_2)]$. H_{box} is a box where each interval is the union hull of the respective intervals of all the boxes in SS, i.e. is the smallest box enclosing all the boxes in SS (line 2). The length of each dimension is the number of ε_2 segments in which the corresponding variable domain can be divided (line 3). Each matrix cell thus obtained is initialized to zero (line 4) and its probability value is computed by summing up the contribution of all hypercubes (boxes) that are aligned with that cell (line 5-9), normalized by the sum of all hypercube contributions (lines 1, 10, 12). Due to the independence assumption, the contribution of an hypercube is the product of each variable contribution, i.e., the integral of its *a priori* p.d.f. (f_i), within the respective box interval (line 8).

Algorithm 1. Calculates marginal *a posteriori* p.d.f.

function $marginalAPosterioriPDF(SS, \varepsilon, m)$

1: $accum \leftarrow 0$
2: $H_{box} \leftarrow \uplus SS$
3: $\forall_{1 \leq i \leq m} \quad l_i \leftarrow H_{box}[i].width/\varepsilon$
4: $\forall_{1 \leq i_1 < l_1} \ldots \forall_{1 \leq i_m < l_m} \quad M[i_1] \ldots [i_m] \leftarrow 0$
5: **while** $SS \neq \oslash$ **do**
6: $\quad G_{box} \leftarrow remove(SS)$
7: $\quad \forall_{1 \leq i \leq m} \quad j_i \leftarrow (G_{box}[i].left - H_{box}[i].left)/\varepsilon$
8: $\quad p_{box} = \prod_{i=1}^{n} \int_{G_{box}[i].left}^{G_{box}[i].right} f_i(x_i)dx_i$
9: $\quad M[j_1] \ldots [j_m] \leftarrow M[j_1] \ldots [j_m] + p_{box}$
10: $\quad accum \leftarrow accum + p_{box}$
11: **end while**
12: $\forall_{1 \leq i_1 < n_1} \ldots \forall_{1 \leq i_m < n_m} \quad M[i_1] \ldots [i_m] \leftarrow M[i_1] \ldots [i_m]/accum$
13: **return** M

6.1 Probabilistic Constraint Approach to Inverse Problems

The application of PCCSPs in the context of inverse problems, based on bounded-error estimation, assumes both prior knowledge on the acceptable parameter ranges and on the uncertainty about the difference between predicted

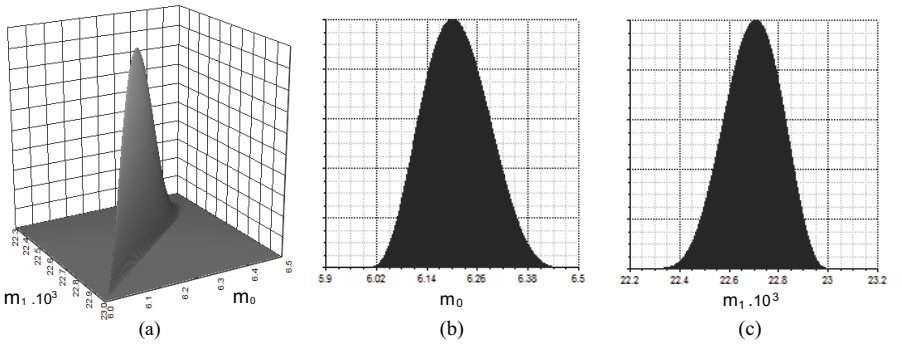


Fig. 3. Exponential model. (a) Joint p.d.f.; (b)(c) marginal p.d.f.s.

and observed data. This knowledge is expressed, respectively, by intervals and explicit *a priori* probability distributions within such intervals. If prior information is unavailable uniform distributions are considered.

When solving the PCCSP, where the set of constraints represents the forward model, a safe approximation of the solution space is obtained, and a projection on the set of model parameters (or any subset of it) provides insight on the *a posteriori* distribution of the resulting narrowed ranges.

In this approach, data parameters cannot be replaced by their respective initial intervals because it is necessary to keep track of their p.d.f.s for computing the *a posteriori* distributions. However, as long as the model parameters are the only shared variables between constraints, the contribution of each constraint, on the model parameters *a posteriori* distribution, may be independently computed and incrementally combined.

Consider again the inverse problem presented in section 2. Suppose that besides accepting a difference δ_i between the i^{th} observation and the respective predicted value, a p.d.f. $\rho_i(d_i)$ is associated to the acceptable interval representing the prior information on such difference. The initial ranges and respective p.d.f.s (possibly uniform distributions) must also be provided for the model parameters m_0 and m_1, characterizing the full *a priori* joint p.d.f.. In this case, the inverse problem may be reformulated as a PCCSP with the following set of constraints (one for each pair $\langle t_i, d_i \rangle$):

$$d_i = m_0 e^{m_1 t_i} \tag{7}$$

where m_0, m_1 and d_0, ... , d_{12} are the variables of the constraint model. The *a posteriori* distribution of the model parameters is computed by solving the PCCSP (projecting the results with respect to m_0 and m_1).

Figure 3 shows the *a posteriori* distribution of the model parameters that is computed with the initial ranges defined in subsection 3.1 and assuming *a priori* uniform distributions for the model parameters and triangular distributions (centered in d_i) for the observable parameters. Besides identifying which value combinations of m_0 and m_1 are consistent, figure 3(a) illustrates its joint

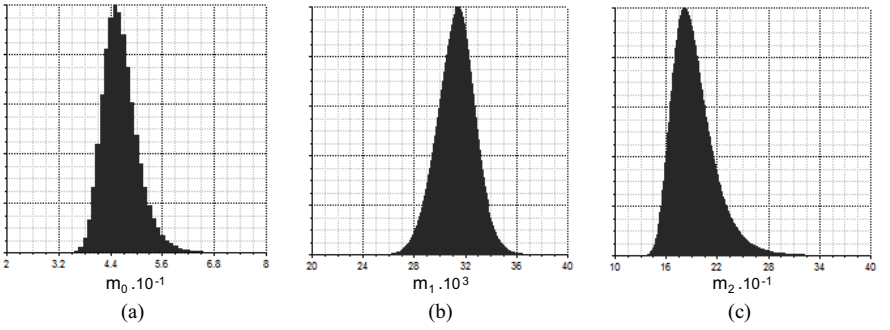


Fig. 4. Logistic model. (a)(b)(c) marginal p.d.f.s.

probability distribution, allowing to identify regions of maximum likelihood. Figures 3(b) and 3(c) are projections on m_0 and m_1 showing the *a posteriori* probability computed for each of the model parameters.

If a logistic model is considered, instead of an exponential model for the population growth, its reformulation as a PCCSP should keep the observable variables but satisfy a new set of constraints (one for each pair $\langle t_i, d_i \rangle$):

$$d_i = \frac{m_2}{1 + m_0 e^{-m_1 t_i}} \tag{8}$$

where m_0, m_1 and m_2 are the variables of the constraint model representing the model parameters. Figure 4 presents the marginal *a posteriori* distributions for each model parameter computed from the joint p.d.f. (with equal assumptions on the observed parameters uncertainty and with initially uniformly distributed $I_0 = [10, 100]$, $I_1 = [0.02, 0.05]$ and $I_2 = [100, 400]$).

The PCCSP associated with a given inverse problem can be easily extended to make predictions on the outcomes of new measurements. For this purpose a new constraint for each new measurement should be included in the model. Such constraints, similar to the other constraints representing the forward model, should include new unknown observable parameters (initially unbounded and uniformly distributed). *A posteriori* distributions for these new variables can be computed by solving the PCCSP and projecting the results with respect to each of them. Figure 5 illustrates the predictions for the population size in 1920 ($t_i = 130$) in the previous problem with both, the exponential model (figure 5(a)) and the logistic model (figure 5(b)). Note that the real observed value for the population size in 1920 was 106.0 (not shown in table 1) which is in accordance with the predictions of the logistic model, but outside the bounds predicted by the exponential model.

An insight about the quality of a particular model for a specific inverse problem may be achieved by analyzing the maximum likelihood regions. The obtained *a posteriori* marginal p.d.f. for the model parameters provides valuable information for inspecting the quality of a particular model. Not only it allows easy identification of maximum likelihood regions as peaks of such p.d.f., but also displays the complete shape of the uncertainty dispersion showing, for instance, if it is unimodal.

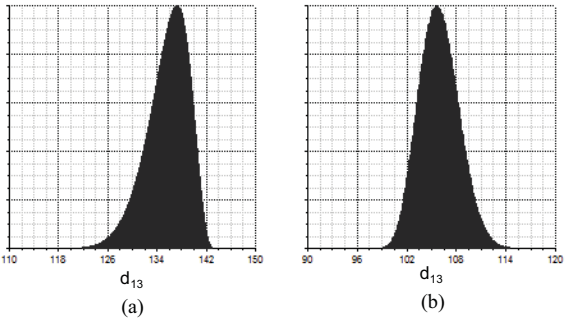


Fig. 5. Expected US population in 1920. (a) Exponential and (b) logistic models.

In the presented example, given the unimodality of the *a posteriori* p.d.f.s for both models, a quantitative measure of their quality may be obtained by evaluating any numerical best-fit criterion (see section 2) at their maximum likelihood points. The boxes that enclose such points for the exponential and the logistic models are, respectively, $\langle[6.159, 6.160], [0.022770, 0.022771]\rangle$ and $\langle[45, 46], [0.0318, 0.0319], [181, 182]\rangle$. The least squares criterion (formula 3 with $\sigma_i = 1$) evaluated at this boxes results, respectively, in $I_1 = [24.6, 18341.6]$ and $I_2 = [0.1, 11.8]$. Since the maximum likelihood points are included in those boxes and any value of I_2 is smaller than any value of I_1, according to the chosen criterion, the logistic model is a better representation for the population growth than the exponential model.

7 Conclusions and Future Work

This paper describes how inverse problems can be cast into the probabilistic continuous constraint framework. The approach introduces new expressive power for modeling the underlying theory about the system behavior and produces appealing graphical results for representing the uncertainty on model parameters and predictions on measurement outcomes. However, it seems to be more adequate to handle inverse problems with a reduced number of parameters. This is particularly true when the model is highly nonlinear, in which case, a smaller granularity is required for pruning the search space. To address the scalability of the approach further experimentation must be done on more realistic inverse problems. Furthermore, we intend to develop an interactive prototype, to improve usability and fully explore the framework capabilities.

References

1. Benhamou, F., McAllester, D., van Hentenryck, P.: Clp(intervals) revisited. In: ISLP, pp. 124–138. MIT Press, Cambridge (1994)
2. Carvalho, E., Cruz, J., Barahona, P.: Probabilistic reasoning with continuous constraints. In: AIP Conference Proceedings, vol. 936(1), pp. 105–108 (2007)

3. Cruz, J., Barahona, P.: Maintaining global hull consistency w/ local search for cc-sps. In: Bliek, C., Jermann, C., Neumaier, A. (eds.) COCOS 2002. LNCS, vol. 2861, pp. 178–193. Springer, Heidelberg (2003)
4. Granvilliers, L., Cruz, J., Barahona, P.: Parameter estimation using interval computations. SIAM J. Scientific Computing 26(2), 591–612 (2004)
5. Jaulin, L., Walter, E.: Set inversion via interval analysis for nonlinear bounded-error estimation. Automatica 29(4), 1053–1064 (1993)
6. Kreinovich, V.: Probabilities, intervals, what next? optimization problems related to extension of interval computations to situations with partial information about probabilities. Journal of Global Optimization 29, 265–280 (2004)
7. Lauritzen, S.L.: Graphical Models. Oxford Statistical Science Series, vol. 17. Clarendon Press, Oxford (1996)
8. Lhomme, O.: Consistency techniques for numeric csps. In: Proc. of the 13th IJCAI, pp. 232–238 (1993)
9. Moler, C.B.: Numerical Computing with Matlab. SIAM (2004)
10. Montanari, U.: Networks of constraints: Fundamental properties and application to picture processing. Information Sciences 7(2), 95–132 (1974)
11. Moore, R.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
12. Mosegaard, K.: Resolution analysis of general inverse problems through inverse monte carlo sampling. Inverse Problems 14(3), 405–426 (1998)
13. Pearl, J.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann, San Francisco (1988)
14. Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice. Springer, Heidelberg (2005)
15. Sam-Haroud, D., Faltings, B.: Consistency techniques for continuous constraints. Constraints 1(1/2), 85–118 (1996)
16. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia, PA, USA (2004)

The Evidential Reasoning Approach for Multi-attribute Decision Analysis under Both Fuzzy and Interval Uncertainty

Min Guo[1], Jian-Bo Yang[1,3], Kwai-Sang Chin[2], and Hongwei Wang[1]

[1] Institute of Systems Engineering
Key Laboratory of Image Processing and Intelligent Control
Huazhong University of Science and Technology, Wuhan, China
guomin_like@sina.com.cn
[2] Department of Manufacturing Engineering and Engineering Management
City University of Hong Kong, Hong Kong, China
mekschin@cityu.edu.hk
[3] Manchester Business School, The University of Manchester, UK
jian-bo.yang@umist.ac.uk

Summary. Many multiple attribute decision analysis (MADA) problems are characterised by both quantitative and qualitative attributes with various types of uncertainties. Incompleteness (or ignorance) and vagueness (or fuzziness) are among the most common uncertainties in decision analysis. The evidential reasoning (ER) and the interval grade evidential reasoning (IER) approaches have been developed in recent years to support the solution of MADA problems with interval uncertainties and local ignorance in decision analysis. In this paper, the ER approach is enhanced to deal with both interval uncertainty and fuzzy beliefs in assessing alternatives on an attribute. In this newly developed FIER approach, local ignorance and grade fuzziness are modelled under the integrated framework of a distributed fuzzy belief structure, leading to a fuzzy belief decision matrix. A numerical example is provided to illustrate the detailed implementation process of the FIER approach and its validity and applicability.

1 Introduction

Many real world multiple attribute decision analysis (MADA) problems are characterised with both quantitative and qualitative attributes. In many circumstances, the attributes, especially qualitative ones, could only be properly assessed using human judgments, which are subjective in nature and are inevitably associated with uncertainties caused due to either or both of the following two phenomenon:

(i) Human being's inability to provide complete judgments, or the lack of information, which is referred to as "ignorance" (incompleteness);
(ii) The vagueness of meanings about attributes and their assessments, which is referred to as "fuzziness" (vagueness).

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 129–140, 2008.
springerlink.com

For decades, many MADA methods have been developed, such as the well-know Analytical Hierarchy Process (AHP) [10] and Multiple Attribute Utility Theory [8, 4]. as well as their extensions [1, 2, 11, 7]. In those methods, MADA problems are modelled using decision matrices, in which an alternative is assessed on each attribute by either a single real number or an interval value. Unfortunately, in many decision situations using a single number or interval to represent a judgement proves to be difficult and may be unacceptable. Information may be lost or distorted in the process of pre-aggregating different types of information, such as a subjective judgement, a probability distribution, or an incomplete piece of information.

Concerning the fuzziness of MADA problems, a large amount of fuzzy MADA methods have been proposed in the literature [3, 15, 16, 17, 9, 6, 13]. Nevertheless, these pure fuzzy MADA approaches are essentially based on traditional evaluation methods and are unable to handle probabilistic uncertainties such as ignorance.

Different from the traditional MADA methods, the Evidential Reasoning (ER) approach [21, 20, 19, 23, 24], which is the combination of the D-S theory [12] with a distributed modelling framework, shed a new line to modelling complex MADA problems. The ER approach uses a distributed modelling framework, in which each attribute is accessed using a set of collectively exhaustive and mutually exclusive assessment grades. Probabilistic uncertainty including local and global ignorance is characterized by a belief structure in the ER approach, which can both model precise data and capture various types of uncertainties such as probabilities and vagueness in subjective judgments. Along with the application of ER modelling, experiences show that decision maker may not always be confident enough to provide subjective assessments to individual grades only, but at times wishes to be able to assess beliefs to sub-sets of adjacent grades. Such ignorance is referred to as local ignorance or interval uncertainty. It is to deal with the local ignorance that the interval grade ER (IER) approach is proposed [14]. Another extension to the basic ER approach is to take account of vagueness or fuzzy uncertainty, i.e. the assessment grades are no longer clearly distinctive crisp sets, but are defined as dependent fuzzy sets. In other words, the intersection of two fuzzy sets may not be empty. Yang et al [22] proposed the fuzzy ER approach.

The aim of this paper is to integrate the main features of the above two approaches, and develop a general ER modelling framework and an attribute aggregation process, referred to as the fuzzy IER (FIER) algorithm, in order to deal with both fuzzy and interval grade assessments in MADA and provide a more powerful means to support the solution of complex MADA problems.

2 The FIER Approach for MADA under Fuzzy Uncertainty

2.1 The New FIER Distributed Modelling Framework using the Fuzzy Belief Structure

Suppose a MADA problem has M alternatives $a_l, l = 1, ..., M$, one upper level attribute, referred to as general attribute, and L lower level attributes

$e_i, i = 1, ..., L$, called basic attributes. The relative weights of the L basic attributes are denoted by $W = (w_1, ..., w_L)$, which are known and satisfy the conditions $0 \leq w_i \leq 1$ and $\sum_{i=1}^{L} w_i = 1$.

Suppose M alternatives are all assessed using a pre-defined set H. In basic ER methodology, set H is defined as a union of N assessment grades $H_p, p = 1, ..., N$, which are mutually exclusive and collectively exhaustive for the assessment of all attributes, and the whole set H_{1N} as follows:

$$H = \{H_1, H_2, ..., H_N, H_{1N}\} \tag{1}$$

According to [14], in the IER (Interval grade ER) methodology, the performances of alternatives can be assessed to an individual grade or a grade interval, the complete set of all individual grades and grade intervals, for assessing each attribute can be represented by

$$H = \begin{Bmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ & H_{22} & \dots & H_{2N} \\ & & \dots & \dots \\ & & & H_{NN} \end{Bmatrix} \tag{2}$$

where $H_{pp}, p = 1, ..., N$ in formula (2) denotes an individual grade. $H_{pq}, p = 1, ..., N, q = p+1, ..., N$ denotes the interval grade which is the union of individual grades $H_{pp}, H_{(p+1)(p+1)}, ..., H_{qq}$.

In the basic ER as well as the IER approach introduced above, all individual and interval assessment grades are assumed to be crisp and independent of each other. However, there are occasions where an assessment grade may represent a vague concept or standard and there may be no clear cut between the meanings of two adjacent grades. In this paper, we will drop the above assumption and allow grades to be vague and adjacent grades to be dependent. To simplify the discussion and without loss of generality, fuzzy sets will be used to characterize vague assessment grades and it is assumed that only two adjacent fuzzy grades have the overlap of meanings. This represents the most common features of fuzzy uncertainty in decision analysis.

In order to generalize the $H_{pq}, p = 1, ..., N, q = p, ..., N$ to fuzzy sets, we assume that a general set of fuzzy individual assessment grades $\{H_{pq}\}, p = 1, ..., N$ are dependent on each other, which may be assumed to be either triangular or trapezoidal fuzzy sets or their combinations for simplifying the discussion and without loss of generality. Assuming that only two adjacent fuzzy individual assessment grades may intersect, we denote by $\{H_{p\wedge(p+1)}\}, p = 1, ..., N-1$ the fuzzy intersection subset of the two adjacent fuzzy individual assessment grades H_{pp} and $H_{(p+1)(p+1)}$(Fig. 1).

Furthermore, we define the sets $H_{pq}, p = 1, ..., N, q = p, ..., N$ as trapezoidal fuzzy sets which include individual grades $H_{pp}, H_{(p+1)(p+1)}, ..., H_{qq}$. If these individual assessment grades are triangular or trapezoidal fuzzy sets, every interval grade will be a trapezoidal fuzzy set (Fig. 1b). And we also define $\{H_{p\wedge(p+1)}\}, p = 1, ..., N-1$ as the fuzzy intersection subset of the two adjacent fuzzy interval assessment grades H_{kp} and $H_{(p+1)q}$, where $k \leq p, q \geq p+1$ (Fig. 2).

Finally, the generalized fuzzy assessment set can be defined as follows.

$$H = H_F = \{H_{pq}, p = 1, ..., N, q = p, ..., N\} \cup \{H_{p \wedge q}, p = 1, ..., N-1\} \tag{3}$$

where H_{pq} is a fuzzy set and $H_{p\wedge(p+1)}$ is the intersection of two adjacent fuzzy sets H_{kp} and $H_{(p+1)q}$, where $k \leq p, q \geq p+1$.

The assessment of an alternative on attribute is then given by

$$S(a_l) = \{C, \beta_i(C)); C \in H, i = 1, ..., L\} \tag{4}$$

where $\sum_{C \in H} \beta_i(C) = 1$, for $i = 1, ..., L$ holds.

The mass functions are defined as follows:

$$m_i(C) = w_i \beta_i(C), i = 1, ..., L, C \neq \emptyset, C \in H \tag{5}$$

$$m_i(\emptyset) = 0 \tag{6}$$

$$m_i(U) = 1 - w_i, i = 1, ..., L \tag{7}$$

where $m_i(U)$ in equation (7) is the remaining probability mass that is unassigned to any evaluation grades in set H after only attribute i has been taken into account. In other words, $m_i(U)$ represents the remaining role that other attributes can play in the assessment. $m_i(U)$ should eventually be assigned back to set H, in a way that is dependent upon the importance of other attributes.

2.2 The New FIER Algorithm under Both Interval Probabilistic and Fuzzy Uncertainties

Based on the fuzzy assessment set H_F, a FIER (Fuzzy Interval grade ER) recursive algorithm is developed as follows using the similar technique used in [21, 22].

$$\widetilde{m}_{I(1)}(H_{pq}) = m_1(H_{pq}), p = 1, ..., N, q = p, ..., N \tag{8}$$

$$\widetilde{m}_{I(1)}(H_{p\wedge(p+1)}) = 0, p = 1, ..., N-1 \tag{9}$$

$$\widetilde{m}_{I(1)}(U) = m_1(U) \tag{10}$$

$$\begin{aligned}
\widetilde{m}_{I(i+1)}(H_{pq}) = & -\widetilde{m}_{I(i)}(H_{pq}) m_{i+1}(H_{pq}) \\
& + \sum_{k=1}^{p} \sum_{l=q}^{N} [\widetilde{m}_{I(i)}(H_{kl}) m_{i+1}(H_{pq}) + \widetilde{m}_{I(i)}(H_{pq}) m_{i+1}(H_{kl})] \\
& + \sum_{k=1}^{p-1} \sum_{l=q+1}^{N} [\widetilde{m}_{I(i)}(H_{kq}) m_{i+1}(H_{pl}) + \widetilde{m}_{I(i)}(H_{pl}) m_{i+1}(H_{kq})] \\
& + \widetilde{m}_{I(i)}(U) m_{i+1}(H_{pq}) + \widetilde{m}_{I(i)}(H_{pq}) m_{i+1}(U), \\
& p = 1, ..., N, q = p, ..., N
\end{aligned} \tag{11}$$

$$\widetilde{m}_{I(i+1)}(H_{p\wedge(p+1)}) = \sum_{k=1}^{p}\sum_{q=p+1}^{N}[\widetilde{m}_{I(i)}(H_{kp})m_{i+1}(H_{(p+1)q}) + \widetilde{m}_{I(i)}(H_{(p+1)q})m_{i+1}(H_{kp})] + \sum_{k=1}^{p+1}\sum_{l=p,l\geq k}^{N}\widetilde{m}_{I(i)}(H_{p\wedge(p+1)})m_{i+1}(H_{kl}) + \widetilde{m}_{I(i)}(H_{p\wedge(p+1)})m_{i+1}(U)$$
$$p = 1, ..., N-1 \tag{12}$$

$$\widetilde{m}_{I(i+1)}(U) = \widetilde{m}_{I(i)}(U)m_{i+1}(U) = \prod_{l=1}^{i+1} m_l(U) \tag{13}$$

$$K = 1/[\sum_{p=1}^{N}\sum_{q=p}^{N}\widetilde{m}_{I(L)}(H_{pq}) + \sum_{p=1}^{N-1}\mu_{p\wedge(p+1)}^{max}\cdot\widetilde{m}_{I(L)}(H_{p\wedge(p+1)}) + \widetilde{m}_{I(L)}(U)] \tag{14}$$

$$m(H_{pq}) = K\cdot\widetilde{m}_{I(L)}(H_{pq}) \tag{15}$$

$$m(\overline{H}_{p\wedge(p+1)}) = K\cdot\mu_{p\wedge(p+1)}^{max}\cdot\widetilde{m}_{I(L)}(H_{pq}) \tag{16}$$

$$m(U) = K\cdot\widetilde{m}_{I(L)}(U), p = 1, ..., N, q = p, ..., N \tag{17}$$

After the L attributes have been combined one-by-one using the above FIER algorithm, the overall assessment of an alternative a_l can be obtained as:

$$\beta(H_{pq}) = \frac{m(H_{pq})}{1 - m(U)}, p = 1, ..., N, q = p, ..., N \tag{18}$$

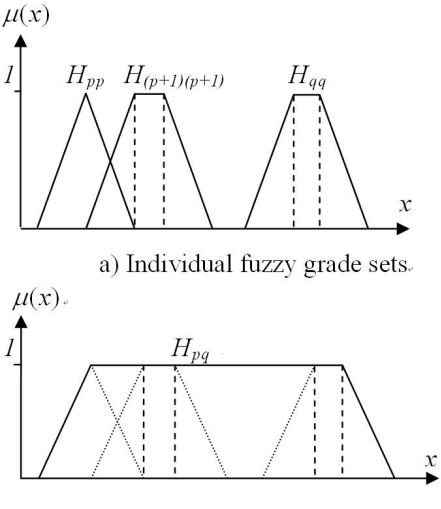
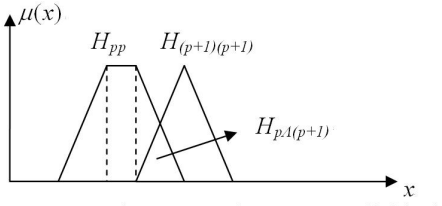
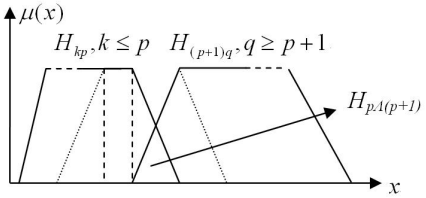


Fig. 1. Fuzzy sets definitions

$$\beta(\overline{H}_{p\wedge(p+1)}) = \frac{m(\overline{H}_{p\wedge(p+1)})}{1 - m(U)}, p = 1, ..., N, q = p, ..., N-1 \tag{19}$$



a) Intersection Sets between Individual Fuzzy Sets



b) Intersection Sets between Interval or Interval Fuzzy Sets

Fig. 2. Intersections between fuzzy assessment grades

3 Fuzzy Expected Utilities for Characterising Alternatives

Utility is one of the most important concepts in decision analysis. In fuzzy MADA, however, utilities corresponding to fuzzy assessment grades can no longer be represented by singleton numerical values because the evaluation grades are fuzzy sets. In general, a fuzzy grade utility should have the same form as its corresponding fuzzy assessment grade. For example, if a fuzzy assessment grade is a triangular fuzzy number, its corresponding fuzzy grade utility should also be a triangular fuzzy number. In the FIER methodology according to the definitions of fuzzy grades in section 2.1, the utility values of an interval fuzzy grade can be calculated from the utility values of the correspondent fuzzy individual grades as shown in Fig. 3. According to the basic ER methodology, the fuzzy expected utility of an aggregated assessment $S(y(a_l))$ for alternative a_l is defined as follows:

$$u(S(y(a_l))) = \sum_{p=1}^{N} \sum_{q=p}^{N} \beta(H_{pq}) u(H_{pq}) + \sum_{p=1}^{N-1} \beta(\overline{H}_{p\wedge(p+1)}) u(\overline{H}_{p\wedge(p+1)}) \tag{20}$$

where $u(H_{pq})$ is the fuzzy grade utility of the assessment grade H_{pq}, and $u(\overline{H}_{p\wedge(p+1)})$ is the fuzzy grade utility of the intersection fuzzy grade set $\overline{H}_{p\wedge(p+1)}$. Without loss of generality, for $p = 1, ..., N-1$, suppose $u(H_{pp})$ is the utility value of the grade H_{pp} with $u(H_{(p+1)(p+1)}) \geq u(H_{pp})$ as it is assumed that the grade $H_{(p+1)(p+1)}$ is preferred to H_{pp}. Suppose H_{11} is the least preferred fuzzy assessment grade, which has the lowest fuzzy grade utility, and H_{NN} is the most preferred fuzzy assessment grade, which has the highest fuzzy grade utility. Suppose $u(H_{pq})$ can take the lower bound value, the upper bound value and the two most possible values as $u_{min}(H_{pq})$, $u_{max}(H_{pq})$,$u_{MPV1}(H_{pq})$ and $u_{MPV2}(H_{pq})$ $(u_{MPV1}(H_{pq}) \leq u_{MPV2}(H_{pq}))$ respectively if all grade sets are triangular or trapezoidal fuzzy sets. It is straightforward that the following equations hold according to the relationships of individual and interval grade sets:

$$u_{min}(H_{pq}) = u_{min}(H_{pp}) \tag{21}$$

$$u_{max}(H_{pq}) = u_{max}(H_{qq}) \tag{22}$$

$$u_{MPV1}(H_{pq}) = u_{MPV1}(H_{pp}) \tag{23}$$

$$u_{MPV2}(H_{pq}) = u_{MPV2}(H_{qq}) \tag{24}$$

where in equation (22), the belief degree $\beta(H_{pq})$ could be assigned to the best grade in the interval grade H_{pq}, which is H_{qq}, and also can be assigned to the worst grade H_{pp} as shown in equation (21).

Similarly, suppose $u(\overline{H}_{p\wedge(p+1)})$ can take the lower bound value, the upper bound value and the two most possible values as $u_{min}(\overline{H}_{p\wedge(p+1)})$, $u_{max}(\overline{H}_{p\wedge(p+1)})$, $u_{MP1}(\overline{H}_{p\wedge(p+1)})$ and $u_{MP2}(\overline{H}_{p\wedge(p+1)})$ respectively, note $u_{MP1}(\overline{H}_{p\wedge(p+1)}) = u_{MP2}(\overline{H}_{p\bigwedge(p+1)})$, and the following equations hold:

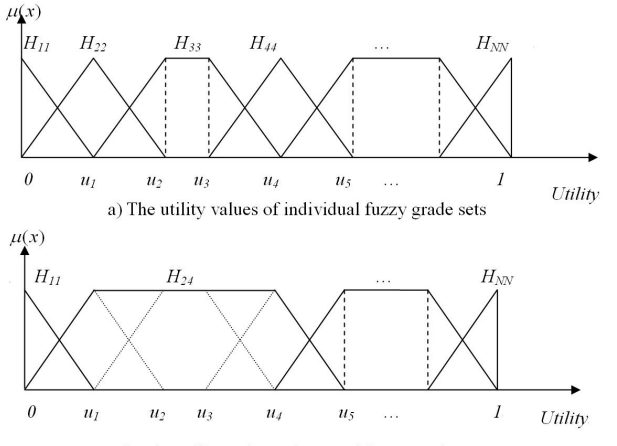


Fig. 3. The utility of fuzzy grades

$$u_{min}(\overline{H}_{p\wedge(p+1)}) = u_{min}(H_{(p+1)(p+1)}) \tag{25}$$

$$u_{max}(\overline{H}_{p\wedge(p+1)}) = u_{max}(H_{pp}) \tag{26}$$

Accordingly, the fuzzy expected utility is also a fuzzy number. From equations (20-26), the maximum utility value of alternative could be calculated as:

$$u_{max}(a_l) = \sum_{p=1}^{N}\sum_{q=p}^{N}\beta(H_{pq})u_{max}(H_{qq}) + \sum_{p=1}^{N-1}\beta(\overline{H}_{p\wedge(p+1)})u_{max}(\overline{H}_{pp}) \tag{27}$$

Similarly, in the worst case, if the uncertainty turned out to be against the assessed alternative, with the belief degree $\beta(H_{pq})$ being assigned to H_{pp} (the worst grade in the interval grade), then the minimum utility value would be given by:

$$u_{min}(a_l) = \sum_{p=1}^{N}\sum_{q=p}^{N}\beta(H_{pq})u_{min}(H_{pp}) + \sum_{p=1}^{N-1}\beta(\overline{H}_{p\wedge(p+1)})u_{min}(\overline{H}_{(p+1(p+1)}) \tag{28}$$

We can also define the two most possible utilities and their average value as follows:

$$u_{MPV1}(a_l) = \sum_{p=1}^{N}\sum_{q=p}^{N}\beta(H_{pq})u_{MPV1}(H_{pp}) + \sum_{p=1}^{N-1}\beta(\overline{H}_{p\wedge(p+1)})u_{MPV1}(\overline{H}_{(p+1(p+1)}) \tag{29}$$

$$u_{MPV2}(a_l) = \sum_{p=1}^{N}\sum_{q=p}^{N}\beta(H_{pq})u_{MPV2}(H_{qq}) + \sum_{p=1}^{N-1}\beta(\overline{H}_{p\wedge(p+1)})u_{MPV2}(\overline{H}_{(p+1(p+1)}) \tag{30}$$

$$u_{AVG-MPV}(a_l) = \frac{u_{MPV1}(a_l) + u_{MPV2}(a_l)}{2} \tag{31}$$

4 Application of the FIER Approach to a New Product Screening Problem

The company concerned is an electronic manufacturer, which manufactures a wide range of electronic entertainment products. Every year, the company identifies market requirements and comes up with a list of potential product development projects. Suppose there are three new computer game projects available: Motor Cycling, Sport Bass Fishin' and Play TV Baseball. However, at a preliminary design phase, the assessment of a project on multiple criteria is mainly based on experts' subject judgments. Experts' opinions may be expressed by belief degrees (or possibility measures) based on basic evaluation grades, i.e. Bad, Poor, Average, Good, and Excellent. As such, the basic evaluation grade set can be defined as a set H as follows:

$$H = \{H_{11}, H_{22}, H_{33}, H_{44}, H_{55}\} = \{\text{Bad, Poor, Average, Good, Excellent}\}$$

Due to the high level of uncertainty involved in this NPD problem, however, these evaluation grades may not be regarded as crisp sets. For example, it is difficult to separate the grade Bad from the grade Poor especially if evaluations need to be given between these two grades. Also it is not surprising that for some evaluations the experts prefer to give the belief degree measures on interval grades. For example, the TIMING for Sport Bass Fishin' is $(H_{34}, 1.0)$, which means that means that 100% belief is given to interval grade H_{34}, i.e. the worst assessment for Sport Bass Fishin' on TIMING is Average and the highest is Good. However, the exact belief degree to each of the two grades is not known. In a similar way, the incomplete opinions of the experts in evaluating this NPD problem can be captured conveniently by the following fuzzy evaluation grades.

$$H = H_F = \{H_{pq}, p = 1, ..., 5, q = p, ..., 5\} \cup \{H_{p\wedge(p+1)}, p = 1, ..., 4\}$$

Based on the experts' opinions, we can approximate all the five individual assessment grades by either triangular or trapezoidal fuzzy numbers as shown in table 3, and the maximum degree of membership for every fuzzy intersection set is 0.5.

By using our proposed FIER methodology, the aggregated performance distribution of all the three alternative projects can be calculated. The expected maximum and minimum utilities can also be calculated according to formulae (27)-(31), as shown in table 5. A final rank order can be obtained as follows.

Sport Bass Fishin' is possibly better than Motor Cycling and Play TV Baseball according to the average MPV values of all the three projects presented. However, it is obviously that Sport Bass Fishin' does not absolutely dominate the other two projects. This is because

$$\begin{aligned} u_{min}(SportBassFisin') = 0.3878 &< u_{max}(MotorCycling) = 0.7420 \\ &< u_{max}(PlayTVBaseball) = 0.7406 \end{aligned}$$

While in the sense of MPV dominance, we can obtain:

$$u_{MPV1}(SportBassFishin') = 0.5686 > u_{MPV2}(MotorCycling) = 0.5425$$

This means that Sport Bass Fishin' is preferred to Motor Cycling in the sense of MPV dominance, or

$$\text{Sport Bass Fishin'} \succ_{MPV} \text{Motor Cycling.}$$

Table 1. Membership functions of the fuzzy assessment grades and their fuzzy utilities

Linguistic term	Worst(W)	Poor(P)	Average(A)	Good(G)	Excellent(E)
Membership functions of fuzzy grade utilities	(0, 0, 0.2)	(0, 0.2, 0.4)	(0.2, 0.4, 0.6,0.8)	(0.6, 0.8, 1)	(0.8, 1, 1)

Table 2. Belief Matrix of the Performance Assessment Problem

Criteria	Weights	Motor Cycling	Sport Bass Fishin'	Play TV Baseball
TIMING	0.1	{(H44, 1.0)}	{(H34, 1.0) }	{(H12, 1.0)}
PRICE	0.1	{(H11, 1.0)}	{(H44, 0.9), (H15, 0.1)}	{(H45,0.9), (H15, 0.1)}
LOGISTICS	0.05	{(H44, 1.0)}	{(H45, 1.0)}	{(H45, 1.0)}
SALES	0.02	{(H33, 1.0)}	{(H33, 1.0)}	{(H22, 1.0)}
MFGTECH	0.02	{(H44, 1.0)}	{(H33, 0.6), (H44, 0.4)}	{(H22,1.0)}
MFGCAP	0.02	{(H44, 1.0)}	{(H44, 1.0)}	{(H45, 1.0)}
SUPPLY	0.05	{(H34, 1.0)}	{(H34, 1.0)}	{(H45, 1.0)}
DESIGN	0.1	{(H11, 1.0)}	{(H44, 0.8), (H15, 0.2)}	{(H45, 0.8), (H15,0.2)}
DIFFADV	0.08	{(H11, 1.0)}	{(H55, 1.0)}	{(H55, 1.0)}
PAYOFFS	0.08	{(H11, 1.0)}	{(H44, 0.8), (H15, 0.2)}	(H45,0.8), (H15, 0.2)
LOSSES	0.08	(H44, 1.0)	(H22, 0.9), (H15, 0.1)	(H23, 0.9), (H15, 0.1)
R&DUNC	0.25	{(H34, 1.0)}	{(H34, 0.9), (H15,0.1)}	{(H12, 0.8), (H15, 0.2)}
NONR&D	0.05	{(H44, 1.0)}	{(H33, 0.8), (H15,0.2)}	{(H12, 0.8), (H15, 0.2)}

Table 3. Fuzzy expected utilities and ranking order of alternatives

	Fuzzy expected utility				
	Lower bound	Most possible value		Upper bound	Avg. of MPV
Motor Cycling	0.2910	0.4273	0.5425	0.7420	0.4849
Sport Bass Fishin'	0.3878	0.5686	0.7809	0.9433	0.6747
Play TV Baseball	0.2831	0.3960	0.6458	0.7406	0.5209

5 Concluding Remarks

Incompleteness and fuzziness are among the most common uncertainties in complex MADA problems. The new development as reported in this paper further extends the capability of the ER approach to utilise information with both local ignorance or interval uncertainty and fuzzy linguistic evaluation grades. Expert judgements can be captured by our proposed FIER method in such a convenient way that the evaluations made by experts, which are incomplete and fuzzy in nature, do not need to be converted to some strictly defined formats that may inevitably lead to the loss of important information, as shown in some classical

MADA methods. In this sense, our FIER method can be used to deal with various types of uncertainties to help the DMs in making more informative decisions.

Similar with the previous ER approach, this FIER method is aimed to generate the preference orders of alternatives without having to gather perfect or complete information as is often done in real life decision making. However, the results obtained using the new methods may be an incomplete preference order as well due to the incompleteness and fuzziness in initial data, as illustrated in the example. In such cases, more information may be needed to support specific decision making such as finding a single winner in a performance assessment problem. Further research is needed to investigate the process of information gathering for sensitivity analysis.

Acknowledgement. The work described in this paper is supported by the Research Grant Council of the Hong Kong SAR, China under CERG Grant No: 9040786 and the UK Engineering and Physical Science Research Council under Grant No.: EPSRC GR/S85498/01. The work described in this paper was also partially supported by the grant from the Natural Science Foundation of China (NSFC) under Grant No.: 70540430127, 70572033 and 70631003, and the grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, project No. CityU 1203/04E and CityU 111906.

References

1. Arbel, A., Vargas, L.G.: The analytic hierarchy process with interval judgments. In: Goicoechea, A., Duckstein, L., Zoints, S. (eds.) Multiple Criteria Decision Making. Proceedings of the Ninth International Conference held in Fairfax, Virginia, pp. 61–70. Springer, New York (1992)
2. Arbel, A., Vargas, L.G.: Preference simulation and preference programming: Robustness issues in priority deviation. European Journal of Operational Research 69, 200–209 (1993)
3. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. Management Science 17(4), 141–164 (1970)
4. Belton, V., Stewart, T.J.: Multiple Criteria Decision Analysis - An Integrated Approach. Kluwer Academic Publishers, Dordrecht (2002)
5. Chin, K.S., et al.: An evidential reasoning-interval based method for new product design assessment. IEEE Transactions on Engineering Management (in press, 2007)
6. Dong, W.M., Wong, F.S.: Fuzzy weighted averages and implementation of the extension principle. Fuzzy Sets and Systems 21, 183–199 (1987)
7. Islam, R., Biswal, M.P., Alam, S.S.: Preference programming and inconsistent interval judgments. European Journal of Operational Research 97, 53–62 (1997)
8. Keeney, R.L., Raiffa, H.: Decisions with Multiple Objectives, 2nd edn. Cambridge University Press, Cambridge (1993)
9. Laarhoven, P.J.M., Pedrycz, W.: A fuzzy extension of Saaty's priority theory. Fuzzy Sets and Systems 11, 229–241 (1983)
10. Saaty, T.L.: The Analytic Hierarchy Process. University of Pittsburgh (1988)

11. Salo, A., Hämäläinen, R.P.: Processing interval judgments in the analytic hierarchy process. In: Goicoechea, A., Duckstein, L., Zoints, S. (eds.) Multiple criteria decision making. Proceedings of the Ninth International Conference held in Fairfax, Virginia, pp. 359–372. Springer, New York (1992)
12. Shafer, G.A.: Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ (1976)
13. Tseng, T.Y., Klein, C.M.: A new algorithm for fuzzy multicriteria decision making. International Journal of Approximate Reasoning 6, 45–66 (1992)
14. Xu, D.L., Yang, J.B., Wang, Y.M.: The evidential reasoning approach for multi-attribute decision analysis under interval uncertainty. European Journal of Operational Research 174(3), 1914–1943 (2005)
15. Yager, R.R.: Multiple objective decision-making using fuzzy sets. International Journal of Man-Machine Studies 9, 375–382 (1977)
16. Yager, R.R.: Fuzzy decision making including unequal objectives. Fuzzy Sets and Systems 1, 87–95 (1978)
17. Yager, R.R.: A new methodology for ordinal multi-objective decisions based on fuzzy sets. Decision Science 12, 589–600 (1981)
18. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems Man Cybernetics 18(1), 183–190 (1988)
19. Yang, J.B.: Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. European Journal of Operational Research 131(1), 31–61 (2001)
20. Yang, J.B., Sen, P.: A general multi-level evaluation process for hybrid MADM with uncertainty. IEEE Transactions on Systems, Man, and Cybernetics 24(10), 1458–1473 (1994)
21. Yang, J.B., Singh, M.G.: An evidential reasoning approach for multiple attribute decision making with uncertainty. IEEE Transactions on Systems, Man, and Cybernetics 24(1), 1–18 (1994)
22. Yang, J.B., et al.: The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties. European Journal of Operational Research 171, 309–343 (2006)
23. Yang, J.B., Xu, D.L.: On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 32(3), 289–304 (2002)
24. Yang, J.B., Xu, D.L.: Nonlinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 32(3), 376–393 (2002)
25. Yen, J.: Generalizing the Dempster-Shafer theory to fuzzy sets. IEEE Transactions on Systems, Man, and Cybernetics 20(3), 559–570 (1990)
26. Zadeh, L.A.: The concepts of a linguistic variable and its application to approximate reasoning (I) (II) (III). Information Science 8, 199–249, 301–357; 9, 43–80 (1975)
27. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1(1), 3–28 (1978)

Modelling and Computing with Imprecise and Uncertain Properties in Object Bases

Tru H. Cao[1], Hoa Nguyen[2], and Ma Nam[1]

[1] Ho Chi Minh City University of Technology
{tru, manam}@cse.hcmut.edu.vn
[2] Ho Chi Minh City Open University
hoa-hanh@hcm.vnn.vn

Summary. Although fuzzy set and probability theories are complementary for dealing with pervasive imprecision and uncertainty in real world problems, object-oriented database models combining the relevance and strength of both the theories appear to be sporadic. This paper introduces our extension of Eiter et al.'s probabilistic object base model with two key features: (1) uncertain and imprecise attribute values are represented as probability distributions on a set of fuzzy set values; and (2) class methods with uncertain and imprecise input and output arguments are formally integrated into the new model. A probabilistic interpretation of relations on fuzzy set values is proposed for their combination with probability degrees. Then the syntax and semantics of fuzzy-probabilistic object base schemas, instances, and selection operation are defined. Furthermore, the soft computing paradigm needs to have real systems implemented to be useful in practice. This paper also presents our development of FPDB4O as a management system for fuzzy and probabilistic object bases of the proposed model.

1 Introduction

For modelling real-world problems and constructing intelligent systems, integration of different methodologies and techniques has been the quest and focus of significant interdisciplinary research effort. The advantages of such a hybrid system are that the strengths of its partners are combined and complementary to each other's weakness.

In particular, object orientation provides a hierarchical data abstraction scheme and an information hiding and inheritance mechanism. Meanwhile, probability theory and fuzzy logic provide measures and rules for representing and reasoning with uncertainty and imprecision in the real world. Many uncertain and fuzzy object-oriented data models have been proposed and developed (cf. [24], [6], [5], [9], [11]).

The key issues in research on extending the classical object-oriented data models to deal with uncertainty and imprecision in the real world are:

1. Modelling partial subclass relationship.
2. Definition of partial class membership.
3. Representation of uncertain and/or imprecise attribute values.
4. Representation and execution of class methods.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 141–159, 2008.
springerlink.com

5. Expression of partial applicability of class properties.
6. Mechanism for inheritance under uncertainty and imprecision.

This paper focuses on the first four issues only.

Regarding partial subclass relationship, there are models defining inclusion degrees between classes ([15], [23], [24]). However, as discussed in [1], a set of classes with a graded inclusion or inheritance relation actually forms a network rather than a hierarchy, because if a class A has some inclusion degree to a class B based on a fuzzy matching of their descriptions, then B usually also has some inclusion degree to A. Moreover, naturally, a concept is usually classified into sub-concepts that are totally subsumed by it, though the sub-concepts can overlap each other (cf. [12]).

Uncertain and imprecise attribute values lead to partial membership of an object to a class, and there are different measures proposed. In [25], for instance, a membership function on a set of objects was defined for each class. In [6], linguistic labels were used to express the strength of the link of an object to a class. In [12], membership was defined as similarity degrees between objects and classes. Meanwhile, [5] mentioned different measures, including probabilistic one, to be used for membership degrees. Nevertheless, it is to be answered how measures of different meanings, such as possibility and probability, on various levels of a model are integrated coherently.

Most of the works on fuzzy object-oriented data models, which are referred in this paper, were mainly based on fuzzy set and possibility theories and used fuzzy sets or possibility distributions to represent imprecise attribute values. In [6] and [5], the authors also modelled uncertainty about an attribute having a particular value. However, much less concern was given for uncertainty over a set of values of an attribute and a foundation to combine probability degrees and fuzzy sets in the same model.

While class attributes were paid much attention and treatment, class methods, as common in object-oriented systems for modelling object behaviours and parameterized properties, were often neglected. In [12] and [5] methods were not considered. The authors of [6] mentioned about methods but did not provide formal representation and explicit manipulation in their model. In [25] and [9] methods were formally defined as Horn clauses and executed as a reasoning process, but those models were thus deductive in contrast to imperative ones.

Recently, [10] reviewed existing proposals and presented recommendations for the application of fuzzy set theory in a flexible generalized object model. Furthermore, [11] focused on representing data as constraints on object attributes and query answering as constraint satisfaction. For realization of fuzzy object-oriented data models, [4] was concerned with implementation of their model on an existing platform.

Meanwhile, [13] introduced a probabilistic model to handle object bases with uncertainty, called POB (Probabilistic Object Bases). For a POB class hierarchy, although a class was assumed to be fully included in its super-classes, the model specified the conditional probability for an object of a class belonging to each of its subclasses. Intuitively, it specified how likely an object of a class belonged to a

subclass of that class. Accordingly, the partial class membership was measured by probability degrees. For each attribute of an object, uncertainty about its value was represented by lower bound and upper bound probability distributions on a set of values. The authors also developed a full-fledged algebra to query and operate on object bases.

However, the two major shortcomings of the POB model are: (1) it does not allow imprecise attribute values; and (2) it does not consider class methods. For instance, in the Plant example therein, the values of the attribute *sun light* are chosen to be only enumerated symbols such as *mild, medium,* and *heavy* without any interpretation. Meanwhile, in practice, those values are inherently vague and imprecise over degrees of sun light. Moreover, without an interpretation, they cannot be measured and their probability distributions calculated.

Since fuzzy set theory and fuzzy logic provide a basis for defining the semantics of, and computing with, linguistic terms ([26]), [8] has applied them to extend the POB model to represent and compute with imprecise attribute values, called FPOB (Fuzzy Probabilistic Object Bases). For instance, the values *mild, medium,* and *heavy* of the attribute *sun light* above could be defined by fuzzy sets. In this paper, as a further extension of FPOB, we integrate class methods into the model.

We argue that, due to complexities of theoretical foundation and practical implementation, no model would be so universal that could include all measures and tackle all aspects of uncertainty and imprecision. Our work is thus not to supersede previous models, but rather to complement them to deal with certain facets of the complex real world. Moreover, the soft computing paradigm needs to have real systems implemented to be useful in practice. So, we have developed FPDB4O as a management system for FPOB, to be used for real world applications.

As a basis for combining probabilities and fuzzy sets, Section 2 introduces a probabilistic interpretation of relations on fuzzy sets and an abstract algebra on fuzzy probabilistic triples. Sections 3 and 4 extend the notions of FPOB types, schemas, and instances with class methods. Section 5 presents the syntax and semantics of FPOB selection operation under imprecision and uncertainty. Section 6 describes FPDB4O with run examples. Finally, Section 7 concludes the paper and suggests further research.

2 Combination of Probabilities and Fuzzy Sets

2.1 Probabilistic Interpretation of Relations on Fuzzy Sets

In this work, for combining fuzzy set values with probabilities, we apply the voting model interpretation of fuzzy sets ([14], [2]). That is, given a fuzzy set A on a domain U, each voter has a subset of U as his/her own crisp definition of the concept that A represents. The membership function value $\mu_A(u)$ is then the proportion of voters whose crisp definitions include u. This model defines a *mass assignment* (i.e., probability distribution) on the power set of U, where the mass (i.e., probability value) assigned to a subset of U is the proportion of voters who

have that subset as a crisp definition for the fuzzy concept A. As such, this mass assignment corresponds to a family of probability distributions on U. We now introduce a probabilistic interpretation of the following binary relations on fuzzy sets. We write $\Pr(E_1 \mid E_2)$ to denote the conditional probability of E_1 given E_2.

Definition 1. *Let A be a fuzzy set on a domain U, B be a fuzzy set on a domain V, and θ be a binary relation from $\{=, \neq, \leq, <, \subseteq, \in\}$ assumed to be valid on $(U \times V)$. The* probabilistic interpretation *of a relation $A\ \theta\ B$, denoted by $prob(A\ \theta\ B)$, is a value in [0, 1] that is defined by $\Sigma_{S \subseteq U, T \subseteq V} Pr(u \theta v \mid u \in S, v \in T).m_A(S).m_B(T)$.*

Intuitively, given fuzzy propositions $x \in A$ and $y \in B$, $prob(A\ \theta\ B)$ is the probability for $x\ \theta\ y$ being true. The rationale of the above probabilistic interpretation is that, given each crisp definition S of A and T of B, the conditional probability $u\ \theta\ v$ given $u \in S$ and $v \in T$ is calculated and weighted by the product of the masses associated with S and T. Then $prob(A\ \theta\ B)$ is the sum of those weighted conditional probability values.

Definition 2. *Let A and B be two fuzzy sets on a domain U. The* probabilistic interpretation *of the relation $A \rightarrow B$, denoted by $prob(A \rightarrow B)$, is a value in [0, 1] that is defined by $\Sigma_{S,T \subseteq U}\ Pr(u \in T \mid u \in S).m_A(S).m_B(T)$.*

The intuitive meaning of $prob(A \rightarrow B)$ is that it is the probability for $x \in B$ being true given $x \in A$ being true. In other words, it is the fuzzy conditional probability of $x \in B$ given $x \in A$. We note that the above probabilistic interpretation can also be adapted for fuzzy sets on continuous domains, using integration instead of addition, as in the definition of fuzzy conditional probability in [3].

2.2 Algebra on Fuzzy Probabilistic Triples

For representing uncertainty about the value of an attribute in POB, [13] introduced the notion of probabilistic triples. In FPOB, [8] has extended that notion to be fuzzy probabilistic triples with fuzzy set values.

Definition 3. *Let $dom(\tau)$ be the set of values of a type τ. A* fuzzy probabilistic triple *of type τ is defined to be of the form $\langle V, \alpha, \beta \rangle$, where $V \in dom(\tau)$ and α and β are lower and upper bound probability distributions on V.*

Intuitively, if the value of an attribute is uncertainly defined by $\langle V, \alpha, \beta \rangle$ then, for each $v \in V$, the probability for that attribute taking the certain value v is between $\alpha(v)$ and $\beta(v)$. A single and certain value v could be considered as a special fuzzy probabilistic triple, written as $\langle \{v\}, u, u \rangle$ where u denotes the uniform distribution, i.e., $u(v) = 1$.

Example 1. Suppose that the time it takes to send a letter by express airmail from Saigon to Hanoi is about 48 or 72 hours with a probability between .4 and .6. Then, it can be represented by the fuzzy probabilistic triple $\langle \{about\_48, about\_72\}, .8u, 1.2u \rangle$, where *about_48* and *about_72* are linguistic

labels of fuzzy sets, u is the uniform distribution, and $.8u$ and $1.2u$ denote the distribution functions $.8u(x) = .8/2 = .4$ and $1.2u(x) = 1.2/2 = .6$ for every x from *about_48, about_72*.

In this work, for introducing methods into the FPOB model, we define an abstract algebra with operations on fuzzy probabilistic triples. We employ alternative probabilistic combination strategies in [18] and [13], where $\otimes$ and $\oplus$ denote alternative conjunction and disjunction operators, respectively.

Definition 4. *Let* $\boldsymbol{\mathcal{V}} = \{\langle V, \alpha, \beta\rangle \mid V \subseteq dom(\tau)\}$ *be a non-empty set of fuzzy probabilistic triples of type* τ. *If* $\boldsymbol{A} = (dom(\tau), o_1, o_2, \ldots, o_n)$ *is a fuzzy set algebra with operations* $o_1, o_2, \ldots, o_n$ *on* $dom(\tau)$, *then* $\boldsymbol{\mathcal{A}} = (\boldsymbol{\mathcal{V}}, o_1, o_2, \ldots, o_n)$ *is a* fuzzy probabilistic triple algebra, *in which the operations* $o_1, o_2, \ldots, o_n$ *on* $\boldsymbol{\mathcal{V}}$ *are derived from* $\boldsymbol{A}$ *as follows:*

(i) $o_i(\langle V_1, \alpha_1, \beta_1\rangle, \langle V_2, \alpha_2, \beta_2\rangle, \ldots, \langle V_{m_i}, \alpha_{m_i}, \beta_{m_i}\rangle) = \langle V, \alpha, \beta\rangle$, *where* $V = \{v = o_i(v_1, v_2, \ldots, v_{m_i}) \mid v_j \in V_j$, *for* j *from 1 to* m_i, *and*

(ii) $[\alpha(v), \beta(v)] = \oplus_{me\ v_1 \in V_1, v_2 \in V_2, \ldots, v_{m_i} \in V_{m_i}, v = o_i(v_1, v_2, \ldots, v_{m_i})} [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)] \otimes \ldots \otimes [\alpha_{m_i}(v_{m_i}), \beta_{m_i}(v_{m_i})]$, *for every* $v \in V$.

In the definition of $[\alpha(v), \beta(v)]$ above, since there can be more than one tuple $(v_1, v_2, \ldots, v_{m_i})$ such that $v = o_i(v_1, v_2, \ldots, v_{m_i})$, the probabilistic conjunctions of those tuples have to be aggregated. This is in agreement with [22]. Also, obviously, the algebra $\boldsymbol{\mathcal{A}}$ has similar properties as the algebra $\boldsymbol{A}$. For example, if o_i's are commutative in $\boldsymbol{A}$, then so are o_i's in $\boldsymbol{\mathcal{A}}$.

Example 2. Let $\{\mathbf{real}\}$ denote the fuzzy real number type, $\boldsymbol{\mathcal{V}} = \{\langle V, \alpha, \beta\rangle \mid V \subseteq dom(\{\mathbf{real}\})\}$ be the set of fuzzy probabilistic triples of type $\{\mathbf{real}\}$, and $\boldsymbol{A} = (dom(\{\mathbf{real}\}), +, \times)$ be an algebra on $dom(\{\mathbf{real}\})$ with two fuzzy arithmetic operations $+$ and $\times$ using the extension principle ([17]). That is, each $v \in V$ is a fuzzy set on real numbers. Then $\boldsymbol{\mathcal{A}} = (\boldsymbol{\mathcal{V}}, +, \times)$ is an algebra on $\boldsymbol{\mathcal{V}}$ with two operations $+$ and $\times$ defined as follows, where $\oplus_{me}$ denotes the mutual exclusion conjunction operator:

(i) $\langle V_1, \alpha_1, \beta_1\rangle + \langle V_2, \alpha_2, \beta_2\rangle = \langle V, \alpha, \beta\rangle$, where $V = \{v = v_1 + v_2 \mid v_1 \in V_1, v_2 \in V_2\}$ and $(v_1 + v_2)(z) = sup_{z=x+y} min[v_1(x), v_2(y)]$, for all real numbers x, y, z, and $[\alpha(v), \beta(v)] = \oplus_{me\ v_1 \in V_1, v_2 \in V_2, v = v_1 + v_2} [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)]$, for every $v \in V$.

(ii) $\langle V_1, \alpha_1, \beta_1\rangle \times \langle V_2, \alpha_2, \beta_2\rangle = \langle V, \alpha, \beta\rangle$, where $V = \{v = v_1 \times v_2 \mid v_1 \in V_1, v_2 \in V_2\}$ and $(v_1 \times v_2)(z) = sup_{z=x\times y} min[v_1(x), v_2(y)]$, for all real numbers x, y, z, and $[\alpha(v), \beta(v)] = \oplus_{me\ v_1 \in V_1, v_2 \in V_2, v = v_1 \times v_2} [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)]$, for every $v \in V$.

3 Fuzzy and Probabilistic Object Properties

3.1 FPOB Class Hierarchy

For FPOB we use the same definition of class hierarchy as for POB. Figure 1 shows an example FPOB hierarchy of postal packages that are classified as being

either letters, boxes, or tubes and, alternatively, as being normal or priority ones. Those subclasses of a class that are connected to a **d** node are mutually disjoint (i.e., an object cannot belong to any two of them at the same time) and they form a cluster of that class. In this example, the class PACKAGE has two clusters, namely LETTER, BOX, TUBE and NORMAL, PRIORITY. A value in [0, 1] associated with the link between a class and one of its immediate subclasses represents the probability for an object of the class belonging to that subclass.

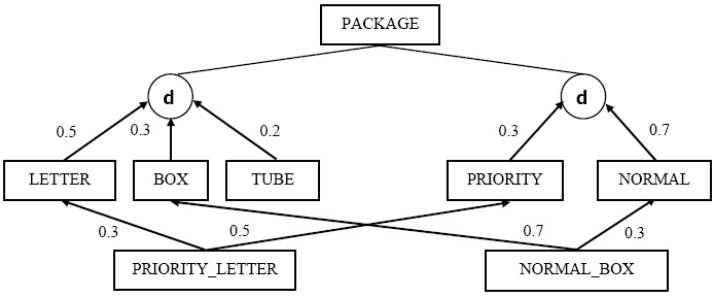


Fig. 1. An example FPOB class hierarchy

3.2 FPOB Attributes and Methods

As in the classical object-oriented model, in [8] each class in FPOB is characterized by a number of attributes whose values are of particular types, including fuzzy set types. Here, we extend that FPOB model further with methods, and use the common term property to refer to both attributes and methods. Each property has its type and value. For a method, its type and value are those of its output, which is defined as a function of the input arguments of the method. For a unified treatment of attributes and methods, an attribute could be considered as a special method with a fixed output, having no input argument. Alternatively, a method could be considered as a parameterized attribute, whose value depends on its input arguments. The following definitions and examples explain these ideas.

Definition 5. *Let $\boldsymbol{\mathcal{P}}$ be a set of properties and $\boldsymbol{\mathcal{T}}$ be a set of atomic types. Then types are inductively defined as follows:*

(i) Every atomic type from $\boldsymbol{\mathcal{T}}$ is a type.
(ii) If τ is a type, then $\{\tau\}$ is the fuzzy set type *of τ.*
(iii) If $P_1, P_2, \ldots, P_k$ are pairwise different properties from $\boldsymbol{\mathcal{P}}, \tau_1, \tau_2, \ldots, \tau_k$, and τ_{ij}'s, for every i from 1 to k and j from 1 to n_i, are types, then $\tau = [P_1(\tau_{11}, \tau_{12}, \ldots, \tau_{1n_1}) : \tau_1, P_2(\tau_{21}, \tau_{22}, \ldots, \tau_{2n_2}) : \tau_2, \ldots, P_k(\tau_{k1}, \tau_{k2}, \ldots, \tau_{kn_k}) : \tau_k]$ is the tuple type over $\{P_1, P_2, \ldots, P_k\}$. One writes $\tau.P_i$ to denote τ_i, and $P_1, P_2, \ldots, P_k$ are called top-level properties of τ.

In the definition above, τ_{ij}'s represent the types of the input arguments of P_i when it is a method, and they are null when it is an attribute. However, τ_{ij}'s can

also be omitted when it is not necessary to mention them in a certain context. Each type has a domain of its values as defined below (cf. [13], [8]).

Definition 6. *Let every atomic type $\tau \in \boldsymbol{\mathcal{T}}$ be associated with a domain $dom(\tau)$. Then values are defined by induction as follows:*

(i) For every $\tau \in \boldsymbol{\mathcal{T}}$, every $v \in dom(\tau)$ is a value of type τ.
(ii) For every $\tau \in \boldsymbol{\mathcal{T}}$, every fuzzy set on $dom(\tau)$ is a value of type $\{\tau\}$.
(iii) If $P_1, P_2, \ldots, P_k$ are pairwise different properties from $\boldsymbol{\mathcal{P}}$ and $v_1, v_2, \ldots, v_k$ are values of types $\tau_1, \tau_2, \ldots, \tau_k$, then $[P_1 : v_1, P_2 : v_2, \ldots, P_k : v_k]$ is a value of type $[P_1 : \tau_1, P_2 : \tau_2, \ldots, P_k : \tau_k]$.

Regarding fuzzy set values, we recall that a crisp set, or a single value, on a domain U could be considered as a special fuzzy set on U. As such, all values of a type could be treated uniformly as fuzzy sets of that type. Therefore, the FPOB model includes the classical object base model as a special case.

In [8], imprecision of attribute values is expressed by fuzzy sets. Meanwhile, uncertainty of attribute values is expressed by probabilistic triples. Here, that imprecision and uncertainty modelling is extended for values of input and output arguments of a method as well.

Definition 7. *Let $P_1, P_2, \ldots, P_k$ be pairwise different properties from $\boldsymbol{\mathcal{P}}$, V_i and V_{ij} be finite sets of values of types τ_i and τ_{ij}, $[\alpha_i, \beta_i]$ and $[\alpha_{ij}, \beta_{ij}]$ be pairs of probability distributions over V_i and V_{ij}, for every i from 1 to k and j from 1 to n_i. Then $fptv = [P_1(\langle V_{11}, \alpha_{11}, \beta_{11}\rangle, \langle V_{12}, \alpha_{12}, \beta_{12}\rangle, \ldots, \langle V_{1n_1}, \alpha_{1n_1}, \beta_{1n_1}\rangle) : \langle V_1, \alpha_1, \beta_1\rangle$, $P_2(\langle V_{21}, \alpha_{21}, \beta_{21}\rangle, \langle V_{22}, \alpha_{22}, \beta_{22}\rangle, \ldots, \langle V_{2n_2}, \alpha_{2n_2}, \beta_{2n_2}\rangle) : \langle V_2, \alpha_2, \beta_2\rangle, \ldots,$ $P_k(\langle V_{k1}, \alpha_{k1}, \beta_{k1}\rangle, \langle V_{k2}, \alpha_{k2}, \beta_{k2}\rangle, \ldots, \langle V_{kn_k}, \alpha_{kn_k}, \beta_{kn_k}\rangle) : \langle V_k, \alpha_k, \beta_k\rangle]$ is a* fuzzy-probabilistic tuple value *of type $[P_1(\tau_{11}, \tau_{12}, \ldots, \tau_{1n_1}) : \tau_1, P_2(\tau_{21}, \tau_{22}, \ldots, \tau_{2n_2}) : \tau_2, \ldots, P_k(\tau_{k1}, \tau_{k2}, \ldots, \tau_{kn_k}) : \tau_k]$ over $\{P_1, P_2, \ldots, P_k\}$. One writes $fptv.P_i$ to denote $\langle V_i, \alpha_i, \beta_i\rangle$.*

As such, a method could be considered as a function whose input and output arguments take fuzzy probabilistic triple values.

3.3 FPOB Schema

FPOB schemas are now defined as follows, extending the definitions in [13] and [8] with methods.

Definition 8. *An* FPOB schema *is a hextuple $(\boldsymbol{\mathcal{C}}, \tau, \Rightarrow, me, p, f)$ where:*

(i) $\boldsymbol{\mathcal{C}}$ is a finite set of classes.
(ii) τ maps each class to a tuple type $\tau(c)$ representing the properties and their types of that class.
(iii) $\Rightarrow$ is a binary relation on $\boldsymbol{\mathcal{C}}$ such that $(\boldsymbol{\mathcal{C}}, \Rightarrow)$ is a directed acyclic graph, whereby each edge $c_1 \Rightarrow c_2$ means c_1 is an immediate subclass of c_2.
(iv) me maps each class $c \in \boldsymbol{\mathcal{C}}$ to a partition of the set of all immediate subclasses of c, such that the classes in each cluster of the partition $me(c)$ are mutually disjoint.

(v) p maps each edge $c_1 \Rightarrow c_2$ in $(\mathcal{C}, \Rightarrow)$ to a rational number $p(c_1 \mid c_2)$ in [0, 1] measuring the conditional probability for an object picked at random uniformly from c_2 belonging to c_1.

(vi) f maps each method $P_i(\tau_{i1}, \tau_{i2}, \ldots, \tau_{in_i}) : \tau_i$ to a function from products of fuzzy probabilistic triples of types τ_{ij}'s to fuzzy probabilistic triples of type τ_i.

Given $c_1 \Rightarrow c_2 \Rightarrow \ldots \Rightarrow c_k$, one can write $c_1 \Rightarrow^* c_k$, and in particular $c \Rightarrow^* c$ for every $c \in \mathcal{C}$.

Example 3. An FPOB schema for the Package example above may be defined as follows:

$\mathcal{C}$ = { PACKAGE, LETTER, BOX, TUBE, PRIORITY, NORMAL, PRIORITY_LETTER, NORMAL_BOX}

τ is given as in Table 1.

$(\mathcal{C}, \Rightarrow)$, *me*, and p are given as in Figure 1.

f defines the methods area and volume using an algebra on fuzzy probabilistic triples introduced above as follows:

LETTER: area([length: $\langle V_1, \alpha_1, \beta_1 \rangle$], [width: $\langle V_2, \alpha_2, \beta_2 \rangle$]): $\langle V, \alpha, \beta \rangle$

1. $\langle V, \alpha, \beta \rangle = \langle V_1, \alpha_1, \beta_1 \rangle \times \langle V_2, \alpha_2, \beta_2 \rangle$
2. **return** $\langle V, \alpha, \beta \rangle$.

BOX: volume([length: $\langle V_1, \alpha_1, \beta_1 \rangle$], [width: $\langle V_2, \alpha_2, \beta_2 \rangle$], [height: $\langle V_3, \alpha_3, \beta_3 \rangle$]) : $\langle V, \alpha, \beta \rangle$

1. $\langle V, \alpha, \beta \rangle = \langle V_1, \alpha_1, \beta_1 \rangle \times \langle V_2, \alpha_2, \beta_2 \rangle \times \langle V_3, \alpha_3, \beta_3 \rangle$
2. **return** $\langle V, \alpha, \beta \rangle$.

TUBE: volume([diameter: $\langle V_1, \alpha_1, \beta_1 \rangle$], [width: $\langle V_2, \alpha_2, \beta_2 \rangle$]): $\langle V, \alpha, \beta \rangle$

1. $\langle V, \alpha, \beta \rangle = (\langle V_1, \alpha_1, \beta_1 \rangle / 2)^2 \times \langle V_2, \alpha_2, \beta_2 \rangle$
2. **return** $\langle V, \alpha, \beta \rangle$.

Table 1. Type assignment of the Package example

c	$\tau(c)$
PACKAGE	[origin: **string**, destination: **string**, time: {**real**}]
LETTER	[length: {**real**}, width: {**real**}, area(length: {**real**}], [width: {**real**}]): {**real**}]
BOX	[length: {**real**}, width: {**real**}, height: {**real**},volume([length: {**real**}], [width: {**real**}], [height: {**real**}]): {**real**}]
TUBE	[diameter: {**real**}, height: {**real**}, volume([diameter: {**real**}], [height: {**real**}]): {**real**}]
PRIORITY	[priority_level: **integer**]
NORMAL	[stop_over: **string**]
PRIORITY_LETTER	[]
NORMAL_BOX	[]

We note that, in Definition 8 and Example 3 of FPOB schemas, the properties specified for a class are only the *top-level* ones of that class, which do not include those properties inherited from its super-classes.

4 Fuzzy and Probabilistic Object Base Instances and Class Extents

4.1 FPOB Instances

Given an FPOB schema, an FPOB instance is defined as a base of objects associated with their classes and fuzzy-probabilistic tuple values in accordance with the schema.

Definition 9. *Let* $\boldsymbol{S} = (\boldsymbol{\mathcal{C}}, \tau, \Rightarrow, me, p, f)$ *be an FPOB schema and* $\boldsymbol{\mathcal{O}}$ *be a set of object identifiers. An* FPOB instance *over* $\boldsymbol{S}$ *is a pair* (π, ν) *where:*

(i) π *maps each* $c \in \boldsymbol{\mathcal{C}}$ *to a finite subset of* $\boldsymbol{\mathcal{O}}$ *such that, for different* $c_1, c_2 \in \boldsymbol{\mathcal{C}}$, $\pi(c_1) \cap \pi(c_2) = \emptyset$.
(ii) For each $c \in \boldsymbol{\mathcal{C}}$, ν *maps each* $o \in \pi(c)$ *to a fuzzy-probabilistic tuple value of type* $\tau(c)$.

In the definition above, $\pi(c)$ denotes only the set of the identifiers of the objects that are defined in the class c. Meanwhile, the set of the identifiers of all the objects that belong to c (i.e., those that are defined in c or its proper subclasses) are denoted by $\pi^*(c) = \cup\{\pi(d) \mid d \in \boldsymbol{\mathcal{C}} \text{ and } d \Rightarrow^* c\}$. Also, one writes $\pi(\boldsymbol{\mathcal{C}})$ to denote $\cup\{\pi(c) \mid c \in \boldsymbol{\mathcal{C}}\}$. We also note that, for ν, the default fuzzy-probabilistic tuple value returned by a method is determined by f.

Table 2. Object mappings π and π^*

c	$\pi(c)$	$\pi^*(c)$
PACKAGE	$\{o_1\}$	$\{o_1, o_2, o_3, o_4\}$
LETTER	{}	$\{o_2, o_3\}$
BOX	{}	$\{o_4\}$
TUBE	{}	{}
PRIORITY	{}	$\{o_2, o_3\}$
NORMAL	{}	$\{o_2\}$
PRIORITY_LETTER	$\{o_2, o_3\}$	$\{o_2, o_3\}$
NORMAL_BOX	$\{o_2\}$	$\{o_2\}$

Example 4. An FPOB instance over the FPOB schema in Example 3 can be (π, ν) where π and π^* are shown in Table 2 and ν in Table 3. Here, for computing the methods area and volume, the extension principle of fuzzy arithmetics and the independence probabilistic conjunction strategy are applied. For example, suppose that *about_20* = (18:0; 20:1; 22:.0), denoting the triangular shape fuzzy set whose three vertices are respectively $(18, 0), (20, 1)$ and $(22, 0)$, is the height of the box denoted by o_4. Then its volume is *about_25200* = (22680:0; 25200:1; 27720:0) or *about_26400* = (23760:0; 26400:1; 29040:0) with a probability between .4 and .7.

Table 3. Value mapping ν

oid	$v(oid)$
o_1	[origin: $\langle\{Hanoi\}, u, u\rangle$, destination: $\langle\{Saigon\}, u, u\rangle$, time: $\langle\{24,\ldots, 240\}, u, u\rangle$]
o_2	[origin: $\langle\{Saigon\}, u, u\rangle$, destination: $\langle\{Hue\}, u, u\rangle$, length: $\langle\{30, 32\}, .8u, 1.8u\rangle$, width: $\langle\{22\}, u, u\rangle$, priority_level: $\langle\{1\}, u, u\rangle$, time: $\langle\{24, 48\}, .9u, 1.2u\rangle$, area: $\langle\{660, 704\}, .8u, 1.8u\rangle$]
o_3	[origin: $\langle\{Hanoi\}, u, u\rangle$, destination: $\langle\{Nhatrang\}, u, u\rangle$, length: $\langle\{22, 24\}, u, u\rangle$, width: $\langle\{11, 12\}, .6u, 1.8u\rangle$, priority_level: $\langle\{2\}, u, u\rangle$, time: $\langle\{48, 72, 96\}, .6u, 1.8u\rangle$, area: $\langle\{242, 264, 288\}, \alpha, \beta\rangle$], where $\alpha(242) = \alpha(288) = .15$, $\alpha(264) = .3$, $\beta(242) = \beta(288) = .45$, $\beta(264) = .9$.
o_4	[origin: $\langle\{Saigon\}, u, u\rangle$, destination: $\langle\{Hanoi\}, u, u\rangle$, length: $\langle\{42, 44\}, .8u, 1.4u\rangle$, width: $\langle\{30\}, u, u\rangle$, height: $\langle\{about\_20\}, u, u\rangle$, stop_over: $\langle\{Nhatrang, Hue\}, .8u, 1.6u\rangle$, time: $\langle\{216, 240\}, .8u, 1.6u\rangle$, volume: $\langle\{about\_25200, about\_26400\}, .8u, 1.4u\rangle$]

4.2 Probabilistic Extents of Classes

In classical object bases, the extent of a class comprises all the objects that belong to that class. In POB as well as FPOB, the probabilistic extent of a class specifies the probability for each object belonging to that class. The following definition is adapted from [13].

Definition 10. *Let (π, ν) be an FPOB instance over an FPOB schema $\boldsymbol{S} = (\boldsymbol{C}, \tau, \Rightarrow, \mathit{me}, p, f)$. Then, for each class $c \in \boldsymbol{C}$, the probabilistic extent of c, denoted by ext(c), maps each $o \in \pi(\boldsymbol{C})$ to a set of rational number in [0, 1] as follows:*

(i) If $o \in \pi^*(c)$ then $ext(c)(o) = \{1\}$.
(ii) If $o \in \pi^*(d)$ and $\epsilon(c) \cap \epsilon(d) = \emptyset$ for every model ϵ of $\boldsymbol{S}$, then $ext(c)(o) = \{0\}$.
(iii) Otherwise, $ext(c)(o) = \{p \mid p$ is the product of the edge probabilities on a path from c up to d where $c \Rightarrow^* d$ with d being minimal and $o \in \pi^*(d)\}$.

For a comparison with relational databases, an FPOB schema corresponds to a relational schema, and each object of an FPOB instance corresponds to a tuple.

5 Selection Operation on Fuzzy and Probabilistic Object Bases

5.1 Syntax of Selection Conditions

As for relational databases and object bases, selection is a basic operation for FPOB. Intuitively, the result of a selection query on an FPOB instance $\boldsymbol{I}$ over an FPOB schema $\boldsymbol{S}$ is another FPOB instance $\boldsymbol{I}$' over $\boldsymbol{S}$ such that the objects of the classes in $\boldsymbol{I}$' and their property values satisfy the selection condition of the query.

Before defining the FPOB selection operation, we present the formal syntax and semantics of selection conditions, extending those definitions in [13] and [8] with class methods and relations on fuzzy set values. We start with the syntax of path expressions and selection expressions.

Definition 11. *Given a type* $\tau = [P_1 : \tau_1, P_2 : \tau_2, \ldots, P_k : \tau_k]$, path expressions *are inductively defined for every i from 1 to k as follows:*

(i) P_i *is a path expression for* τ.
(ii) If E_i *is a path expression for* τ_i, *then* $P_i.E_i$ *is a path expression for* τ.

For selection expressions on FPOBs, stating constraints on values of the last properties in path expressions, we employ the binary relations on fuzzy sets defined in Section 2.

Definition 12. *Let* $\boldsymbol{S} = (\boldsymbol{C}, \tau, \Rightarrow, me, p, f)$ *be an FPOB schema and* $\boldsymbol{\mathcal{X}}$ *be a set of object variables. Then* fuzzy selection expressions *are inductively defined as having one of the following forms:*

(i) $x \in c$, *where* $x \in \boldsymbol{\mathcal{X}}$ *and* $c \in \boldsymbol{C}$.
(ii) $x.P\ \theta\ v$, *where* $x \in \boldsymbol{\mathcal{X}}$, P *is a path expression,* θ *is a binary relation from* $\{=, \neq, \leq, <, \subseteq, \in, \rightarrow\}$, *and* v *is a value.*
(iii) $x.P_1 =_{\otimes} x.P_2$, *where* $x \in \boldsymbol{\mathcal{X}}$, P_1 *and* P_2 *are path expressions, and* $\otimes$ *is a probabilistic conjunction strategy of combining the probabilities for* $x.P_1 = v_1$ *and* $x.P_2 = v_2$ *such that* $v_1 = v_2$.
(iv) $E_1 \otimes E_2$, *where* E_1 *and* E_2 *are selection expressions over the same object variable and* $\otimes$ *is a probabilistic conjunction operator combining the probabilities for* E_1 *and* E_2 *being true.*
(v) $E_1 \oplus E_2$, *where* E_1 *and* E_2 *are selection expressions over the same object variable and* $\oplus$ *is a probabilistic disjunction operator combining the probabilities for* E_1 *and* E_2 *being true.*

Those of the first three forms are called atomic fuzzy selection expressions. *Different probabilistic conjunction and disjunction strategies mentioned in Section 2 could be used.*

Selection conditions are now defined as selection expressions to be satisfied with a probability in a given interval.

Definition 13. Fuzzy selection conditions *are inductively defined as follows:*

(i) If E *is a fuzzy selection expression and* $[l, u]$ *is a sub-interval of [0, 1], then* $(E)[l, u]$ *is a fuzzy selection condition.*
(ii) If ϕ *and* ψ *are fuzzy selection conditions, then* $\neg\phi$, $(\phi \wedge \psi)$, *and* $(\phi \vee \psi)$ *are fuzzy selection conditions.*

5.2 Semantics of Selection Conditions

For defining the semantics of selection conditions, interpretations of path expressions, fuzzy selection expressions and conditions are introduced.

Definition 14. *Given a type* $\tau = [P_1 : \tau_1, P_2 : \tau_2, \ldots, P_k : \tau_k]$ *and a value* $v = [P_1 : v_1, P_2 : v_2, \ldots, P_k : v_k]$*, the interpretation of a path expression* E *for* τ *under* v*, denoted by* $v.E$*, is inductively defined as follows:*

(i) If $E = P_i$*, then* $v.E = v_i$.
(ii) If $E = P_i.E_i$ *where* E_i *is a path expression for* τ_i*, then* $v.E = v_i.E_i$.

Definition 15. *Let* $\boldsymbol{S} = (\boldsymbol{\mathcal{C}}, \tau, \Rightarrow, \mathit{me}, p, f)$ *be an FPOB schema,* $\boldsymbol{I} = (\pi, \nu)$ *be an FPOB instance over* $\boldsymbol{S}$*, and* $o \in \pi(\boldsymbol{\mathcal{C}})$*. The probabilistic interpretation with respect to* $\boldsymbol{S}$*,* $\boldsymbol{I}$*, and* o*, denoted by* $prob_{\boldsymbol{S},\boldsymbol{I},o}$*, is the partial mapping from the set of all fuzzy selection expressions to the set of all closed sub-intervals of [0, 1] that is inductively defined as follows:*

(i) $prob_{\boldsymbol{S},\boldsymbol{I},o}(x \in c) = [min(ext(c)(o)),\ max(ext(c)(o))]$.
(ii) $prob_{\boldsymbol{S},\boldsymbol{I},o}(x.E\ \theta\ v) =$
$[\Sigma_{w \in V}\alpha(w).Pr(w.E'\ \theta\ v), min(1,\ \Sigma_{w \in V}\beta(w).Pr(w.E'\ \theta\ v))]$, *where* $E = P.E'$, $\nu(o).P = \langle V, \alpha, \beta \rangle$.
(iii) $prob_{\boldsymbol{S},\boldsymbol{I},o}(x.E_1 =_{\otimes} x.E_2) =$
$[\Sigma_{w \in V}\alpha(w).Pr(w_1.E_1'\,\theta\ w_2.E_2'), min(1, \Sigma_{w \in V}\beta(w).Pr(w_1.E_1'\,\theta\ w_2.E_2'))]$, *where* $E_1 = P_1.E_1'$, $\nu(o).P_1 = \langle V_1, \alpha_1, \beta_1 \rangle$, $E_2 = P_2.E_2'$, $\nu(o).P_2 = \langle V_2, \alpha_2, \beta_2 \rangle$, *and* $[\alpha(w), \beta(w)] = [\alpha_1(w_1), \beta_1(w_1)] \otimes [\alpha_2(w_2), \beta_2(w_2)]$ *for all* $w = (w_1, w_2) \in V_1 \times V_2$.
(iv) $prob_{\boldsymbol{S},\boldsymbol{I},o}(E_1 \otimes E_2) = prob_{\boldsymbol{S},\boldsymbol{I},o}(E_1) \otimes prob_{\boldsymbol{S},\boldsymbol{I},o}(E_2)$.
(v) $prob_{\boldsymbol{S},\boldsymbol{I},o}(E_1 \oplus E_2) = prob_{\boldsymbol{S},\boldsymbol{I},o}(E_1) \oplus prob_{\boldsymbol{S},\boldsymbol{I},o}(E_2)$.

Intuitively, $prob_{\boldsymbol{S},\boldsymbol{I},o}(x \in c)$ is the interval of the probability for o belonging to c, $prob_{\boldsymbol{S},\boldsymbol{I},o}(x.P.E'\ \theta\ v)$ is the interval of the probability for the property P of o having a value w such that $w.E'\ \theta\ v$. Also, $prob_{\boldsymbol{S},\boldsymbol{I},o}(x.P_1.E_1' =_{\otimes} x.P_2.E_2')$ is the interval of the probability for the properties P_1 and P_2 of o (whose mutual dependency is reflected in the selected $\otimes$) having values w_1 and w_2, respectively, such that $w_1.E_1' = w_2.E_2'$. We note that E', E_1', and E_2' can be empty. We now define when a selection condition is satisfied under a probabilistic interpretation.

Definition 16. *Let* $\boldsymbol{S} = (\boldsymbol{\mathcal{C}}, \tau, \Rightarrow, \mathit{me}, p, f)$ *be an FPOB schema,* $\boldsymbol{I} = (\pi, \nu)$ *be an FPOB instance over* $\boldsymbol{S}$*, and* $o \in \pi(\boldsymbol{\mathcal{C}})$*. The satisfaction of fuzzy selection conditions under* $prob_{\boldsymbol{S},\boldsymbol{I},o}$ *is defined as follows:*

(i) $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= (E)[l, u]$ *if and only if* $prob_{\boldsymbol{S},\boldsymbol{I},o}(E) \subseteq [l, u]$.
(ii) $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \neg\phi$ *if and only if* $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \phi$ *does not hold.*
(iii) $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= (\phi \wedge \psi)$ *if and only if* $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \phi$ *and* $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \psi$.
(iv) $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= (\phi \vee \psi)$ *if and only if* $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \phi$ *or* $prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \psi$.

Definition 17. *Let* $\boldsymbol{S} = (\boldsymbol{\mathcal{C}}, \tau, \Rightarrow, \mathit{me}, p, f)$ *be an FPOB schema,* $\boldsymbol{I} = (\pi, \nu)$ *be an FPOB instance over* $\boldsymbol{S}$*, and* ϕ *be a fuzzy selection condition over an object variable* x*. The selection on* $\boldsymbol{I}$ *with respect to* ϕ*, denoted by* $\sigma_\phi(\boldsymbol{I})$*, is the FPOB instance* $\boldsymbol{I}' = (\pi', \nu')$ *over* $\boldsymbol{S}$ *such that* $\pi'(c) = \{o \in \pi(c) \mid prob_{\boldsymbol{S},\boldsymbol{I},o} \mid\!= \phi\}$ *and* ν' *is* ν *restricted to* $\pi'(c)$.

Example 5. Let $\boldsymbol{I} = (\pi, \nu)$ be the instance in Example 4. The query "Find all the packages whose volume is *about 25000* with a probability of at least .3 and which can be sent within 10 days (240 hours) with a probability of at least .8" can be answered by the selection operation $\sigma_\phi(\boldsymbol{I})$ where $\phi = (x.\mathsf{volume} \rightarrow \mathit{about\_25000})[.3, 1] \wedge (x.\mathsf{time} \leq 240)[.8, 1]$. One can verify that $\sigma_\phi(\boldsymbol{I})$ only includes o_4, because $prob_{\boldsymbol{S},\boldsymbol{I},o_4}(x.\mathsf{time} \leq 240) = [.8, 1] \subseteq [.8, 1]$ and $prob_{\boldsymbol{S},\boldsymbol{I},o_4}(x.\mathsf{volume} \rightarrow \mathit{about\_25000}) = [.469, .821] \subseteq [.3, 1]$.

6 FPDB4O: A Fuzzy and Probabilistic Object Base Management System

6.1 Overview of FPDB4O

We have developed a management system for FPOB, called FPDB4O. It is built over DB4O, an open source database library written in Java. One can develops software using Java and employing DB4O for storing and retrieving objects in the classical object-oriented data model ([16]).

Small size, high performance, and simplicity in managing object persistence are advantages of DB4O. Therefore, it is getting popular in industry as well as research. For example, in [20], DB4O is used for storing reservation timetables of a local reservation system. Meanwhile, in [21], [19], and [27], the authors rely on DB4O for persistent layers in their systems.

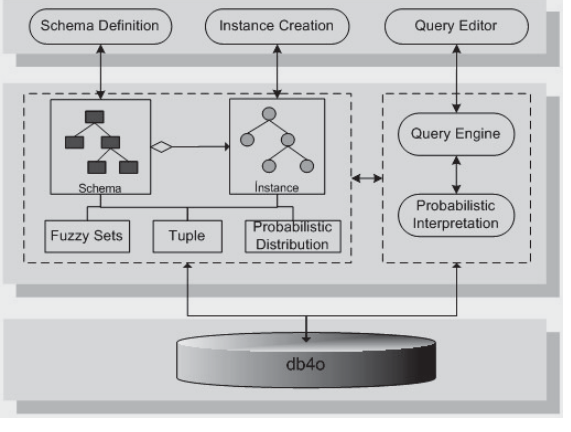


Fig. 2. FPDB4O architecture

Therefore, we have chosen DB4O as the underlying database management system. On top of it, we have defined classes to represent the probabilistic and fuzzy features of the FPOB model as presented above, and to express and execute FPOB queries. Figure 2 shows the architecture of FPDB4O comprising three layers:

(i) GUI Layer: this is for user interface, allowing users to draw an FPOB class hierarchy and define attributes for classes (Schema Definition), to create a new FPOB instance by creating and modifying data objects (Instance Creation), and to query on an FPOB (Query Editor).
(ii) Core Layer: this is the main component of the architecture. It includes two main blocks, one for representing the FPOB model, and one for executing queries. In more detail, the former determines the structure for representing schemas, data objects, probabilistic distributions of attribute values, and fuzzy sets. The latter parses queries, performs probabilistic interpretations, and gives answers.
(iii) Data Layer: this provides functions for storing and retrieving FPOB data objects, employing related functions of DB4O.

6.2 Implementation of FPOB Types and Schemas

Types in an FPOB schema, as defined above, include atomic types, fuzzy set types, and tuple types. Each type is implemented as a Java class. For atomic types, the system provides some available types in Java like **string**, **integer**, **double**, in addition to enumerated types defined by users. In case of an enumerated type, it is compiled into a corresponding class with its domain values, which then behaves in the same way as other built-in types like **string** and **integer**.

For fuzzy set types, we define two classes, namely, DiscreteFuzzySet for representing fuzzy sets on discrete domains and ContinuousFuzzySet for representing those on continuous domains. A discrete fuzzy set is defined by a set of domain values and their membership degrees. A continuous fuzzy set is assumed to be of the trapezoid shape, defined by four corner points. Fuzzy sets of both types are associated with linguistic labels for further reference.

A tuple type comprises some attributes and their types, which can be recursively tuple types as well. Therefore, we implement a tuple type in a tree structure, in which sub-nodes of a node function as sub-types of a tuple type. Names of classes representing types are created automatically and used internally in the system.

To define a schema, one needs to draw a class hierarchy and specify a finite set of attributes for each class. FPDB4O provides tools to draw a graph and edit attributes visually. Each class type has a table to store its attributes, each row of which contains the name and type of one attribute, and the list of its sub-types. Besides, for FPOB, a table is constructed to store the sub-class relationship and probability degrees between classes. Based on this table, the inherited attributes and probabilistic extent of an object to an FPOB class can be computed.

After an FPOB schema is completely defined and submitted, it is checked for syntactic and semantic correctness. The types, attributes and classes of a correct schema are then compiled into actual Java classes and saved for creating FPOB instances later. Figure 3 shows FPDB4O interface windows after a schema is created. The left panel is the probabilistic class hierarchy of the Package example above. Like most object-oriented systems, currently FPDB4O does not allow

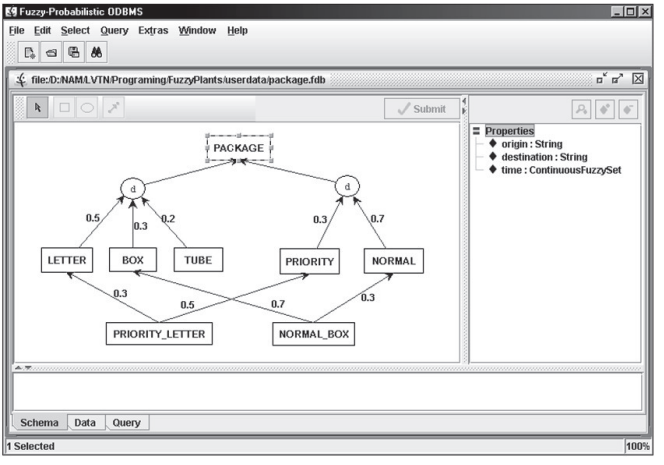


Fig. 3. Creating an FPOB schema in FPDB4O

multiple super-classes of a class to have a common property. When selecting a class in the left panel, one can see its attributes in the right panel in a tree structure. On this panel, users can edit, add or remove attributes of a class.

6.3 Implementation of FPOB Instances

An FPOB instance over an FPOB schema contains a set of objects instantiated from classes in the schema. Fuzzy-probabilistic triple values of object attributes are implemented by the generic class `Triple`$\langle$`type`$\rangle$, where `type` is the parameter representing the type of a particular attribute as defined in the schema. The class `Triple` is for representing fuzzy-probabilistic triple values of the form $\langle\{v_1, v_2, \ldots, v_k\}, \alpha, \beta\rangle$, by holding the list $\{v_1, \alpha(v_1), \beta(v_1)\}$,$\{v_2, \alpha(v_2), \beta(v_2)\}$,$\ldots$,$\{v_k, \alpha(v_k), \beta(v_k)\}$.

Figure 4 shows the interface of FPDB4O for creating an FPOB instance over a defined schema. The first panel displays all classes of the schema. The second lists all objects of a selected class. The third shows the attributes and their values of a selected object, displayed in a tree structure to be easily viewed and edited.

The system maintains a set of predefined fuzzy set values for each FPOB instance. Users can add, remove or edit any of those fuzzy sets. Each value of an attribute of a fuzzy set type has a reference to one of those predefined fuzzy set values. Removing a fuzzy set is permitted only if there is no reference to it.

When one wants to edit a tuple type value, the system traverses its whole value tree to get all the leaf nodes, which represent the values of the most inner attributes of that tuple value. Those values must be primitive, i.e., atomic or fuzzy set values. All of them are displayed on the same panel, and one can edit a tuple type value at one time.

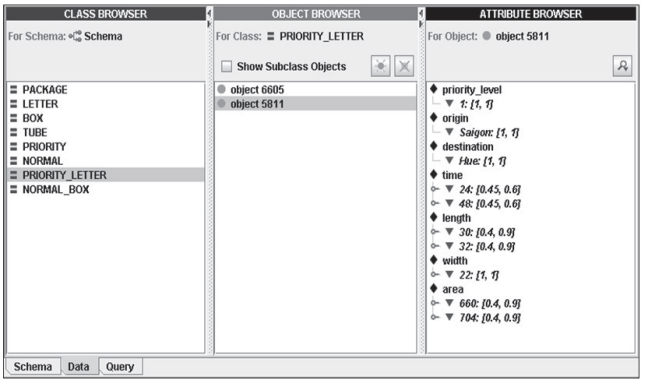


Fig. 4. Creating an FPOB instance in FPDB4O

6.4 Implementation of FPOB Selection Operation

FPDB4O supports the FPOB selection query as defined above, with a text editor for posing a query. For example, the query to select "Find all the packages whose volume is about 25000 with a probability of at least 0.3 and which can be sent within 10 days (240 hours) with a probability of at least 0.8" can be entered as follows:

$$(x.\mathsf{volume} \rightarrow \mathit{about\_25000})[.3,\ 1]\ \text{AND}\ (x.\mathsf{time} \leq 240)[.8,\ 1]$$

After parsing the fuzzy selection condition of a query, the system computes its probabilistic interpretation on each object in the FPOB instance of discourse, as in Definition 15. For FPOB, there are three atomic fuzzy selection expressions and two forms for expression combination, each of which is interpreted by a method as follows:

(i) The interpretation of $x \in c$ is computed based on the value of $ext(o)(c)$ for each object o. Those values are obtained from the component that manages the schema of discourse.
(ii) The interpretation of $x.P\ \theta\ v$ is computed by tracing the path expression P to get the actual attribute of an object, then comparing its fuzzy-probabilistic triple value to v in regards of θ.
(iii) The interpretation of $x.P_1 =_{\otimes} x.P_2$ is computed similarly. For the path expressions P_1 and P_2, every pair of corresponding values are compared and weighted by their probability distributions.
(iv) The interpretations of $E_1 \otimes E_2$ and $E_1 \oplus E_2$ are computed recursively using the probabilistic conjunction and disjunction strategies of discourse.

In the current version of FPDB4O, continuous fuzzy sets are transformed into discrete ones before fuzzy selection conditions are interpreted, with the overall computation accuracy less than 0.1%. Also, crisp attribute values are treated as special fuzzy sets in fuzzy relations.

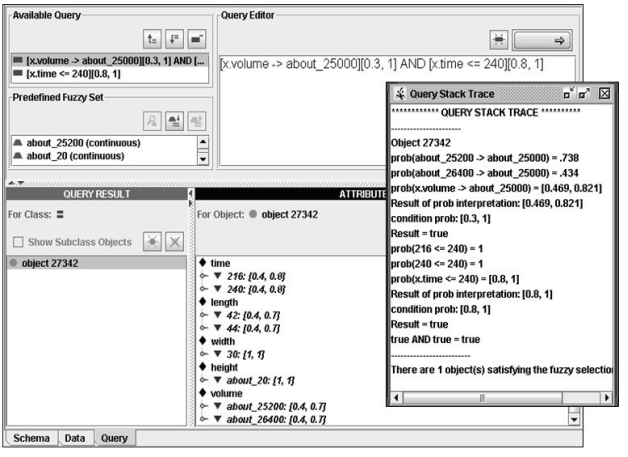


Fig. 5. Executing a query in FPDB4O

Figure 5 demonstrates execution of the query (x.volume $\rightarrow$ *about_25000*) [.3, 1] AND (x.time $\leq$ 240)[.8, 1], entered using the Query Editor. A query can be saved for later use. The Predefined Fuzzy Set panel displays all available fuzzy sets and allow users to edit them or add new ones. The result is shown on two bottom panels, with the left panel showing answer objects and the right panel showing their attribute values. When execution of a query is completed, the Query Stack Trace panel appears to show the probabilistic interpretations of the query with respect to each object. As shown in the figure, the object with internal ID 27342 is an answer to the query.

7 Conclusion

We have presented the object base model FPOB, as a fuzzy extension of the probabilistic model introduced by Eiter et al. In FPOB, uncertain and imprecise object property values are represented by probability distributions on fuzzy set values. Class methods are supported and realized as functions on fuzzy probabilistic triples.

A probabilistic interpretation of relations on fuzzy sets has been proposed as a basis to integrate probability and fuzzy set values into a unified model. The selection operation has been defined to query object bases, taking into account uncertain and imprecise values of class attributes and methods.

We have developed FPDB4O as an FPOB management system for real world applications, based on the open source object-oriented database platform DB4O. Implementation of FPOB types, schemas, instances, and selection operation have been presented.

On the one hand, FPDB4O is currently enhanced with computation on continuous fuzzy sets. On the other hand, the FPOB model is being extended further

with uncertain applicability of class properties and inheritance under uncertainty as discussed in [7]. Thus FPDB4O is to be advanced with those new features and the other operations, such as join and union, of the full FPOB algebra.

References

1. Baldwin, J.F., et al.: Toward soft computing object-oriented logic programming. In: Proceedings of the 9th IEEE International Conference on Fuzzy Systems, pp. 768–773 (2000)
2. Baldwin, J.F., Lawry, J.M., Martin, T.P.: A mass assignment theory of the probability of fuzzy events. International Journal of Fuzzy Sets and Systems 83, 353–367 (1996)
3. Baldwin, J.F., Lawry, J.M., Martin, T.P.: A note on probability/possibility consistency for fuzzy events. In: Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 521–525 (1996)
4. Berzal, F., et al.: A framework to build fuzzy object-oriented capabilities over an existing database system. In: Ma, Z. (ed.) Advances in fuzzy object-oriented database: modeling and applications, pp. 177–205. Idea Group Publishing, USA (2005)
5. Blanco, I., et al.: Softening the object-oriented database model: imprecision, uncertainty and fuzzy types. In: Proceedings of the 1st International Joint Conference of the International Fuzzy Systems Association and the North American Fuzzy Information Processing Society, pp. 2323–2328 (2001)
6. Bordogna, G., Pasi, G., Lucarella, D.: A fuzzy object-oriented data model managing vague and uncertain information. International Journal of Intelligent Systems 14, 623–651 (1999)
7. Cao, T.H.: Uncertain inheritance and recognition as probabilistic default reasoning. International Journal of Intelligent Systems 16, 781–803 (2001)
8. Cao, T.H., Nguyen, H.: Fuzzy and probabilistic object bases. In: Ma, Z. (ed.) Advances in fuzzy object-oriented databases: modelling and applications, pp. 46–84. Idea Group Publisher, USA (2005)
9. Cao, T.H., Rossiter, J.M.: A deductive probabilistic and fuzzy object-oriented database language. International Journal of Fuzzy Sets and Systems 140, 129–150 (2003)
10. Cross, V.V.: Defining fuzzy relationships in object models: Abstraction and interpretation. International Journal of Fuzzy Sets and Systems 140, 5–27 (2003)
11. De Tré, G., De Caluwe, R.: A constraint based fuzzy object-oriented database model. In: Ma, Z. (ed.) Advances in fuzzy object-oriented databases: Modelling and applications, pp. 1–45. Idea Group Publisher, USA (2005)
12. Dubitzky, W., et al.: Towards concept-oriented databases. Data & Knowledge Engineering 30, 23–55 (1999)
13. Eiter, T., et al.: Probabilistic object bases. ACM Transactions on Database Systems 26, 264–312 (2001)
14. Gaines, B.R.: Fuzzy and probability uncertainty logics. Journal of Information and Control 38, 154–169 (1978)
15. George, R., Buckles, B.P., Petry, F.E.: Modelling class hierarchies in the fuzzy object-oriented data model. International Journal of Fuzzy Sets and Systems 60, 259–272 (1993)

16. Grehan, R.: Complex object structures, persistence, and DB4O. Series of db4o whitepaper. db4objects Inc. (2005)
17. Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic - theory and applications. Prentice Hall PTR, Englewood Cliffs (1995)
18. Lakshmanan, L.V.S., et al.: ProbView: A flexible probabilistic database system. ACM Transactions on Database Systems 22, 419–469 (1997)
19. Mohamedally, D., et al.: MIKE's PET: A participant-based experiment tracking tool for HCI practitioners using mobile devices. In: Proceedings of SPIE's 18^{th} Annual Symposium for Electronic Imaging, pp. 216–224 (2006)
20. Nakada, H., et al.: Design and implementation of a local scheduling system with advance reservation for co-allocation on the grid. In: Proceedings of the 6^{th} IEEE International Conference on Computer and Information Technology, pp. 217–222 (2006)
21. Pfeifer, D.: Flexible object-oriented views using method propagation. In: Proceedings of the 8^{th} International Conference on Object-Oriented Information Systems, pp. 521–535 (2002)
22. Ross, R., Subrahmanian, V.S.: Aggregate Operators in Probabilistic Databases. Journal of the ACM 52, 54–101 (2005)
23. Rossazza, J.-P., Dubois, D., Prade, H.: A hierarchical model of fuzzy classes. In: De Caluwe, R. (ed.) Fuzzy and uncertain object-oriented databases: Concepts and models, pp. 21–61. World Scientific, Singapore (1997)
24. Van Gyseghem, N., De Caluwe, R.: The UFO database model: Dealing with imperfect information. In: De Caluwe, R. (ed.) Fuzzy and uncertain object-oriented databases: Concepts and models, pp. 123–185. World Scientific, Singapore (1997)
25. Yazici, A., George, R.: Fuzzy database modelling. Studies in Fuzziness and Soft Computing, vol. 26. Physica-Verlag, Heidelberg (1999)
26. Zadeh, L.A.: PRUF - A meaning representation language for natural languages. International Journal of Man-Machine Studies 10, 395–460 (1978)
27. Zhang, X., et al.: A usage-based authorization framework for collaborative computing system. In: Proceedings of ACM Symposium on Access Control Models and Technologies, pp. 180–189 (2006)

Part IV

Rough Sets and Belief Functions

Several Reducts in Dominance-Based Rough Set Approach

Masahiro Inuiguchi and Yukihiro Yoshioka

Department of Systems Innovation
Graduate School of Engineering Science, Osaka University
1-3, Machikaneyama, Toyonaka, Osaka 560-8531, Japan
inuiguti@sys.es.osaka-u.ac.jp, yoshioka@inulab.sys.es.osaka-u.ac.jp

Summary. In this paper, we investigate reducts preserving a structure induced from dominance-based rough sets as well as a structure from variable-precision dominance-based rough sets. It is shown that three kinds of reducts are obtained in the dominance-based rough set approach. For each kind of reduct, a discernibility matrix is given for the enumeration of the reducts. Moreover, it is shown that four kinds of reducts are obtained in the variable-precision dominance-based rough set approach. A counter-example showing the impossibility of reduct enumeration in the case of variable-precision dominance-based rough set approach.

1 Introduction

Attribute reduction in decision tables is one of major advantages of rough set analysis [1]. Important attributes are found by the attribute reduction. Attribute reduction is used in pattern recognition, data mining, decision analysis, and so on. The variety of attribute reduction is discussed by Ślęzak [2]. A reduced attribute set is called a *reduct*. Some reducts are defined by a measure while others are defined by structure induced from rough set analysis. In this paper, we focus on the structure-based reducts.

The well-known conventional reduct can be seen as a reduct preserving lower approximations. Since other reducts are defined by Ślęzak [2] and Inuiguchi and Tsurumi [3], we call the conventional reduct *reduct preserving lower approximations* or simply *L-reduct*. Ślęzak [2] proposed a reduct preserving boundary regions by the name of *rough decision reduct*. We call this reduct simply *B-reduct*. On the other hand, Inuiguchi and Tsurumi [3] proposed a reduct preserving upper approximations. We call it *U-reduct*. It is shown that a B-reduct is a U-reduct and vice versa [3]. Moreover, a U-reduct preserves lower approximations. Then, we obtain two kinds of reducts, L- and U-reducts based on the structure induced from rough sets.

This discussion is extended to the case of variable precision rough set models by Inuiguchi [4]. In this case, we can prove neither that a U-reduct preserves lower approximations nor that B- and U-reducts are equivalent. Moreover, we

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 163–175, 2008.
springerlink.com

may have objects which are not included in any upper approximations of decision classes. The set of such objects is called the *unpredictable region*. Because of these complexities, we can define seven kinds of structure-preserving reducts. The relations of those reducts are investigated by Inuiguchi [4].

In this paper, we would like to extend the investigation into dominance-based rough set approach (DRSA) [5, 6]. In DRSA, a reduct based on a measure called the quality of approximation has been proposed by Susmaga et al. [7]. They have not yet discussed the structure induced by DRSA. By the nature of DRSA, the lower and upper approximations are defined for upward and downward unions. Therefore, we can discuss structures based on those unions as well as based on decision classes. We treat the structure based on upward/downward unions in this paper. We define reducts related to the structure defined by unions and discuss the relationships among them in DRSA as well as variable-precision dominance-based rough set approach (VP-DRSA) [8]. Moreover, for each kind of reduct, we discuss the enumeration of reducts based on discernibility matrices. We show that this is impossible in VP-DRSA while possible in DRSA.

This paper is organized as follows. In next section, DRSA and VP-DRSA are briefly reviewed. In Section 3, reducts in DRSA are investigated. The relations among different kinds of reducts and discernibility matrices enumerating all reducts are given. In Section 4, reducts in VP-DRSA are studied. The relations among different kinds of reducts are shown and the impossibility of reduct enumeration is explained by a counter-example.

2 Dominance-Based Rough Set Approach

2.1 Decision Table with Dominance Relations

Consider a decision table $\mathcal{T} = \langle U, C \cup \{d\}, V, \rho \rangle$ shown in Table 1. A decision table $\mathcal{T}$ is characterized by an object set U, a condition attribute set C and a decision attribute d, an attribute value set $V = \bigcup_{a \in C \cup \{d\}} V_a$ (V_a is a set of all values of attribute a) and an information function $\rho : U \times C \cup \{d\} \to V$. In Table 1, we have $U = \{\mathrm{S1}, \mathrm{S2}, \ldots, \mathrm{S15}\}$, $C = \{\text{Mathematics } (Math), \text{Literature } (Lit)\}$, $d =$ Passing Status (PS) and $V = \{\text{Utterly Bad (UB)}, \text{Very Bad (VB)}, \text{Bad (B)}, \text{Medium (M)}, \text{Good (G)}, \text{Very Good (VG)}, \text{Excellent (E)}, \text{Yes (Y)}, \text{No (N)}\}$. The information function ρ is characterized by the table so that we know, for example, $\rho(\mathrm{S2}, Math) = \mathrm{E}$ and $\rho(\mathrm{S11}, PS) = \mathrm{N}$.

In cases such as Table 1, we assume that the better condition attribute values are, the better the decision value is. Namely, in Table 1, we assume a student having better evaluations in $Math$ and Lit, he/she can have a better value in PS. However, an inconsistency with this monotonicity is found in Table 1. For example, an inconsistency is found in evaluation between S4 and S9. S4 takes much better evaluations in $Math$ and Lit but a worse result in PS than S9. Such an inconsistency can occur (a) when decision maker have a hesitation in the evaluation, (b) when some related condition attribute is missing, (c) when the condition attribute data are substituted ones (e.g., trial examination scores

Table 1. A decision table of student evaluation

Student	Mathematics	Literature	Passing Status
S1	Excellent	Very Good	Yes
S2	Excellent	Medium	Yes
S3	Very Good	Excellent	Yes
S4	Very Good	Very Good	No
S5	Very Good	Good	Yes
S6	Very Good	Utterly Bad	No
S7	Good	Bad	Yes
S8	Medium	Very Good	Yes
S9	Medium	Bad	Yes
S10	Bad	Medium	No
S11	Bad	Very Bad	No
S12	Very Bad	Very Bad	No
S13	Very Bad	Utterly Bad	No
S14	Utterly Bad	Medium	No
S15	Utterly Bad	Bad	No
S16	Utterly Bar	Very Bad	No
S17	Utterly Bad	Utterly Bad	No

in this example) for those used for the evaluation of the decision attribute, (d) when data are recorded mistakenly, and so on. Such inconsistencies included in given decision tables lead to counter-intuitive results in the classical rough set analysis. To avoid such counter-intuitive results, the dominance-based rough set approach (DRSA) has been proposed by Greco et al. [5, 6]. In DRSA, we can treat nominal and ordinal condition attributes at the same time but in this paper, for the sake of simplicity, we consider a case that all condition attributes are ordinal. By this simplification, we do not loose the essence of the proposed approach.

2.2 DRSA

Let Cl_k, $k = 1, 2, \ldots, n$ be decision classes. Namely, to each decision attribute value v_{d_k}, we define $Cl_k = \{x \in U \mid \rho(x, d) = v_{d_k}\}$. We assume a total order for decision attribute values such that $v_{d_1} \prec v_{d_2} \prec \cdots \prec v_{d_n}$, where $v_{d_k} \prec v_{d_j}$ means that v_{d_j} is better than v_{d_k}. According to this total order we write $Cl_1 \prec Cl_2 \prec \cdots \prec Cl_n$. We also assume a dominance relation on condition attribute values. A dominance relation with respect to condition attribute p is denoted by $\succ_p$ and "$v_1 \succ_p v_2$" means that v_1 dominates (is better than) v_2. In Table 1, we have N $\prec$ Y for decision attribute and E $\succ_p$ VG $\succ_p$ G $\succ_p$ M $\succ_p$ B $\succ_p$ VB $\succ_p$ UB ($p = Math, Lit$) for condition attributes.

In order to reflect the total order and dominance relations, the following upward and downward unions of decision classes are considered:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \qquad Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s. \tag{1}$$

Then, we have

$$Cl_1^{\geq} = Cl_n^{\leq} = U, \qquad Cl_1^{\leq} = Cl_1, \qquad Cl_n^{\geq} = Cl_n, \tag{2}$$

$$Cl_t^{\geq} = U - Cl_{t-1}^{\leq}, \qquad Cl_t^{\leq} = U - Cl_{t+1}^{\geq}, \tag{3}$$

where we define $Cl_0^{\leq} = Cl_{n+1}^{\geq} = \emptyset$ so that the second equalities are valid for $t = 1, 2, \ldots, n$.

On the other hand, using dominance relations on condition attribute values, a dominance relation between objects with respect to a set of condition attributes $P \subseteq C$ is defined by

$$xD_Py \Leftrightarrow \rho(x,p) \succeq_p \rho(y,p) \text{ for all } p \in P, \tag{4}$$

where $v_1 \succeq_p v_2$ if and only if $v_1 \succ_p v_2$ or $v_1 = v_2$. Obviously, D_P is reflexive and transitive. Given $P \subseteq C$ and $x \in U$, we define

$$D_P^+(x) = \{y \in U \mid yD_px\}, \qquad D_P^-(x) = \{y \in U \mid xD_py\}. \tag{5}$$

Given $P \subseteq C$, for $t = 1, 2, \ldots, n$, P-lower and P-upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as follows:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^+(x) \subseteq Cl_t^{\geq}\}, \quad \overline{P}(Cl_t^{\geq}) = \bigcup\{D_p^+(x) \mid x \in Cl_t^{\geq}\}, \tag{6}$$

$$\underline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^-(x) \subseteq Cl_t^{\leq}\}, \quad \overline{P}(Cl_t^{\leq}) = \bigcup\{D_p^-(x) \mid x \in Cl_t^{\leq}\}. \tag{7}$$

Moreover, for $t = 1, 2 \ldots, n$, boundary regions $Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$ can be defined by

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \quad Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}). \tag{8}$$

Using those upper and lower approximations, decision tables with dominance relations can be analyzed in the same way as the classical rough set analysis. We have the following properties:

$$\underline{P}(Cl_1^{\geq}) = \overline{P}(Cl_1^{\geq}) = U, \quad \underline{P}(Cl_n^{\leq}) = \overline{P}(Cl_n^{\leq}) = U, \tag{9}$$

$$\underline{P}(Cl_{n+1}^{\geq}) = \overline{P}(Cl_{n+1}^{\geq}) = \emptyset, \quad \underline{P}(Cl_0^{\leq}) = \overline{P}(Cl_0^{\leq}) = \emptyset, \tag{10}$$

$$\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \quad \underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}), \tag{11}$$

$$\overline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\},$$

$$\overline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}, \tag{12}$$

$$\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq}), \quad \underline{P}(Cl_t^{\leq}) = U - \overline{P}(Cl_{t-1}^{\geq}), \tag{13}$$

$$Bn_P(Cl_t^{\geq}) = Bn_P(Cl_{t-1}^{\leq}) \text{ for } t = 1, \ldots, n. \tag{14}$$

Moreover, when $Q \subseteq P \subseteq C$, we have the following monotonicity:

$$\underline{Q}(Cl_t^{\geq}) \supseteq \underline{P}(Cl_t^{\geq}), \quad \underline{Q}(Cl_t^{\leq}) \supseteq \underline{P}(Cl_t^{\leq}), \tag{15}$$

$$\overline{Q}(Cl_t^{\geq}) \subseteq \overline{P}(Cl_t^{\geq}), \quad \overline{Q}(Cl_t^{\leq}) \subseteq \overline{P}(Cl_t^{\leq}). \tag{16}$$

2.3 VP-DRSA

The given decision table is sometimes inconsistent with the monotonicity. Such an inconsistency is caused by errors in recording, measurement, observation, and so on. From this point of view, the authors proposed variable-precision dominance-based rough set approach. In the approach, we define the precision of $x \in Cl_t^{\geq}$ by

$$\beta = \frac{|D_P^-(x) \cap Cl_t^{\geq}|}{|D_P^-(x) \cap Cl_t^{\geq}| + |D_P^+(x) \cap Cl_{t-1}^{\leq}|}. \tag{17}$$

The precision β can be interpreted as follows. For any $y \in D_P^-(x)$, from the dominance relation D_P, we may infer that $\rho(x,d) \succeq_d \rho(y,d)$, i.e., x is included in a decision class not worse than the decision class to which y belongs. Thus for any $y \in D_P^-(x) \cap Cl_t^{\geq}$, we may infer $x \in Cl_t^{\geq}$. Hence, $|D_P^-(x) \cap Cl_t^{\geq}|$ is the number of objects which endorses $x \in Cl_t^{\geq}$. On the contrary, by the same consideration, for any $z \in D_P^+(x) \cap Cl_{t-1}^{\leq}$, we may infer $x \in Cl_{t-1}^{\leq} = U - Cl_t^{\geq}$. Hence $|D_P^+(x) \cap Cl_{t-1}^{\leq}|$ is the number of objects which endorses $x \notin Cl_t^{\geq}$. Other objects endorse neither $x \in Cl_t^{\geq}$ nor $x \notin Cl_t^{\geq}$. Therefore, β is the ratio of objects endorsing $x \in Cl_t^{\geq}$ to all objects endorsing $x \in Cl_t^{\geq}$ or $x \notin Cl_t^{\geq}$.

Then, given a precision level $l \in (0.5, 1]$, a P-positive region of $Cl_t^{\geq}$ with respect to $P \subseteq C$ is defined as a set of objects $x \in U$ whose degrees of precision are not less than l, i.e.,

$$\mathrm{POS}_P^l(Cl_t^{\geq}) = \left\{ x \in U \;\middle|\; \frac{|D_P^-(x) \cap Cl_t^{\geq}|}{|D_P^-(x) \cap Cl_t^{\geq}| + |D_P^+(x) \cap Cl_{t-1}^{\leq}|} \geq l \right\}. \tag{18}$$

In variable-precision case, we use a terminology 'positive region' instead of 'lower approximation' following the convention. Accordingly, we use a terminology 'non-negative region' instead of 'upper approximation'.

Similarly, a P-positive region of $Cl_t^{\leq}$ with respect to $P \subseteq C$ is defined by

$$\mathrm{POS}_P^l(Cl_t^{\leq}) = \left\{ x \in U \;\middle|\; \frac{|D_P^+(x) \cap Cl_t^{\leq}|}{|D_P^+(x) \cap Cl_t^{\leq}| + |D_P^-(x) \cap Cl_{t+1}^{\geq}|} \geq l \right\}. \tag{19}$$

By using the duality, P-nonnegative regions of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ with respect to $P \subseteq C$ can be defined by

$$\begin{aligned} \mathrm{NNG}_P^l(Cl_t^{\geq}) &= U - \mathrm{POS}_P^l(U - Cl_t^{\geq}) = U - \mathrm{POS}_P^l(Cl_{t-1}^{\leq}) \\ &= \left\{ x \in U \;\middle|\; \frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x) \cap Cl_t^{\geq}| + |D_P^+(x) \cap Cl_{t-1}^{\leq}|} > 1 - l \right\}, \end{aligned} \tag{20}$$

$$\begin{aligned} \mathrm{NNG}_P^l(Cl_t^{\leq}) &= U - \mathrm{POS}_P^l(U - Cl_t^{\leq}) = U - \mathrm{POS}_P^l(Cl_{t+1}^{\geq}) \\ &= \left\{ x \in U \;\middle|\; \frac{|D_P^+(x) \cap Cl_t^{\leq}|}{|D_P^+(x) \cap Cl_t^{\leq}| + |D_P^-(x) \cap Cl_{t+1}^{\leq}|} > 1 - l \right\}. \end{aligned} \tag{21}$$

We can define P-boundary regions of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ with respect to $P \subseteq C$ as follows:

$$\mathrm{BND}_P^l(Cl_t^{\geq}) = \mathrm{NNG}_P^l(Cl_t^{\geq}) - \mathrm{POS}_P^l(Cl_t^{\geq}), \tag{22}$$
$$\mathrm{BND}_P^l(Cl_t^{\leq}) = \mathrm{NNG}_P^l(Cl_t^{\leq}) - \mathrm{POS}_P^l(Cl_t^{\leq}). \tag{23}$$

For positive and nonnegative regions defined above satisfy the following relations:

$$\mathrm{POS}_P^l(Cl_1^{\geq}) = \mathrm{NNG}_P^l(Cl_1^{\geq}) = U,\ \mathrm{POS}_P^l(Cl_n^{\leq}) = \mathrm{NNG}_P^l(Cl_n^{\leq}) = U, \tag{24}$$
$$\mathrm{POS}_P^l(Cl_{n+1}^{\geq}) = \mathrm{NNG}_P^l(Cl_{n+1}^{\geq}) = \emptyset,\ \mathrm{POS}_P^l(Cl_0^{\leq}) = \mathrm{NNG}_P^l(Cl_0^{\leq}) = \emptyset, \tag{25}$$
$$\mathrm{POS}_P^l(Cl_t^{\geq}) \subseteq \mathrm{NNG}_P^l(Cl_t^{\geq}), \quad \mathrm{POS}_P^l(Cl_t^{\leq}) \subseteq \mathrm{NNG}_P^l(Cl_t^{\leq}), \tag{26}$$
$$\mathrm{BND}_P^l(Cl_t^{\geq}) = \mathrm{BND}_P^l(Cl_{t-1}^{\leq}) \text{ for } t = 1, \ldots, n. \tag{27}$$

In the case of VP-DRSA, the monotonicity expressed by (15) and (16) cannot be always satisfied. Moreover, by virtue of (24), there is no unpredictable region.

3 Union-Based Reducts in DRSA

Lower and upper approximations and boundary regions of upward and downward unions can be considered as a structure over a given object set U. From this point, we define several structure-preserving reducts. The following reducts are conceivable:

Reduct preserving lower approximations of upward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving lower approximations of upward unions' or simply an '$\mathrm{L}^{\geq}$-reduct' if and only if
($\mathrm{L1}^{\geq}$) $\underline{P}(Cl_t^{\geq}) = \underline{C}(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$, and
($\mathrm{L2}^{\geq}$) $\nexists Q \subseteq C$, $\underline{Q}(Cl_t^{\geq}) = \underline{C}(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$.

Reduct preserving lower approximations of downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving lower approximations of downward unions' or simply an '$\mathrm{L}^{\leq}$-reduct' if and only if
($\mathrm{L1}^{\leq}$) $\underline{P}(Cl_t^{\leq}) = \underline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, and
($\mathrm{L2}^{\leq}$) $\nexists Q \subseteq C$, $\underline{Q}(Cl_t^{\leq}) = \underline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$.

Reduct preserving upper approximations of upward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving upper approximations of upward unions' or simply an '$\mathrm{U}^{\geq}$-reduct' if and only if
($\mathrm{U1}^{\geq}$) $\overline{P}(Cl_t^{\geq}) = \overline{C}(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$, and
($\mathrm{U2}^{\geq}$) $\nexists Q \subseteq C$, $\overline{Q}(Cl_t^{\geq}) = \overline{C}(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$.

Reduct preserving upper approximations of downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving upper approximations of downward unions' or simply an '$\mathrm{U}^{\leq}$-reduct' if and only if
($\mathrm{U1}^{\leq}$) $\overline{P}(Cl_t^{\leq}) = \overline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, and
($\mathrm{U2}^{\leq}$) $\nexists Q \subseteq C$, $\overline{Q}(Cl_t^{\leq}) = \overline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$.

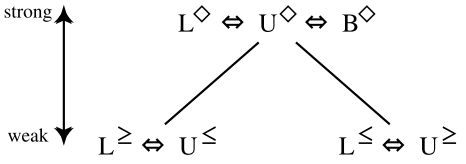


Fig. 1. Strong-weak relation of union-based reducts in DRSA

Reduct preserving lower approximations of upward and downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving lower approximations of upward and downward unions' or simply an '$L^{\Diamond}$-reduct' if and only if

(L1$^{\Diamond}$) $\underline{P}(Cl_t^{\geq}) = \underline{C}(Cl_t^{\geq})$, $\underline{P}(Cl_t^{\leq}) = \underline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, and

(L2$^{\Diamond}$) $\nexists Q \subseteq C$, $\underline{Q}(Cl_t^{\leq}) = \underline{C}(Cl_t^{\leq})$, $\underline{Q}(Cl_t^{\leq}) = \underline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$.

Reduct preserving upper approximations of upward and downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving upper approximations of upward and downward unions' or simply an '$U^{\Diamond}$-reduct' if and only if

(U1$^{\Diamond}$) $\overline{P}(Cl_t^{\leq}) = \overline{C}(Cl_t^{\leq})$, $\overline{P}(Cl_t^{\leq}) = \overline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, and

(U2$^{\Diamond}$) $\nexists Q \subseteq C$, $\overline{Q}(Cl_t^{\leq}) = \overline{C}(Cl_t^{\leq})$, $\overline{Q}(Cl_t^{\leq}) = \overline{C}(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$.

Reduct preserving boundary regions of upward/downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving boundary regions of upward/downward unions' or simply an '$B^{\Diamond}$-reduct' if and only if

(B1$^{\Diamond}$) $Bn_P(Cl_t^{\leq}) = Bn_C(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, or equivalently, $Bn_P(Cl_t^{\geq}) = Bn_C(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$, and

(B2$^{\Diamond}$) $\nexists Q \subseteq C$, $Bn_Q(Cl_t^{\leq}) = Bn_C(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, or equivalently, $\nexists Q \subseteq C$, $Bn_Q(Cl_t^{\geq}) = Bn_C(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$.

From (9) to (14), we know that $L^{\geq}$-reduct and $U^{\leq}$-reduct are equivalent. Similarly, $L^{\leq}$-reduct and $U^{\geq}$-reduct are equivalent. Therefore, $L^{\Diamond}$-reduct is equivalent to $U^{\Diamond}$-reduct. Moreover, since condition (L1$^{\Diamond}$) implies conditions (L1$^{\geq}$) and (L1$^{\leq}$), any $L^{\Diamond}$-reduct is a superset of an $L^{\geq}$-reduct and also a superset of an $L^{\leq}$-reduct. Similarly, since condition (U1$^{\Diamond}$) implies conditions (U1$^{\geq}$) and (U1$^{\leq}$), any $U^{\Diamond}$-reduct is a superset of a $U^{\geq}$-reduct and also a superset of a $U^{\leq}$-reduct.

We have the following theorem.

Theorem 1. *$B^{\Diamond}$-reduct is equivalent to $L^{\Diamond}$-reduct, or equivalently, $U^{\Diamond}$-reduct.*

Proof. By (8) and (13), it is straightforwardly obtained that (L1$^{\Diamond}$) implies (B1$^{\Diamond}$). Then it suffices to prove that (B1$^{\Diamond}$) implies (L1$^{\Diamond}$). Suppose (B1$^{\Diamond}$) holds but (L1$^{\Diamond}$) does not. Under this supposition, two cases are possible: (a) (L1$^{\geq}$) does not hold and (b) (L1$^{\leq}$) does not hold. Consider case (a). In this case, by the monotonicity, there exists at least one object such that $x \in \underline{C}(Cl_t^{\geq})$ but $x \notin \underline{P}(Cl_t^{\geq})$. To satisfy (B1$^{\Diamond}$), we should have $x \notin \overline{P}(Cl_t^{\geq})$. However, we have $x \in \underline{C}(Cl_t^{\geq}) \subseteq \overline{C}(Cl_t^{\geq}) \subseteq \overline{P}(Cl_t^{\geq})$. This fact contradicts with the previous fact. Then we never have case

Table 2. An example showing the difference $\mathrm{L}^{\diamond}$ and $\mathrm{L}^{\geq}$-reducts

Student	Mathematics	Literature	Passing Status
S1	Very Good	Good	Yes
S2	Good	Very Good	Yes
S3	Medium	Medium	Yes
S4	Good	Bad	No
S5	Bad	Very Bad	Yes
S6	Very Bad	Utterly Bad	No

(a). In case (b), in the same way, we obtain the contradiction. Therefore, the supposition is never valid. Then we have that (B1$^{\diamond}$) implies (L1$^{\diamond}$). □

As the result of the discussion, we obtain three different reducts based on the structure induced from rough set operations on unions. They are represented by $\mathrm{L}^{\geq}$-reduct, $\mathrm{L}^{\leq}$-reduct and $\mathrm{L}^{\diamond}$-reduct. The strong-weak relations among those reducts are depicted in Figure 1. Moreover, a simple example showing the difference between $\mathrm{L}^{\geq}$- and $\mathrm{L}^{\diamond}$-reducts is given by a decision table shown in Table 2. In Table 2, we have $C = \{Math, Lit\}$ and consider 'Yes' $\succ$ 'No'. Let $P = \{Lit\}$. We have $\underline{C}(Cl^{\geq}_{\mathrm{Yes}}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}\}$, $\overline{C}(Cl^{\geq}_{\mathrm{Yes}}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}\}$, $\underline{P}(Cl^{\geq}_{\mathrm{Yes}}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}\}$ and $\overline{P}(Cl^{\geq}_{\mathrm{Yes}}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}\}$. Then P can be an $\mathrm{L}^{\geq}$-reduct but not an $\mathrm{L}^{\diamond}$-reduct.

3.1 Discernibility Matrices for Reducts

The usual reduct defined in the classical rough set model can be calculated by using a decision matrix. In this subsection, we present discernibility matrices that can be used to calculate the union-based reducts.

Consider two objects in U. If one object belongs to a lower approximation of a union and the other object does not, we should take an attribute which can explain this difference. From the opposite viewpoint, if no such an attribute is included in $P \subseteq C$, the difference of the two objects cannot be expressed. Moreover, when any two objects in a lower approximation of a union become indiscernible by the reduction of condition attribute, their memberships to the lower approximation are preserved.

From these observations, the (i, j, k)-component of the discernibility matrix for enumerating $\mathrm{L}^{\geq}$-reducts is defined by

$$\delta^{\geq}_{ijk} = \begin{cases} \{a \in C \mid \rho(x_i, a) \succ \rho(x_j, a)\}, & \text{if } x_i \in \underline{C}(Cl^{\geq}_k),\ x_j \notin \underline{C}(Cl^{\geq}_k), \\ *, & \text{otherwise}, \end{cases} \tag{28}$$

where $*$ stands for 'don't care'. The preservation of lower approximations of upward unions is characterized by the following logical function:

$$F^{\mathrm{L}^{\geq}} = \bigwedge_{k=1,2,\ldots,n} \bigwedge_{\substack{i,j \\ x_i, x_j \in U}} \bigvee \delta^{\geq}_{ijk}. \tag{29}$$

Thus, all $\mathrm{L}^{\geq}$-reducts can be obtained as prime implicants of $F^{\mathrm{L}^{\geq}}$.

Similarly, the (i,j,k)-component of the discernibility matrix for enumerating $\mathrm{L}^{\leq}$-reducts is defined by

$$\delta^{\leq}_{ijk} = \begin{cases} \{a \in C \mid \rho(x_i,a) \prec \rho(x_j,a)\}, & \text{if } x_i \in \underline{C}(Cl^{\leq}_k),\ x_j \notin \underline{C}(Cl^{\leq}_k), \\ *, & \text{otherwise.} \end{cases} \tag{30}$$

The preservation of lower approximations of upward unions is characterized by the following logical function:

$$F^{\mathrm{L}^{\leq}} = \bigwedge_{k=1,2,\ldots,n} \bigwedge_{\substack{i,j \\ x_i,x_j \in U}} \bigvee \delta^{\leq}_{ijk}. \tag{31}$$

Thus, all $\mathrm{L}^{\leq}$-reducts can be obtained as prime implicants of $F^{\mathrm{L}^{\leq}}$.

Finally, because we have $(\mathrm{L}^{\diamond}) \Leftrightarrow (\mathrm{L}^{\geq}) \wedge (\mathrm{L}^{\leq})$, all $\mathrm{L}^{\diamond}$-reducts can be obtained as prime implicants of $F^{\mathrm{L}^{\geq}} \wedge F^{\mathrm{L}^{\leq}}$.

4 Union-Based Reducts in VP-DRSA

Now let us discuss reducts in VP-DRSA. The seven reducts corresponding to $\mathrm{L}^{\geq}$-, $\mathrm{L}^{\leq}$-, $\mathrm{U}^{\geq}$-, $\mathrm{U}^{\leq}$-, $\mathrm{L}^{\diamond}$-, $\mathrm{U}^{\diamond}$- and $\mathrm{B}^{\diamond}$-reducts in DRSA are conceivable. However, considering the relations (20) and (21), we select the following four reducts:

Reduct preserving positive regions of upward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving positive regions of upward unions' or simply an '$\mathrm{L}^{\geq}_l$-reduct' if and only if

($\mathrm{L}1^{\geq}_l$) $\mathrm{POS}^l_P(Cl^{\geq}_t) = \mathrm{POS}^l_C(Cl^{\geq}_t)$, $t = 1,2,\ldots,n$, and

($\mathrm{L}2^{\geq}_l$) $\nexists Q \subseteq C$, $\mathrm{POS}^l_Q(Cl^{\geq}_t) = \mathrm{POS}^l_C(Cl^{\geq}_t)$, $t = 1,2,\ldots,n$.

Reduct preserving positive regions of downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving positive regions of downward unions' or simply an '$\mathrm{L}^{\leq}_l$-reduct' if and only if

($\mathrm{L}1^{\leq}_l$) $\mathrm{POS}^l_P(Cl^{\leq}_t) = \mathrm{POS}^l_C(Cl^{\leq}_t)$, $t = 1,2,\ldots,n$, and

($\mathrm{L}2^{\leq}_l$) $\nexists Q \subseteq C$, $\mathrm{POS}^l_Q(Cl^{\leq}_t) = \mathrm{POS}^l_C(Cl^{\leq}_t)$, $t = 1,2,\ldots,n$.

Reduct preserving positive regions of upward and downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving positive regions of upward and downward unions' or simply an '$\mathrm{L}^{\diamond}_l$-reduct' if and only if

($\mathrm{L}1^{\diamond}_l$) $\mathrm{POS}^l_P(Cl^{\geq}_t) = \mathrm{POS}^l_C(Cl^{\geq}_t)$, $\mathrm{POS}^l_P(Cl^{\leq}_t) = \mathrm{POS}^l_C(Cl^{\leq}_t)$, $t = 1,2,\ldots,n$, and

($\mathrm{L}2^{\diamond}_l$) $\nexists Q \subseteq C$, $\mathrm{POS}^l_Q(Cl^{\leq}_t) = \mathrm{POS}^l_C(Cl^{\leq}_t)$, $\mathrm{POS}^l_Q(Cl^{\leq}_t) = \mathrm{POS}^l_C(Cl^{\leq}_t)$, $t = 1,2,\ldots,n$.

Reduct preserving boundary regions of upward/downward unions: A set of condition attributes, $P \subseteq C$ is called a 'reduct preserving boundary regions of upward/downward unions' or simply an '$\mathrm{B}^{\diamond}_l$-reduct' if and only if

(B1$_l^\diamond$) $\mathrm{BND}_P^l(Cl_t^{\leq}) = \mathrm{BND}_C^l(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, or equivalently, $\mathrm{BND}_P^l(Cl_t^{\geq}) = \mathrm{BND}_C^l(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$, and

(B2$^\diamond$) $\nexists Q \subseteq C$, $\mathrm{BND}_Q^l(Cl_t^{\leq}) = \mathrm{BND}_C^l(Cl_t^{\leq})$, $t = 1, 2, \ldots, n$, or equivalently, $\nexists Q \subseteq C$, $\mathrm{BND}_Q^l(Cl_t^{\geq}) = \mathrm{BND}_C^l(Cl_t^{\geq})$, $t = 1, 2, \ldots, n$.

Note that other three reducts are included in those four reducts. More concretely, $\mathrm{U}_l^{\geq}$-reduct defined corresponding to $\mathrm{U}^{\geq}$-reduct is equivalent to $\mathrm{L}_l^{\leq}$-reduct and $\mathrm{U}_l^{\leq}$-reduct defined corresponding to $\mathrm{U}^{\leq}$-reduct is equivalent to $\mathrm{L}_l^{\geq}$-reduct. Moreover, $\mathrm{U}_l^{\diamond}$-reduct defined corresponding to $\mathrm{U}^{\diamond}$-reduct is equivalent to $\mathrm{L}_l^{\diamond}$-reduct.

Obviously, we have that condition (L1$_l^\diamond$) implies conditions (L1$_l^{\geq}$) and (L1$_l^{\leq}$). Therefore, any $\mathrm{L}_l^{\diamond}$-reduct is a superset of an $\mathrm{L}_l^{\geq}$-reduct and a superset of an $\mathrm{L}_l^{\leq}$-reduct. Moreover, conditions (B1$_l^\diamond$) and (L1$_l^{\geq}$) imply condition (L1$_l^\diamond$) and conditions (B1$_l^\diamond$) and (L1$_l^{\leq}$) imply condition (L1$_l^\diamond$). Therefore, the reduct based on (B1$_l^\diamond$) and (L1$_l^{\geq}$) and that based on (B1$_l^\diamond$) and (L1$_l^{\leq}$) are both reduced to an $\mathrm{L}_l^{\diamond}$-reduct.

In DRSA, Theorem 1 holds. However in VP-DRSA, the theorem corresponding to Theorem 1 does not hold. We only have that (L1$_l^\diamond$) implies (B1$_l^\diamond$). This implies that any $\mathrm{L}_l^{\diamond}$-reduct is a superset of an $\mathrm{B}_l^{\diamond}$-reduct. The difference between $\mathrm{L}_l^{\diamond}$-reduct and $\mathrm{B}_l^{\diamond}$-reduct can be shown by a simple example with a decision table given in Table 1. Let $P = \{Math\}$ and $l = 0.7$. We obtain

$$\mathrm{POS}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}\}, \tag{32}$$

$$\mathrm{POS}_P^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}\}, \tag{33}$$

$$\mathrm{NNG}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S7}, \mathrm{S8}, \mathrm{S9}\}, \tag{34}$$

$$\mathrm{NNG}_P^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}, \mathrm{S7}, \mathrm{S8}, \mathrm{S9}\}, \tag{35}$$

$$\mathrm{BND}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S7}, \mathrm{S8}, \mathrm{S9}\}, \tag{36}$$

$$\mathrm{BND}_P^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S7}, \mathrm{S8}, \mathrm{S9}\}, \tag{37}$$

$$\mathrm{POS}_C^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S6}, \mathrm{S10}, \mathrm{S11}, \mathrm{S12}, \mathrm{S13}, \mathrm{S14}, \mathrm{S15}, \mathrm{S16}, \mathrm{S17}\}, \tag{38}$$

$$\mathrm{POS}_P^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S10}, \mathrm{S11}, \mathrm{S12}, \mathrm{S13}, \mathrm{S14}, \mathrm{S15}, \mathrm{S16}, \mathrm{S17}\}, \tag{39}$$

$$\mathrm{NNG}_C^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S6}, \mathrm{S7}, \mathrm{S8}, \mathrm{S9}, \mathrm{S10}, \mathrm{S11}, \mathrm{S12}, \mathrm{S13}, \mathrm{S14}, \mathrm{S15}, \mathrm{S16}, \mathrm{S17}\}, \tag{40}$$

$$\mathrm{NNG}_P^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S7}, \mathrm{S8}, \mathrm{S9}, \mathrm{S10}, \mathrm{S11}, \mathrm{S12}, \mathrm{S13}, \mathrm{S14}, \mathrm{S15}, \mathrm{S16}, \mathrm{S17}\}, \tag{41}$$

$$\mathrm{BND}_C^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S7}, \mathrm{S8}, \mathrm{S9}\}, \tag{42}$$

$$\mathrm{BND}_P^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S7}, \mathrm{S8}, \mathrm{S9}\}. \tag{43}$$

Thus, $\mathrm{BND}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \mathrm{BND}_P^l(Cl_{\mathrm{Yes}}^{\geq})$ and $\mathrm{BND}_C^l(Cl_{\mathrm{No}}^{\leq}) = \mathrm{BND}_P^l(Cl_{\mathrm{No}}^{\leq})$ are valid. Then P is a $\mathrm{B}_l^{\diamond}$-reduct. However, we have $\mathrm{POS}_C^l(Cl_{\mathrm{Yes}}^{\geq}) \neq \mathrm{POS}_P^l(Cl_{\mathrm{Yes}}^{\geq})$, $\mathrm{NNG}_C^l(Cl_{\mathrm{Yes}}^{\geq}) \neq \mathrm{NNG}_P^l(Cl_{\mathrm{Yes}}^{\geq})$, $\mathrm{POS}_C^l(Cl_{\mathrm{No}}^{\leq}) \neq \mathrm{POS}_P^l(Cl_{\mathrm{No}}^{\leq})$ and $\mathrm{NNG}_C^l(Cl_{\mathrm{No}}^{\leq}) \neq \mathrm{NNG}_P^l(Cl_{\mathrm{No}}^{\leq})$. Therefore, P is not an $\mathrm{L}_l^{\diamond}$-reduct.

The strong-weak relations among those reducts are depicted in Figure 2. Even in VP-DRSA, the relation of union-based reducts becomes simple.

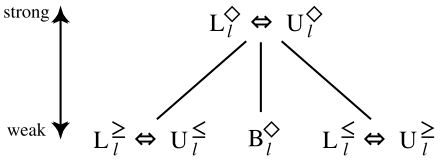


Fig. 2. Strong-weak relation of union-based reducts in VP-DRSA

Table 3. A counter-example

Student	Mathematics	Physics	Literature	Passing Status
S1	Good	Very Good	Very Good	Yes
S2	Very Good	Very Good	Good	Yes
S3	Good	Good	Good	Yes
S4	Bad	Bad	Bad	Yes
S5	Good	Very Good	Bad	No
S6	Good	Medium	Medium	No
S7	Medium	Good	Bad	Yes
S8	Medium	Medium	Medium	Yes
S9	Bad	Very Bad	Bad	No
S10	Very Bad	Bad	Bad	No
S11	Very Bad	Very Bad	Bad	No

Being different from the classical variable precision rough set model, no union-based reduct in VP-DRSA can be enumerated by a discernibility matrix. This is because their memberships to the positive region (resp. boundary region) are not always preserved when an attribute is dropped keeping the discernibility between positive and non-positive regions of unions. A counter-example is given as follows.

Consider a decision table given in Table 3. Let $l = 0.65$. Then we obtain

$$\mathrm{POS}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}, \mathrm{S7}, \mathrm{S8}\}, \tag{44}$$

$$\mathrm{NNG}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}, \mathrm{S7}, \mathrm{S8}\}, \tag{45}$$

$$\mathrm{BND}_C^l(Cl_{\mathrm{Yes}}^{\geq}) = \emptyset, \tag{46}$$

$$\mathrm{POS}_C^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S9}, \mathrm{S10}, \mathrm{S11}\}, \tag{47}$$

$$\mathrm{NNG}_C^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S9}, \mathrm{S10}, \mathrm{S11}\}, \tag{48}$$

$$\mathrm{BND}_C^l(Cl_{\mathrm{No}}^{\leq}) = \emptyset. \tag{49}$$

In order to keep the difference between $\mathrm{POS}_C^l(Cl_{\mathrm{Yes}}^{\geq})$ and $\mathrm{NNG}_C^l(Cl_{\mathrm{No}}^{\leq})$, we cannot drop $Math$ from C since dropping $Math$ makes S4 and S10 indiscernible in condition attributes. Similarly, we cannot drop Phy from C since dropping Phy makes S4 and S9 indiscernible. Then $P = \{Math, Phy\}$ can be a candidate of $\mathrm{L}_l^{\geq}$-reduct (also a candidate of any other reduct). However, we have

$$\mathrm{POS}_P^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}, \mathrm{S7}\}, \tag{50}$$

$$\mathrm{NNG}_P^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}, \mathrm{S7}, \mathrm{S8}\}, \tag{51}$$

$$\mathrm{BND}_P^l(Cl_{\mathrm{Yes}}^{\geq}) = \{\mathrm{S8}\}, \tag{52}$$

$$\mathrm{POS}_P^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S9}, \mathrm{S10}, \mathrm{S11}\}, \tag{53}$$

$$\mathrm{NNG}_P^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S8}, \mathrm{S9}, \mathrm{S10}, \mathrm{S11}\}, \tag{54}$$

$$\mathrm{BND}_C^l(Cl_{\mathrm{No}}^{\leq}) = \{\mathrm{S8}\}. \tag{55}$$

Namely, we have $\mathrm{POS}_C^l(Cl_{\mathrm{Yes}}^{\geq}) \neq \mathrm{POS}_P^l(Cl_{\mathrm{Yes}}^{\geq})$. Then P cannot be an $\mathrm{L}_l^{\geq}$-reduct.

On the contrary, if we have an additional object S12 such that $\rho(\mathrm{S12}, Math) =$ Bad, $\rho(\mathrm{S12}, Phy) =$ Medium, $\rho(\mathrm{S12}, Lit) =$ Bad and $\rho(\mathrm{S12}, PS) =$ Good in the decision table, P is an $\mathrm{L}_l^{\geq}$-reduct.

From this observation, we know that we cannot always enumerate all union-based reducts in VP-DRSA based on a discernibility matrix.

Therefore, to enumerate all union-based reducts in VP-DRSA surely, an exhaustive enumeration would be necessary.

5 Concluding Remarks

We have investigated several reducts based on downward and upward unions in DRSA and VP-DRSA. Three different reducts have been obtained in DRSA whereas four different reducts have been obtained in VP-DRSA. The three kinds of reducts in DRSA can be enumerated by a decision matrix method while the four kinds of reducts cannot. The selection of a kind of reduct depends on the application. For example, an $\mathrm{L}^{\geq}$-reduct can be useful to know necessary attributes in inferring the certain lower bounds of classes while an $\mathrm{L}^{\leq}$-reduct can be useful to know necessary attributes in inferring the certain upper bounds of classes.

The investigations of reducts based on classes and the relations with reducts described in this paper would be our future topics.

Acknowledgments

This work has been partially supported by the Grant-in-Aid for Scientific Research (B) No.17310098.

References

1. Pawlak, Z.: Rough sets. International Journal of Information and Computer Sciences 11, 341–356 (1982)
2. Ślęzak, D.: Various approaches to reasoning with frequency based decision reducts: A survey. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) Rough set methods and applications, pp. 235–285. Physica-Verlag, Heidelberg (2000)

3. Inuiguchi, M., Tsurumi, M.: Measures based on upper approximations of rough sets for analysis of attribute importance and interaction. International Journal of Innovative Computing, Information and Control 2(1), 1–12 (2006)
4. Inuiguchi, M.: Attribute reduction in variable precision rough set model. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 14(4), 461–479 (2006)
5. Greco, S., Matarazzo, B., Słowiński, R.: The use of rough sets and fuzzy sets in MCDM. In: Gal, T., Stewart, T.J., Hanne, T. (eds.) Multicriteria decision making: Advances in MCDM models, algorithms, theory, and applications, pp. 14.1–14.59. Kluwer Academic Publishers, Boston (1999)
6. Greco, S., Matarazzo, B., Słowiński, R.: Decision rule approach. In: Figueira, J., Greco, S., Ehrgott, M. (eds.) Multiple criteria decision analysis, pp. 507–561. Springer, New York (2005)
7. Susmaga, R., et al.: Generation of reducts and rules in multi-attribute and multi-criteria classification. Control and Cybernetics 29(4), 969–988 (2000)
8. Inuiguchi, M., Yoshioka, Y.: Variable-precision dominance-based rough set approach. In: Greco, S., et al. (eds.) RSCTC 2006. LNCS (LNAI), vol. 4259. Springer, Heidelberg (2006)

Topologies of Approximation Spaces of Rough Set Theory

Milan Vlach[1,2]

[1] Kyoto College of Graduate Studies for Informatics
7 Monzen-cho, Tanaka, Sakyo-ku, Kyoto, 606-8225 Japan
m_vlach@kcg.ac.jp
[2] Charles University, School of Mathematics and Physics
Malostranské náměstí 25, 118 00 Praha 1, Czech Republic
milan.vlach@mff.cuni.cz

Summary. The main aim of this brief note is to explain relations between the classic approach to set approximations and recent proposals appearing in the literature on rough sets. In particular, relations between the standard topological concepts and basic concepts of rough set theory are considered.

1 Introduction

The theory of rough sets, as introduced by Pawlak, deals with situations in which knowledge about elements of a finite nonempty set is restricted by a collection of equivalence relations on that set. The basic idea is to use the available knowledge for introducing the notion of a definable subset and to approximate every undefinable subset through definable sets.

Various formulations and a number of generalizations of Pawlak's approach have been proposed and studied, and several notions and facts known in various branches of mathematics have been rediscovered and introduced under different names by the rough set community.

The main purpose of the present paper is to show that there is an intimate connection between the approximation operators of rough set theory and operators commonly met in topological spaces and their extensions. In particular, I would like to attract attention to the rarely noticed connection with the system of extended topology introduced and studied by Preston C. Hammer in a series of papers published in the early 1960's, and recently by Gniłka [2].

2 Preliminaries

2.1 Relations

Throughout the paper, U denotes a fixed nonempty (not necessarily finite) set and $\mathcal{P}(U)$ denotes the power set of U. If R is a binary relation in U, then the

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 176–186, 2008.
springerlink.com

converse R^{-1} of R is defined by $R^{-1} = \{(x, y) : (y, x) \in R\}$. The *identity* relation (also called the *diagonal*) in U is the set of all pairs of the form (x, x) for $x \in U$. If R and S are binary relation in U, then the *composite* of R and S is denoted by $R \circ S$; it is defined to be the set of all pairs (x, z) such that for some y it is true that $(x, y) \in S$ and $(y, z) \in R$.

A reflexive binary relation in U is called

- a *tolerance* if it is symmetric,
- a *preorder* if it transitive,
- an *equivalence* if it is transitive and symmetric.

By a *partition* of U we mean a collection of pairwise disjoint nonempty subsets of U whose union is U. Let us recall that partitions and equivalence relations are closely connected.

Let $\mathcal{E}$ be the mapping from the set of all partitions of U into the set of binary relations in U defined as follows. If D is a partition of U, then $\mathcal{E}(D)$ is the binary relation such that $(x, y) \in \mathcal{E}(D)$ just in case x and y belong to the same member of partition D. We shall call $\mathcal{E}(D)$ the relation *induced* by the partition D. It can easily be shown that $\mathcal{E}(D)$ is an equivalence relation.

On the other hand, every equivalence relation in U induces a partition as follows. If E is an equivalence relation between elements of U, then an *equivalence class* of E in U is a subset X of U such that

- every two elements in X are equivalent in the sense of E, and
- each element of U that is equivalent to some element of X belongs to X.

The equivalence class of an element x of U is the equivalence class to which x belongs. The equivalence class of x will be denoted by x/E, and the collection of all equivalence classes of E will be denoted by $\mathcal{D}(E)$. It turns out that the collection $\mathcal{D}(E)$ is a partition of U. Moreover, for each equivalence E and each partition D, we have

$$\mathcal{D}(\mathcal{E}(D)) = D \quad \text{and} \quad \mathcal{E}(\mathcal{D}(E)) = E.$$

2.2 Topologies

A *topology* for U is a collection τ of subsets of U satisfying the following conditions:

- The empty set and U belong to τ.
- The union of the members of each sub-collection of τ is a member of τ.
- The intersection of the members of each finite sub-collection of τ is a member of τ.

If τ is a topology for U then we say that the pair (U, τ) or U is a topological space. Let τ be a topology for U and X be a subset of U.

- The members of τ are called the open sets of (U, τ).
- The complements of the open sets are called the closed sets of (U, τ).
- The interior of X is the largest open subset of X.
- The closure of X is the smallest closed subset that includes X.

2.3 Uniformities

A *quasiuniformity* for U is a nonempty collection ϱ of subsets of $U \times U$ such that

(a) Each member of ϱ contains the diagonal.
(b) The intersection of any two members of ϱ also belongs to ϱ.
(c) If R is a member of ϱ and S is a subset of $U \times U$ such that $R \subset S$, then S also belongs to ϱ.
(d) For each $R \in \varrho$ there is an $S \in \varrho$ such that $S \circ S \subset R$.

The quasiuniformity ϱ for U is called a *uniformity* for U if the following additional condition is satisfied:

(e) If R is in ϱ, then the inverse R^{-1} of R is also in ϱ.

If ϱ is a quasiuniformity or uniformity for U, then the pair (U, ϱ) is said to be a *quasiuniform* or *uniform* space, respectively.

Every quasiuniformity ϱ on U yields a topology for U by taking as open sets the sets A with the property: if $x \in A$ then there is R in ϱ such that $\{y : (x, y) \in R\} \subset A$.

3 Definability in Rough Set Theory

3.1 Definability Based on Equivalences

Consider information in Table 1 taken from one of Pawlak's examples.

Table 1. A data table

Patient	Headache	Muscular pain	Temperature	Flue
1	no	yes	high	yes
2	yes	no	high	yes
3	yes	yes	very high	yes
4	no	yes	normal	no
5	yes	no	high	no
6	no	yes	very high	yes

First, notice that some patients are indiscernible with respect to some attributes or sets of attributes. For instance, patients 2, 3 and 5 are indiscernible with respect to Headache, patients 3 and 6 are indiscernible with respect to the pair Muscular pain and Flue. Also notice that every set of attributes divides the set of patients into a collection of pairwise disjoint nonempty subsets. For example,

Headache: {1,4,6}, {2,3,5};
Temperature: {1,2,5}, {3,6}, {4};
Muscular pain and Temperature: {2,5}, {3,6}, {4}, {1}.

Consider some set of patients, for example, the set $X = \{1, 3, 6\}$. It can be described by means of the pair of attributes Muscular pain and Temperature as the set of the patients feeling muscular pain and having high or very high temperature. However, it cannot be described exactly by the attribute Temperature.

Intuitively, this leads us to consider as definable subsets (with respect to a given equivalence relation) only the equivalence classes and unions of some equivalence classes.

Formally, we obtain the family of definable subsets as follows. Let E be an equivalence relation in U and let $\mathcal{D}(E)$ be the partition of U induced by E. The lower and upper approximations, and definability of a subset X of U considered by Pawlak [9] can be introduced as follows:

- The *E-lower approximation* $\underline{E}X$ of X is the union of those members of $\mathcal{D}(E)$ that are subsets of X.
- The *E-upper approximation* $\overline{E}X$ of X is the union of those members of $\mathcal{D}(E)$ that contain at least one element of X.
- A subset X of U is *E-definable* if $\underline{E}X = \overline{E}X$.

It can easily be verified that

- $\underline{E}X \subset X \subset \overline{E}X$,
- $\overline{E}X = U \setminus \underline{E}(U \setminus X)$ and $\underline{E}X = U \setminus \overline{E}(U \setminus X)$,
- $X \subset Y$ implies $\underline{E}X \subset \underline{E}Y$ and $\overline{E}X \subset \overline{E}Y$,
- $\underline{E}(X \cap Y) = \underline{E}X \cap \underline{E}Y$ and $\overline{E}(X \cup Y) = \overline{E}X \cup \overline{E}Y$,
- $\underline{E}(\underline{E}X) = \underline{E}X$ and $\overline{E}(\overline{E}X) = \overline{E}X$.

Also notice that

$$\underline{E}(\underline{E}X) = \overline{E}(\underline{E}X) \text{ and } \underline{E}(\overline{E}X) = \overline{E}(\overline{E}X).$$

In other words, for each X, Pawlak's lower and upper approximations $\underline{E}X$ and $\overline{E}X$ are definable sets.

3.2 Definability Based on Tolerances

For a tolerance relation T, we define x/T to be the set of those elements y in U for which $(x, y) \in T$.

If T is a transitive tolerance, then T is an equivalence and x/T is the equivalence class of x. As a consequence, for every x and y, if x/T and y/T are not equal, then they are disjoint. However, if a tolerance T is not transitive, then two distinct x/T and y/T may have common elements. Therefore, the collection $\{x/T : x \in U\}$ of subsets of U is not necessarily a partition of U. By the reflexivity of T, $\{x/T : x \in U\}$ is always a covering of U. We denote it by $\mathcal{C}(T)$.

As an example, let us consider Table 2 taken from Järinen [5].

From the table we extract the following tolerance relation:

$$T = \Delta \cup \{(1,2), (2,1), (1,3), (3,1), (5,4), (4,5), (5,6), (6,5), (5,7), (7,5)\}$$

where Δ is the identity relation in the set of patients. The corresponding sets x/T are given in Table 3.

Table 2. A data table

Patient	Blood pressure	Hemoglobin	Temperature
1	103/65	125	39.3
2	97/60	116	39.1
3	109/71	132	39.2
4	150/96	139	37.1
5	145/93	130	37.3
6	143/95	121	37.8
7	138/83	130	36.7

Table 3. Tolerance classes

$1/T$	$2/T$	$3/T$	$4/T$	$5/T$	$6/T$	$7/T$
{1,2,3}	{1,2}	{1,3}	{4,5}	{4,5,6,7}	{5,6}	{5,7}

Following Pawlak's approach, several authors (see, for example, [5, 6, 15]) have introduced the lower approximations, upper approximations and definability of sets with respect to tolerances as follows.

Let T be a tolerance relation in U, let $\mathcal{C}(T)$ be the covering of U defined by T, and let X be a subset of U.

- The *T-lower approximation* $\underline{T}X$ of X is the union of those members of $\mathcal{C}(T)$ that are subsets of X.
- The *T-upper approximation* $\overline{T}X$ of X is the union of those members of $\mathcal{C}(T)$ that contain at least one element of X.
- A subset X of U is *T-definable* if $\underline{T}X = \overline{T}X$.

It can easily be verified that

- $\underline{T}X \subset X \subset \overline{T}X$,
- $\overline{T}X = U \setminus \underline{T}(U \setminus X)$ and $\underline{T}X = U \setminus \overline{T}(U \setminus X)$,
- $X \subset Y$ implies $\underline{T}X \subset \underline{T}Y$ and $\overline{T}X \subset \overline{T}Y$.

Let T be a tolerance in U and let E_T be the intersection of all equivalences in U that include T. It turns out that E_T is an equivalence, and the collection of T-definable sets is the same as the collection of E_T-definable sets.

Nevertheless, there is a significant difference between approximations based on tolerances and those based on equivalences. For an equivalence, both lower and upper approximation of a set is always a definable set, while this is not true for tolerances. For instance, in the example above, the upper approximation $\overline{T}\{2\}$ is T-undefinable, because $\overline{T}\{2\} = \{1, 2\}$, and $\underline{T}\{1, 2\} = \{2\}$, $\overline{T}\{1, 2\} = \{1, 2, 3\}$.

3.3 Definability Based on Preorders

Let Q be a preorder in U and x be an arbitrary point in U. Let $Q^{\leftarrow}(x)$ and $Q^{\rightarrow}(x)$ be the subsets of U defined by

$$Q^{\leftarrow}(x) = \{y \in U : (y, x) \in Q\}$$

$$Q^{\rightarrow}(x) = \{y \in U : (x, y) \in Q\}$$

If we wish to define the upper and lower and upper approximations with respect to preorders analogously to the cases of equivalences and tolerances, we have to take into account that preorders may be unsymmetrical; that is, in general, $Q^{\leftarrow}(x) \neq Q^{\rightarrow}(x)$. Hence we have several possibly nonequivalent variants how to define the lower and upper approximations $\overline{Q}$ and $\underline{Q}$. Namely,

$$\overline{Q}X = \{x \in U : Q^{\rightarrow}(x) \cap X \neq \emptyset\} \tag{1}$$
$$\underline{Q}X = \{x \in U : Q^{\rightarrow}(x) \subset X\} \tag{2}$$
$$\overline{Q}X = \{x \in U : Q^{\leftarrow}(x) \cap X \neq \emptyset\} \tag{3}$$
$$\underline{Q}X = \{x \in U : Q^{\leftarrow}(x) \subset X\} \tag{4}$$
$$\overline{Q}X = \{x \in U : Q^{\rightarrow}(x) \cap X \neq \emptyset\} \tag{5}$$
$$\underline{Q}X = \{x \in U : Q^{\leftarrow}(x) \subset X\} \tag{6}$$
$$\overline{Q}X = \{x \in U : Q^{\leftarrow}(x) \cap X \neq \emptyset\} \tag{7}$$
$$\underline{Q}X = \{x \in U : Q^{\rightarrow}(x) \subset X\} \tag{8}$$

For a preorder Q, we define the upper approximations $\overline{Q}X$ and lower approximations $\underline{Q}X$ of X by

$$\overline{Q}X = \{x \in U : Q^{\rightarrow}(x) \cap X \neq \emptyset\} \tag{9}$$
$$\underline{Q}X = \{x \in U : Q^{\leftarrow}(x) \subset X\} \tag{10}$$

and we say that a subset X of U is *Q-definable* if

$$\underline{Q}X \;=\; \overline{Q}X.$$

To illustrate reasons for this choice, we consider the following example taken from Järinen and Kortelainen [6].

Let U be the set $\{1, 2, 3\}$, and consider the preorder

$$Q = \{(1,1), (2,2), (3,3), (1,2), (1,3), (2,3)\}.$$

Then we obtain Tables 4–6.

The last table shows that the richest collection of definable sets is obtained for our choice or for the "dual" choice

$$\overline{Q}X = \{x \in U : Q^{\leftarrow}(x) \cap X \neq \emptyset\} \tag{11}$$
$$\underline{Q}X = \{x \in U : Q^{\rightarrow}(x) \subset X\}. \tag{12}$$

Table 4. Images and inverse images under Q

x	$Q^{\leftarrow}(x)$	$Q^{\rightarrow}(x)$
1	{1}	{1,2,3}
2	{1,2}	{2,3}
3	{1,2,3}	{3}

Table 5. Lower and upper approximations

X	$\overline{Q}X, \rightarrow$	$\overline{Q}X, \leftarrow$	$\underline{Q}X, \rightarrow$	$\underline{Q}X, \leftarrow$
{1}	{1}	{1,2,3}	$\emptyset$	{1}
{2}	{1,2}	{2,3}	$\emptyset$	$\emptyset$
{3}	{1,2,3}	{3}	{3}	$\emptyset$
{1,2}	{1,2}	{1,2,3}	$\emptyset$	{1,2}
{1,3}	{1,2,3}	{1,2,3}	$\emptyset$	{1}
{2,3}	{1,2,3}	{2,3}	{2,3}	$\emptyset$

Table 6. Lower and upper approximations for (5) and (6)

$\underline{Q}\emptyset$	=	$\emptyset$	$\overline{Q}\emptyset$	=	$\emptyset$
$\underline{Q}\{1\}$	=	$\{1\}$	$\overline{Q}\{1\}$	=	$\{1\}$
$\underline{Q}\{2\}$	=	$\emptyset$	$\overline{Q}\{2\}$	=	$\{1,2\}$
$\underline{Q}\{3\}$	=	$\emptyset$	$\overline{Q}\{3\}$	=	$\{1,2,3\}$
$\underline{Q}\{1,2\}$	=	$\{1,2\}$	$\overline{Q}\{1,2\}$	=	$\{1,2\}$
$\underline{Q}\{2,3\}$	=	$\emptyset$	$\overline{Q}\{2,3\}$	=	$\{1,2,3\}$
$\underline{Q}\{1,3\}$	=	$\{1\}$	$\overline{Q}\{1,3\}$	=	$\{1,2,3\}$
$\underline{Q}\{1,2,3\}$	=	$\{1,2,3\}$	$\overline{Q}\{1,2,3\}$	=	$\{1,2,3\}$

3.4 General Case

Let R be an arbitrary relation in U and x be an arbitrary point in U. Let $R^{\leftarrow}(x)$ and $R^{\rightarrow}(x)$ be the subsets of U defined by

$$R^{\leftarrow}(x) = \{y \in U : (y, x) \in R\},$$

$$R^{\rightarrow}(x) = \{y \in U : (x, y) \in R\}.$$

Let us introduce the upper approximations $\overline{R}X$ and lower approximations $\underline{R}X$ of X by

$$\overline{R}X = \{x \in U : R^{\rightarrow}(x) \cap X \neq \emptyset\},$$
$$\underline{R}X = \{x \in U : R^{\leftarrow}(x) \subset X\}.$$

Then we say a subset X of U is *R-definable* if

$$\underline{R}X = \overline{R}X.$$

Given that equivalences and tolerances are symmetric, it is immediate that these approximations, and consequently the definability, coincide with those described above for equivalences, tolerances and preorders.

4 Topologies of Approximation Spaces

Let D be a partition of U. It can easily be seen that the collection of all sets that can be written as unions of some members of D together with the empty set is a topology for U. This topology is called the *partition topology* generated by D. Because every equivalence E in U defines a partition of U, it also generates a topology for U; namely, the partition topology generated by the partition $\mathcal{D}(E)$. We denote it by τ_E and, if there is no danger of misunderstanding, we omit references to E.

It is worth mentioning that partition topologies are characterized by the fact that every open set is also closed, and that every partition topology is an Alexandrov topology. That is, a topology for which the intersection of the members of each (not only finite) collection of open sets is open.

In the rough set literature, the ordered pair (U, E), where U is a (finite) nonempty set and E is an equivalence in U, is called an approximation space or Pawlak's approximation space. In order to see clearly how Pawlak's approximation spaces are intimately related with the topological spaces, we observe that:

- A subset X of U is E-definable if and only if it is either empty or it can be written as the union of some members of the partition induced by E.
- A subset X of U is τ_E-open if and only if it is either empty or it can be written as the union of some members of the partition induced by E.

Moreover, if E is an equivalence in U, then the collection ϱ of subsets of $U \times U$ defined by

$$\varrho = \{R : R \subset U \times U, E \subset R\}$$

is a uniformity for U and the topology for U induced by this uniformity coincides with topology τ_E.

Consequently,

- Pawlak's approximation spaces are uniform spaces whose uniform topologies coincide with partition topologies.
- These topologies can be characterized by the fact that every open set is also closed.
- These topologies are *Alexandrov topologies*.

Hence the difference between Pavlak's approximation space (U, E) and the topological space (U, τ_E) is only terminological. In particular, we have

- X is definable if and only if it is open.
- X is definable if and only if it is closed.

- The lower approximation of X is the interior of X.
- The upper approximation of X is the closure of X.
- X is definable if and only if its interior is equal to its closure.

For translation of some other terms of Pawlak's terminology into the standard language of general topology, see [14].

Let T be a tolerance in U and let E_T be the intersection of all equivalences in U that include T. It turns out that E_T is an equivalence, and the collection of T-definable sets is the same as the collection of E_T-definable sets. Therefore, for each tolerance T, the collection of T-definable sets is a partition topology. Moreover, the collection ϱ of subsets of $U \times U$ defined by $\varrho = \{R : R \subset U \times U, T \subset R\}$ is a quasiuniformity for U and the topology for U induced by this quasiuniformity coincides with the partition topology generated by τ_{E_T}.

Let Q be a preorder in $\mathcal{U}$. Järvinen and J. Kortelainen [6] showed that the collection of Q-definable sets is again a topology for U. We denote this topology by τ_Q. This topology is not necessarily a partition topology but it is always an Alexandrov topology.

Conversely, let τ be an Alexandrov topology for U, and let Q_τ be a binary relation in U defined by $(x, y) \in Q_\tau$ if and only if $N_x \subset N_y$ where N_x and N_y are the neighborhood systems of x and y, respectively.

It turns out that τ_Q is an Alexadrov topology if Q is a preorder, Q_τ is a preorder if τ is an Alexandrov topology, and

$$\tau_{Q_\tau} = \tau \text{ and } Q_{\tau_Q} = Q.$$

5 Hammer's Extended Topology

If R is a binary relation in U, then the function

$$X \mapsto \{y \in U : (x, y) \in R \text{ for some } x \in X\}$$

maps the power set $\mathcal{P}(U)$ of U into itself. Therefore, we can obtain an insight into the system of approximations of rough set theory also by studying properties of set-valued set-functions.

Basic properties of such functions have been studied in detail, for example, by Preston C. Hammer [4] in a series of papers on the extended topology published in the early 1960's; see also Gniłka [2].

Let

$$f : \mathcal{P}(U) \to \mathcal{P}(U)$$

be an arbitrary function. Define the *dual* f_d of f by

$$f_d(X) = U \setminus f(U \setminus X).$$

Since the lower and upper approximations used in rough set theory are *isotonic* functions in the sense that

$$X \subset Y \text{ implies } f(X) \subset f(Y)$$

we focus our attention on the class of isotonic functions.

The following properties are equivalent:

$$f \text{ is isotonic,}$$

$$f(X) \cup f(Y) \subset f(X \cup Y) \text{ for all } X, Y \in U,$$

$$f(X) \cap f(Y) \supset f(X \cap Y) \text{ for all } X, Y \in U,$$

$$f_d \text{ is isotonic,}$$

$$f_d(X) \cup f_d(Y) \supset f_d(X \cup Y) \text{ for all } X, Y \in U,$$

$$f_d(X) \cap f_d(Y) \subset f_d(X \cap Y) \text{ for all } X, Y \in U,$$

$$X \cap Y = \emptyset \text{ implies } f(X) \cap f_d(Y) = \emptyset.$$

In Hammer's system of extended topology the isotonic functions that are also non-shrinking in the sense

$$X \subset f(X) \text{ for all } X$$

are called *expansive* functions. It can easily be seen that if f is expansive, then f_d is *contractive*; that is, isotonic and non-enlarging in the sense

$$X \supset f_d(X) \text{ for all } X.$$

We have seen that the upper approximations are expansive and lower approximations are contractive. Moreover the upper approximations are *subadditive* in the sense that

$$f(X) \cup f(Y) \supset f(X \cup Y) \text{ for all } X, Y \in U,$$

and the lower approximations are *submultiplicative* in the sense

$$f(X) \cap f(Y) \subset f(X \cap Y) \text{ for all } X, Y \in U.$$

The fact that the isotonic approximations are exactly those with the property

$$X \cap Y = \emptyset \text{ implies } f(X) \cap f_d(Y) = \emptyset$$

is important in optimization theory, because the optimality of a feasible solution can be expressed by stating that the intersection of certain sets is empty.

In contrast to the approximations used in the theory of rough sets, the approximations used in optimization are usually neither expansive nor contractive. Properties of such approximations have been studied in Vlach (1980, 1983), where the isotonic approximations with the property

$$f(X) = X \text{ for } X \in \{\emptyset, U\}$$

are called *external* approximations, provided they are subadditive, and *internal* approximations, if they are submultiplicative.

6 Conclusions

The paper shows that there exists an intimate relationship between the theory rough sets and the theory of topological spaces. Namely: if the underlying indiscernibility relation is an equivalence, then the collection of definable sets is a uniformity whose topology is a partition topology, that is, a topology in which every open set is also closed and vice versa; if the collection of definable sets is a tolerance, then the collection of definable sets is a quasiuniformity whose topology is again a partition topology; if the underlying indiscernibility relation is a preorder, then the collection of definable sets is again a topology but not necessarily a partition topology. However, in all mentioned cases, the collection of definable sets is an Alexandrov topology, that is, an arbitrary intersection of definable sets is a definable set.

References

1. Alexandroff, P.: Discrete Räume. Mat. Sb. 2, 501–518 (1937)
2. Gniłka, S.: On extended topologies. I: Closure operators. Ann. Soc. Math. Pol., Ser. I, Commentat. Math. 34, 81–94 (1994)
3. Halmos, P.R.: Naive Set Theory. Springer, New York (1987)
4. Hammer, P.C.: Extended topology and systems. Math. Systems Theory 1, 135–142 (1967)
5. Järvinen, J.: Approximations and rough sets based on tolerances. In: Ziarko, W., Yao, Y. (eds.) RSCTC 2000. LNCS (LNAI), vol. 2005, pp. 182–189. Springer, Heidelberg (2001)
6. Järvinen, J., Kortelainen, J.: A note on definability in rough set theory. In: De Baets, B., et al. (eds.) Current Issues in Data and Knowledge Engineering, Problemy Współczesnej Nauki, Teoria i Zastosowania, Informatyka, Warsaw, Poland, pp. 272–277 (2004)
7. Kelley, J.L.: General Topology. D. Van Nostrand Company, London (1967)
8. Lin, T.Y.: Neighborhood Systems: A qualitative theory for fuzzy and rough sets. In: Wang, P. (ed.) Advances in Machine Intelligence and Soft Computing, vol. IV, pp. 132–155 (1997)
9. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
10. Stopher Jr., E.C.: Point set operators and their interrelations. Bull. Amer. Math. Soc. 45, 758–762 (1939)
11. Vlach, M.: Approximation Operators in Optimization Theory. Zeitschrift für Operations Research 25, 15–23 (1981)
12. Vlach, M.: On nonenlarging closures in affine spaces. In: Novak, J. (ed.) General Topology and its Relations to Modern Analysis and Algebra V, pp. 653–656. Helderman-Verlag, Berlin (1982)
13. Vlach, M.: Closures and neighbourhoods induced by tangential approximations. In: Hammer, G., Pallaschke, D. (eds.) Selected Topics in Operations Research and Mathematical Economics, pp. 119–127. Springer, Berlin (1984)
14. Vlach, M.: Rough sets, neighborhoods, and set approximations. In: Noguchi, H., Ishii, H., Inuiguchi, M. (eds.) Proceedings of 7th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, Japan, pp. 155–157 (2004)
15. Wybraniec-Skardovska, U.: On a generalization of approximation space. Bulletin of The Polish Academy of Sciences: Mathematics 37, 51–62 (1989)

Uncertainty Reasoning in Rough Knowledge Discovery

Yaxin Bi[1], Xuhui Shen[2], and Shengli Wu[1]

[1] School of Computing and Mathematics, University of Ulster
Newtownabbey, Co. Antrim, BT37 0QB, UK
{y.bi, s.wu}@ulster.ac.uk
[2] Institute of Earthquake Science, China Earthquake Administration
Beijing, 100036 China
shenxh@seis.ac.cn

Summary. In this paper we develop a rough set-based solution to dealing with the inconsistent decision classes of instances in inducing decision rules and an evidential reasoning method to resolve such inconsistent conclusions encountered in determining class decisions for instances. The distinguishing aspects of our method are to exploit the rough boundary region in inducing rules and to aggregate multiple conclusions in classifying instances. We present our proposed method and use an example to illustrate how our method can be applied to classification problems along with its advantage.

1 Introduction

The starting point of rough set theory is the observation that instances may be indiscernible due to limited information available [1]. In such a situation, two instances with the same attribute values may be classified into two different decision classes. Thus the decision classes relating to these instances may be inconsistent and uncertain. In the rough set-based inductive learning approaches, such inconsistency is not corrected. Instead, it will be replaced by the lower and upper approximations of the decision classes [2]. On the basis of these approximations, two sets of rules are generated: certain and uncertain rules. The certain rules can be used as in conventional rule-based systems, and the uncertain rules may be used to cope with non-deterministic instances in classification.

In a rule-based classification system, classifying instances is to select single rules in which their conditions either exactly or partially match instances and their conclusions will be assigned to the instances. For a given instance, there often exist more than one rule to be activated, this implies that one instance may belong to more than one decision class. These classes may potentially be in inconsistence or conflict. To deal with such an issue, the conventional approaches used in most rule-based systems is to select one and only one rule in inference process to avoid conflict cases, these include 1) ordering rules and picking the first or last rule matched; 2) starting from the first rule and taking the earliest matched rule; and 3) putting a default rule at the end of rule list in case no rule

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 187–200, 2008.
springerlink.com

is activated [3, 4]. In the context of rough set-based inductive systems, such a issue has not drawn a great deal of attention and there is a little effort devoted to how conflicting cases could be resolved.

Given the above two cases of conflicting decision classes arising from rule-based systems, we propose a novel approach. For the former we suggest to make use of the rough set boundary of instances to generate uncertain rules instead of using the upper approximation. We argue that inconsistent decision classes should be treated as a set of classes with a certain degree of confidence, because they cannot ascertain any individual decision classes. For the later case we put forward an effective solution to resolve inconsistent decisions by accumulating evidence derived from multi-conclusions. This solution is developed on the basis of the Dempster-Shafer theory of evidence.

The rest of paper is organized as follows. Section 2 details the rough sets-based algorithms for computing minimal sets of attributes called reducts for computing lower and upper approximations, which are, in turn, used to generate certain and uncertain rules. Section 3 provides a method for modeling multiple conclusions drawn as uncertainty reasoning problem. The final summary is given in Section 4.

2 A Rough Sets-Based Method for Rule Induction

To properly understand how rules can be induced using a rough sets-based learning algorithm, let us start with a formal model for rule induction [1, 5].

2.1 Preliminary

In rough sets, objects / instances are organized into an *information system*, denoted by $\mathcal{I} =< U, A >$, where

1. $U = \{u_1, u_2, \cdots, u_{|U|}\}$ is a finite non-empty set, called the *universe* or *object space*; elements of U are called *objects/instances*;
2. $A = \{a_1, a_2, \cdots, a_{|A|}\}$ is also a finite non-empty set; elements of A are called *attributes*;
3. for every $a \in A$ there is a mapping for a from U into some space $a : \ U \rightarrow a(U)$, and $a(U) = \{a(u) \mid u \in U\}$ is called the *domain* of attribute a.

An information system $\mathcal{I}$ can be expressed intuitively in terms of an *information table* as follows:

$$\frac{U \backslash A \mid a_j}{u_i \quad \mid a_j(u_i)} \tag{1}$$

where attribute a_j has domain $a_j(U) = \{a_j(u_i) \mid i = 1, 2, ..., |U|\}$. When there is no repetition of objects, the information table is a relation, and attributes can be divided into two categories of condition and decision based on their dependencies. An information system $< U, A >$ is called a decision system, if we have $A = C \cup D$ and $C \cap D = \emptyset$, where attributes in C are called condition attributes and attributes in D are called decision attributes.

Given a decision system, a range of operations can be defined below.

Table 1. A decision system: h = headache, m = muscle-pain, t = temperature, d = flu or non-flu, Y = yes, N = no, H = high, VH = very high

U/Y	h	m	t	d
1	N	Y	H	Y
2	Y	N	H	Y
3	Y	Y	VH	Y
4	N	Y	N	N
5	Y	N	H	N
6	N	Y	VH	Y

$$U/h = \{W_1, W_2\} = \{\{1,4,6\},\{2,3,5\}\}$$
$$U/m = \{W_1, W_2\} = \{\{1,3,4,6\},\{2,5\}\}$$
$$U/t = \{W_1, W_2, W_3\} = \{\{1,2,5\},\{3,6\},\{4\}\}$$
$$U/d = \{W_1, W_2\} = \{\{1,2,3,6\},\{4,5\}\}$$

Fig. 1. Partitions of attributes h,m, t and d

Definition 1. Let U be a universe U and $a \in A$ be an attribute, a partition operation over U is defined by $U/\{a\} = \{W_1, \cdots, W_K\}$, simply denoted by U/a, where for any $u_1, u_2 \in W_i$, $u_1 \neq u_2$, $a(u_1) = a(u_2)$, and W_i is called a partition.

Definition 2. Let W_1 and W_2 be two subsets of U, an intersection operation is defined by $W_1 \otimes W_2$ below:

$$W_1 \otimes W_2 = \{u \in U |\ u \in W_1,\ u \in W_2\} \tag{2}$$

To make the above concepts more clear, let us look at a well-known example drawn from the work of Pawlak as follows [6]. The example is represented as a decision system $< U, A >$ in Table 1, containing information about patients' symptoms and the diseases from which patients suffer, $U = \{1,2,3,4,5,6\}$ and $A = C \cup D$, where $C = \{h, m, t\}$ are condition attributes and $D = \{d\}$ is a decision attribute. The partitions of the attributes are illustrated in Figure 1 and the intersections of partitions with respect to attributes h and m are presented in Figure 2.

$$U/hm = U/h \otimes U/m$$
$$= \{\{1,4,6\},\{2,3,5\}\} \otimes \{\{1,3,4,6\},\{2,5\}\}$$
$$= \{W_1, W_2, W_3\} = \{\{1,4,6\},\{3\},\{2,5\}\}$$

Fig. 2. Intersection of partitions of attributes h and m

2.2 Decision Rules

In general, decision rules as knowledge representation can be used to express various relations and semantics between condition attributes (or the antecedent)

and decision attributes (or the consequent). For instance, a rule can be used to express a cause-effect relation: if you have headache and a high temperature, then you may have caught flu. It can express semantics, e.g. if $A =$"*gender*" and $B =$"*sex*", then A and B are semantically equivalent. It also can express an association discovered from a supermarket transaction, e.g., if someone buys breads, then he or she probably buys milks as well. According to the types of outputs which inductive learning algorithms have, decision rules can be grouped into three categories as shown in Figure 3. In this figure, $c_i \in C$ is a attribute,

$$\begin{array}{|l|}\hline 1)IF(c_1/v_1) \wedge \cdots \wedge (c_M/v_M)THENd_1|stg \\ 2)IF(c_1/v_1) \wedge \cdots \wedge (c_M/v_M)THEN\{d_1|stg_1, \cdots, d_s|stg_s\} \\ 3)IF(c_1/v_1) \wedge \cdots \wedge (c_M/v_M)THEN\{d_1, \cdots, d_s\}|stg_s \\ \hline \end{array}$$

Fig. 3. Three types of decision rules

$v_i \in c_i(U)$ is an attribute value of c_i and $d_i \in d(U)$ is a decision attribute value, and stg_i is a degree of belief about possible conclusions, called a rule strength and also denoted by $stg(d_i) = stg_i$. It lies in a range [0, 1]. The intuitive interpretation of rule execution is straightforward. Given an instance, if it exactly matches the condition of a rule, then three possible conclusions could be drawn from three types of rules: 1) the conclusion is d_1 with the degree of belief stg_1; 2) the conclusion is a set of decisions along with their degrees of belief, and 3) the conclusion is a set of decisions with one degree of belief. This division of rules is slightly different from that given by Song [7]. We merge the three types of rules described by Song into the first two and add the third type of rule that can be used to deal with a set of conclusions in decision making under uncertainty. In this study, we use Type 1 and 3 of rules to represent different uncertain relations between condition and decision attributes in different situations.

To illustrate the process of rule induction, let us look at Table 1. Since these instances do not belong to the same decision class, we need a way to split them into subsets based on the condition attribute values. First, we consider partitions on attributes h and d to see whether all the instances can be classified into a decision class *flu* or *not-flu*.

With partitions $U/h = \{W_1^h, W_2^h\}$ and $U/d = \{W_1^d, W_2^d\}$, it can be seen that no subsets of U/h belongs to any subsets of U/d. This means that instances with respect to attribute h cannot be classified into the decision class *flu* or *not-flu*. Hence we need to add another attribute to h to classify the instances in U. For example, we add attribute m into h to yield another set of partitions $U/hm = \{W_1^{hm}, W_2^{hm}, W_3^{hm}\}$, we notice $W_2^{hm} \subseteq W_1^d$, but the remaining partitions still do not belong to U/d, i.e. some instances still cannot be classified into *flu* or *not-flu* with attributes h and m. Thus we need to add another attribute to hm to partition U. This process will be repeated for each of the remaining attributes until all the instances in U are classified into the decision classes of *flu* or *not-flu*.

Unfortunately, in some situations, some instances in U cannot be classified into the decision classes even if all the combinations of attributes are considered.

For example, instances 2 and 5 in Table 1, although the two instances have the same attribute values, the decision class to which instance 2 belongs is different from that of instance 5. That means that two instances cannot be characterized in terms of the information available, resulting in inconsistency or uncertainty, in determining a decision class for instances 2 and 5. This leads to a difficult situation to induce decision rules.

The rough set theory as a mathematical tool provides an effective mechanism to cope with inconsistencies or uncertainty in determining decision classes for instances. Unlike the approaches widely used in other inductive learning algorithms, such inconsistencies are corrected [2] or instances are removed. Instead, these inconsistent classes will be replaced by a pair of the lower and/or upper approximations of the decision classes. On the basis of the approximations, two sets of rules are induced: certain and uncertain. The certain rules can be used as in conventional rule-based systems, and the possible rules may be used for dealing with uncertainty. In the following sections, we will see how the two types of rules can be induced by the rough set-based algorithms and how they can be exploited for classification.

2.3 An Algorithm for Computing Multiple Reducts

Inductive learning can be seen as a process of synthesizing evidential mappings from a sample space consisting of individual instances (objects). The result often is to reduce the space containing individual instances in size, forming a new smaller space containing a set of representative instances. These remaining instances serve the same role as the original ones. In contrast, the rough set-based inductive learning is aimed at learning a subset of attributes in terms of a reduct which is a minimal sufficient subset of a set of condition attributes. It preserves the dependency degree with respect to a set of decision attributes and has the same ability to discriminate instances as a full set of attributes. Prior to developing algorithms, we introduce several important concepts below [1, 5].

Definition 3. Suppose that a subset of attributes $X \subseteq A$, if there is $Q \subseteq X$, and Q is minimal among all the subsets of X with the condition of $U/X = U/Q$, then Q is called a reduct of X.

Notice that attributes within a reduct are significant such that none of them can be omitted, and there may be more than one reduct for a given subset X.

Definition 4. For each subset $W \in U/Y = \{W_1, \cdots, W_K\}$ with respect to a decision attribute $Y \subseteq D$, and a subset of condition attributes $X \subseteq C$, we associate two subsets with W:

$$\underline{X}W = \cup\{v \subseteq U/X \mid v \subseteq W\} \tag{3}$$

$$\overline{X}W = \cup\{v \subseteq U/X \mid v \cap W \neq \emptyset\} \tag{4}$$

In the above notation, $\underline{X}W$ is called a *lower* approximation of W and $\overline{X}W$ is called a *upper* approximation of W. In addition $POS_X = \underline{X}W_1 \cup \underline{X}W_2 \cup \cdots \cup \underline{X}W_K$ is called a X-positive region and $BN_X = \overline{X}W - \underline{X}W$ is called a boundary.

$$\begin{array}{l}\underline{X}W = \underline{X}W_1 \cup \underline{X}W_2 = \{1,3,6\} \cup \{4\} = \{1,3,4,6\} \\ \overline{X}W = \overline{X}W_1 \cup \overline{X}W_2 = \{1,3,6,2,5\} \cup \{4,2,5\} \\ \qquad\qquad\qquad\qquad = \{1,2,3,4,5,6\} \\ BN_X = \overline{X}W - \underline{X}W = \cup_{i=1}^{2}\overline{X}W_i - \cup_{i=1}^{2}\underline{X}W_i = \{2,5\}\end{array}$$

Fig. 4. Lower and upper approximations, and boundary

$$\begin{array}{l}spt_X(W_1) = spt_{\{m,t\}}(W_1) = \frac{|\{1,3,6\}|}{|U|} = \frac{3}{6} \\ spt_X(W_2) = spt_{\{m,t\}}(W_2) = \frac{|\{4\}|}{|U|} = \frac{1}{6} \\ \gamma_X(Y) = \gamma_{\{m,t\}}(Y) = spt_X(W_1) + spt_X(W_2) = \frac{4}{6}\end{array}$$

Fig. 5. Support and dependency degree

To quantify relation between a subset of condition attributes X and a decision attribute Y, we define two measures: *support*, denoted by $spt_X(W)$, and *dependency degree*, denoted by $\gamma_X(W)$ (being equivalent to *stg* defined in Fig. 3), below.

$$spt_X(W) = \frac{|\underline{X}W|}{|U|},\ W \in U/X \tag{5}$$

$$\gamma_X(Y) = \sum_{W \in U/Y} spt_X(W) \tag{6}$$

Formula (6) expresses the dependency degree between X and Y. If $\gamma_X(Y) = 1$, Y is totally dependent on X. This means all the instances in U can be classified by using a subset of attributes X. If $0 < \gamma_X(Y) < 1$, then Y is partially dependent on X. In other words, only those instances which fall in the positive region can be classified by using X. When $\gamma_X(Y) = 0$, none of instances can be classified using a subset of attributes X.

As an example, we look at Table 1 again. Suppose we are given a subset of attributes $X = \{m,t\}, Y = \{d\}$, and $U/mt = \{\{1\},\{3,6\},\{4\},\{2,5\}\}$ $U/d = \{W_1, W_2\} = \{\{1,2,3,6\},\{4,5\}\}$, by using formulas (3)-(6), we can obtain lower and upper approximations, dependency as shown in Figure 4 and the corresponding support and the degree of dependency as depicted in Figure 5.

With the definition of the degree of dependency between two sets of attributes, we proceed to introduce an algorithm for calculating reducts. In our work, we have implemented an algorithm which is similar to the algorithms developed using single covering approaches [2, 8]. The key idea of the algorithm is to search subsets of attributes with the maximum dependency degree from a set of attributes C in terms of reducts. It starts with an empty subset Q and then computes an attribute with the largest dependency to be added into Q. Removing Q from C, the algorithm continues to calculate another attribute with the second largest dependency, which will be added into Q. This process will be repeated until the degree of dependency of Q subset of attributes does not increase. Figure 6 illustrates a simplified algorithm for generating reducts.

Algorithm: generate reducts

```
1   Input C a set of condition attributes,
     D a set of decision attributes
2   Output Q: holds a reduct, Q ⊆ C
3   Q ← ∅,  X ← ∅
4   while (γ_{Q∪{x}}(D) ≠ γ_T(D))
5      T ← Q
6      for each x ∈ (C − Q) do
7         if (γ_{Q∪{x}}(D) > γ_T(D)) then
8            T ← Q ∪ {x}
9            X ← {x}
10        endif
11        Q ← T
12     endfor
13  endwhile
14  output Q
15  C ← C − X
16  if (|C| > 1) go to 2
17  end
```

Fig. 6. A rough set for generating reducts

2.4 Rough Sets for Rule Induction

Using rough sets-based methods, rules can be generated from the instances belonging to the lower approximation, the upper approximation and the boundary of different decision classes [2]. Given the lower approximation of a reduct, a set of rules can be generated in the form of Type 1 as illustrated in Figure 3, and their strengths can be calculated by formula (5). These rules are regarded as certain rules in this context. Rules generated from the other regions are regarded as uncertain rules, which can be used for coping with vagueness and uncertainty in classification.

The approaches to inducing uncertain rules from an upper approximation have been used in most of the rough sets-based systems [3]. However there are few suggestions about how the uncertain rules can be derived from the boundary and how uncertain rules derived in this way can be exploited. In the boundary case, the consequent part of such rules consists of a set of contradicting decision classes in the disjunctive form. However choosing the best supported decision from them is cumbersome because we have no criterion for identifying that best individual, all we could do is to pick one at random. In this study we propose that such contradicting decision classes could be treated as a set of decision classes, representing an undeterministic status of rule induction. Figures 7 and 8 illustrate two algorithms for rule construction from low approximation and boundary, respectively.

Returning the example and calculating the dependencies for all possible subsets of C, we can obtain two reducts $\{m, t\}$ and $\{h, t\}$. Inputting the two reducts

```
Algorithm: generating rules
1   Input: U/d = {W_1, ..., W_K} a set of partitions
       with respect to decision attribute d;
       X a reduct; X̲W_i a lower approximation
       mask[] an index of attributes
2   Output: R_i a rule
3   S ← X̲W_i
4   for (i = 1to|S|) do
5       stg ← 1
6       for (j = i + 1to|S|) do
7           if (¬mask[j]) then
8               if (X(S[i] == X(S[j])) then
9                   stg ← stg + 1
10                  mask[j] ← true;
11              endif
12          endif
13      endfor
14      R_i ← IF
15      for each (x ∈ X)
16          R_i ← R_i ∧ (x/x(S[i]))
17          R_i ← R_i + THEN d_i|stg/|U|
18    endfor
18  endfor
19  output R_i
20  end
```

Fig. 7. Constructing rules from lower approximation

and boundary into the algorithms of rule constructions, two sets of rules can be generated as shown in Figures 9 and 10.

In the following section, we examine the use of rules induced from the lower approximation and the corresponding boundary in evidence accumulation, i.e. retaining a number of decision classes with support/belief, which are used to contribute evidence accumulation.

3 Uncertainty Reasoning for Classification

Given the methods and techniques described in the previous sections, now we are ready to examine how they can be applied to uncertainty reasoning for classification.

3.1 Matching Process

In rule-based classification systems, rules can be activated in different ways such as 'M-of-N rule' [9], low and high match degrees for dynamic reasoning [7], etc. In our work, we adopt a probabilistic model developed for information retrieval to determine how a rule can be activated in reasoning process [10]. Given an

Algorithm: generating rules

1 **Input:** $U/d = \{W_1, \cdots, W_K\}$ a set of partitions
with respect to decision attribute d;
X a reduct; $BN_X(W_i)$ a lower approximation
2 **Output**: R_i a rule
3 $S \leftarrow BN_X(W_i)$
4 **for** $(i = 1to|S|)$ **do**
5 $T \leftarrow T \cup d(S[i])$
6 endfor
7 $stg \leftarrow |BN_X(W_i)|/|U|$
8 $R_i \leftarrow$ IF
9 **for** each $(x \in X)$
10 $R_i \leftarrow R_i \wedge (x/x(S[i]))$
11 $R_i \leftarrow R_i + THENT|stg$
12 **endfor**
13 **output** R_i
14 **end**

Fig. 8. Constructing rules from boundary

$r_{11} : IF(h/N) \wedge (t/H)THENd/Y|\frac{|\{u_1\}|}{|U|} = \frac{1}{6}$
$r_{12} : IF(h/Y) \wedge (t/VH)THENd/Y|\frac{|\{u_3\}|}{|U|} = \frac{1}{6}$
$r_{13} : IF(h/N) \wedge (t/VH)THENd/Y|\frac{|\{u_6\}|}{|U|} = \frac{1}{6}$
$r_{14} : IF(h/N) \wedge (t/N)THENd/N|\frac{|\{u_4\}|}{|U|} = \frac{1}{6}$
$r_{15} : IF(h/Y) \wedge (t/h)THEN\{Y, N\}|\frac{|\{u_2,u_5\}|}{|U|} = \frac{2}{6}$

Fig. 9. Rules derived by using reduct $\{h, t\}$

$r_{21} : IF(m/Y) \wedge (t/H)THENd/Y|\frac{|\{u_1\}|}{|U|} = \frac{1}{6}$
$r_{22} : IF(m/Y) \wedge (t/VH)THENd/Y|\frac{|\{u_3,u_6\}|}{|U|} = \frac{2}{6}$
$r_{23} : IF(m/Y) \wedge (t/VH)THENd/N|\frac{|\{u_4\}|}{|U|} = \frac{1}{6}$
$r_{24} : IF(m/N) \wedge (t/h)THEN\{Y, N\}|\frac{|\{u_2,u_5\}|}{|U|} = \frac{2}{6}$

Fig. 10. Rules derived by using reduct $\{m, t\}$

instance or a piece of evidence e and a set of rules $R = \{r_1, r_2, \cdots, r_{|R|}\}$, determining whether e is matched against the conditions of a rule r_i is simply modelled by a logic implication, denoted by $P(e \rightarrow r_i)$ and the conditions of each rule are denoted by $\{c_{i1}/v_{i1}, c_{i2}/c_{i2}, \cdots, c_{ik}/c_{ik}\}$. By interpreting such a logical implication as a conditional probability, we have $P(e \rightarrow r_i) = P(e|r_i)$, which can be calculated by the following formula:

$$\mu(e) = \frac{|\{e\} \cap \{c_{i1}/v_{i1}, c_{i2}/c_{i2}, \cdots, c_{ik}/c_{ik}\}|}{|\{c_{i1}/v_{i1}, c_{i2}/c_{i2}, \cdots, c_{ik}/c_{ik}\}|} \tag{7}$$

Equation (7) will be used for computing the matching degree of a given instance and a set of rules. When $\mu\{e\} = 1$ means that the evidence is exactly matched to the conditions of a rule, and $\mu\{e\} < 1$ indicates that the evidence is partially matched to the conditions of a rule.

It notices that for a given evidence, more than one rule may be activated, in particular, partial matching is involved in reasoning process, this implies that more than one conclusion may be inferred. Therefore in order to obtain a single conclusion, it is desirable to study an effective method to combine those inferred conclusions, so that a final conclusion can be achieved. Prior to detailing our proposed method, we briefly overview several important concepts of the DS theory of evidence which have been used in our method [11].

3.2 Dempster-Shafer (DS) Theory of Evidence

Definition 5. Let Θ be a finite nonempty set, and call it the *frame of discernment*. Let $[0, 1]$ denote the interval of real numbers from zero to one, inclusive: $[0, 1] = \{x|0 \leq x \leq 1\}$. A function $m : 2^{\Theta} \rightarrow [0, 1]$ is called a *mass function* if it satisfies:

$$1) \;\; m(\emptyset) = 0; \qquad 2) \sum_{X \subseteq \Theta} m(X) = 1.$$

A mass function is a basic probability assignment (*bpa*) to all subsets X of Θ. A subset A of a frame Θ is called a *focal element* of a mass function m over Θ if $m(A) > 0$ and it is called a *singleton* if it contains only one element. The definition applies to propositions as well as subsets.

Definition 6. Let m_1 and m_2 be mass functions on the frame Θ, and for any subsets $A \subseteq \Theta$, the *orthogonal sum* $\oplus$ of two mass functions on A is defined as

$$m(A) = (1/N) \sum_{X \cap Y = A} m_1(X) m_2(Y) \tag{8}$$

where $N = 1 - \sum_{X \cap Y \neq \emptyset} m_1(X) m_2(Y)$ and $K = 1/N$ is called the normalization constant of the orthogonal sum $m_1 \oplus m_2$. The orthogonal sum is a fundamental operation of evidential reasoning and it is often called Dempster's rule of combination. There are two conditions to ensure the orthogonal sum exists: 1) $N \neq 0$; 2) two mass functions must be independent of each other. We often allocate some mass to undeterministic status by means of *ignorance*.

3.3 Defining Rule Mass Function

Conclusions inferred from different rules cannot be directly combined using Dempster's rule of combination since for an unseen instance, it is a rare case or even impossible that the whole set of rules will be satisfied, in this situation, the sum of the rule strengths associated with these conclusions will not be equal to 1, which does not meet the condition given in Definition 5. To apply Dempster's rule of to combine conclusions, we need to normalize all the rule strengths

attached to the inferred conclusions and ensure that the normalization still fully preserve the information provided with these conclusions.

Formally let $D = \{d_1, d_2, \cdots, d_{|D|}\}$ be a frame of discernment, and let $R = \{r_1, r_2, \cdots, r_{|R|}\}$ be a set of rules, given a new evidence e, if q rules are activated, i.e. $r_{i1}, r_{i2}, \cdots, r_{iq}$, then q conclusions are inferred from R. The inference process can be expressed by $r_{i1}(e) \rightarrow d_{i1}|stg(d_{i1}), r_{i2}(e) \rightarrow d_{i2}|stg(d_{i2}), \cdots, r_{iq}(e) \rightarrow d_{iq}|stg(d_{iq})$. As a result, we have a set of conclusions, denoted by $H = \{d_{i1}, d_{i2}, \cdots, d_{iq}\}$, where $H \in 2^D$. Due to $q < |R|$, $stg(d_{i1}) + stg(d_{i2}) + \cdots + stg(d_{iq}) < 1$, thus H does not meet the condition of applying Dempster's rule as given Definition 5. Now we describe our method for normalizing a set of conclusions below.

First, we remove duplicate conclusions and add up rule strengths, and then construct a new set of the conclusions. Specifically, for any two d_{ij} and d_{ij}, if $d_{ij} = d_{ij}$ and $ii \neq ij$, then $stg(d_{ii}) \leftarrow stg(d_{ii}) + stg(d_{ij})$, and d_{ij} is eliminated. After processing, a frame of conclusions is reconstructed, denoted by $H = \{d_1, d_2, \cdots, d_{|H|}\}$, apparently $H \in 2^D$. Notice that d_i is a singleton decision since it is inferred by using the rules of Type 1 (see Table 4), but it can be a set of decisions when the rules of Type 3 are used. To avoid confusion in accommodating more complicate situations, we denote $H = \{h_1, h_2, \cdots, h_{|H|}\}$, where $h_j = d_{ij}$ or $h_j \subseteq D$. Therefore a rule mass function for H is defined below:

Definition 7. A rule mass function is defined as a mapping $m : H \rightarrow [0, 1]$. There are two situations based on the inclusive relations between H and D.

1) if $D \in H$, then a rule mass function is defined as follows:

$$m(\{h_i\}) = \frac{stg(h_i)}{\sum_{j=1}^{|H|} stg(h_j)} \quad (1 \leq i \leq |H|) \tag{9}$$

2) if $D \notin H$ and $|H| > 2$, then $H \leftarrow H \cup D$ and a rule mass function is defined as follows:

$$m(\{h_i\}) = stg(h_i) \quad (1 \leq i \leq |H| - 1) \tag{10}$$

$$m(\{h_i\}) = 1 - \sum_{j=1}^{|H|-1} stg(h_j) \tag{11}$$

We have provided a proof that rule strengths satisfy the condition of a mass function [12, 13]. However, as in the first case above, some conclusions cannot be individually inferred from a piece of evidence, this means those conclusions remain unknown. Therefore, it is necessary to redistribute rule strengths over the known conclusions in order to meet the conditions given in Definition 5.

The second case means that the added D represents the unknown state of hypotheses in inference processes in terms of *ignorance*. It absorbs the unassigned portion of the belief after committing to H. The addition of ignorance about the likelihood of future hypotheses provides us with the partial information we need for the inference process. This also means that our method does not require complete knowledge about all potential hypotheses since we represent an

implicit set of unmodeled future hypotheses by including an additional unknown state D.

To demonstrate how a threshold can be used for partial matching, and how rule strengths can be calculated, normalized and then combined using Dempster's rule, let us look at the following section.

3.4 An Example

Suppose a new instance $e = \{h/Y, m/Y\}$ and a threshold $\mu(e) = 0.5$, the process of decision making starts with activating the rules as shown in Figures 9 and 10.

$$\mu(\{h/Y\}) = \frac{|\{h/Y\} \cap \{h/Y, t/VH\}|}{|\{h/Y, t/VH\}|} = \frac{|\{h/Y\}}{|\{h/Y, t/VH\}|} = \frac{1}{2}$$

$$\mu(\{h/Y\}) = \frac{|\{h/Y\} \cap \{h/Y, t/VH\}|}{|\{h/Y, t/H\}|} = \frac{|\{h/Y\}}{|\{h/Y, t/H\}|} = \frac{1}{2}$$

$$\mu(\{m/Y\}) = \frac{|\{m/Y\} \cap \{m/Y, t/H\}|}{|\{m/Y, t/H\}|} = \frac{|\{m/Y\}}{|\{m/Y, t/H\}|} = \frac{1}{2}$$

$$\mu(\{m/Y\}) = \frac{|\{m/Y\} \cap \{m/Y, t/VH\}|}{|\{m/Y, t/VH\}|} = \frac{|\{m/Y\}}{|\{m/Y, t/VH\}|} = \frac{1}{2}$$

$$\mu(\{m/Y\}) = \frac{|\{m/Y\} \cap \{m/Y, t/N\}|}{|\{m/Y, t/N\}|} = \frac{|\{m/Y\}}{|\{m/Y, t/N\}|} = \frac{1}{2}$$

By $\mu(e) = 0.5$, two sets of rules are activated: R_1 and R_2, where $R_1 = \{r_{12}, r_{15}\}$ and $R_2 = \{r_{21}, r_{22}, r_{23}\}$. With these rules, we can obtain two groups of conclusions, denoted by $H_1$and H_2, respectively, where $H_1 = \{h_{11}, h_{12}\} = \{yes, \{yes, no\}\}$, and $H_2 = \{h_{21}, h_{22}, h_{23}\} = \{yes, yes, no\}$, performing a normalization operation on them, we can have $H_1 = \{yes, \{yes, no\}\}$ and $H_2 = \{yes, no, \{yes, no\}\}$, and the calculation of rule mass functions are given below.

For mass function m_1 over H_1, we have

$$m_1(\{yes\}) = \frac{stg(\{yes\})}{stg(\{yes\}) + stg(\{no\})} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{2}{6}} = \frac{1}{3}$$

$$m_1(\{yes, no\}) = \frac{stg(\{yes, no\})}{stg(\{yes\}) + stg(\{yes, no\})} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{2}{6}} = \frac{2}{3}$$

For mass function m_2 over H_2, we have

$$m_2(\{yes\}) = stg(\{yes\}) + stg(\{yes\}) = \frac{1}{6} + \frac{2}{6} = \frac{3}{6}$$

$$m_2(\{no\}) = stg(\{no\}) = \frac{1}{6}$$

$$m_2(\{yes, no\}) = 1 - m_2(\{yes\}) - m_2(\{no\}) = 1 - \frac{1}{6} - \frac{3}{6} = \frac{2}{6}$$

By using Dempster's rule to aggregate the rule strengths from two sets of rules, we have $(m_1 \oplus m_2)(\{yes\}) = 0.647$, $(m_1 \oplus m_2)(\{no\}) = 0.118$ and $(m_1 \oplus m_2)(\{yes, no\}) = 0.235$. The final decision made on basis of the combined rule strengths *flu* = *yeas* with confidence 0.647.

This example illustrates how conflict conclusions can be resolved by aggregating these conclusions through Dempster's rule. Although this example is a special case where the matching degrees are equal, the degrees associated with every single rule are different from each other. Thus it is sufficient to warrant the merit of our proposed method.

4 Conclusion

In this paper, we have introduced a rough set-based method for rule induction, including the induction of certain decision rules and uncertain decision rules. In particular, we provide an effective solution to coping with inconsistencies among the decision classes and to how these inconsistent instances are organized into a set of decision classes with a single degree of belief. These solutions underpins the method for making use of multiple pieces of evidence in uncertainty reasoning for classification.

In addition we have developed a measure in terms of rule strengths to represent quantitative relation between condition and decision attributes in a decision system. It has been theoretically proved that rule strengths meet the conditions of mass functions. But in most situations, the rule strengths attached to the conclusions drawn are necessarily normalized so that Dempster's rule of combination can be applied.

We use a simplified example to illustrate the process of decision making with a partial matching mechanism, where inferred conclusions may be in conflict. The significant aspect of our method is that we can achieve a sensible result even if an instance is not exactly matched with the antecedents of rules throughout. The other important aspect is that when more than one rule is activated, these rules can be combined to infer a conclusion, so that the generalization capability may be improved in a classification systems. These two aspects distinguish our method from most of the knowledge-based methods which are conventionally used in classification systems.

References

1. Pawlak, Z.: Rough Set: Theoretical aspects of reasoning about data. Kluwer Academic, Dordrecht (1991)
2. Grzymala-Busse, J.: LERS - A System for Learning from Examples Based on Rough Sets. In: Slowinski, R. (ed.) Intelligent Decision Support, pp. 3–17. Kluwer Academic, Dordrecht (1992)
3. Stefanowski, J.: On Rough Set Based Approaches to Induction of Decision Rules. In: Polkowski, L., Skowron, A. (eds.) Studies in Fuzziness and Soft Computing, vol. 1, pp. 500–529. Physica-Verlag (1998)

4. Cohen, W.W., Singer, Y.: Simple, Fast, and Effective Rule Learner in AAAI 1999 (1999)
5. Guang, J.W., Bell, D.: Rough computational methods for information systems. Artificial Intelligence 105, 77–103 (1998)
6. Pawlak, Z.: Rough Set Elements. In: Polkowski, L., Skowron, A. (eds.) Studies in Fuzziness and Soft Computing, vol. 1, pp. 10–30. Physica-Verlag (1998)
7. Song, S.: A Strategy Of Dynamic Reasoning In Knowledge- Based System With Fuzzy Production Rules. Journal of intelligent information systems 19(3), 303–318 (2002)
8. Chouchoulas, A., Shen, Q.: Rough Set-Aided Keyword Reduction for Text Categorization. Applied Artificial Intelligence (2001)
9. Towell, G., Shavlik, J.: Using Symbolic Learning to Improve Knowledge-based Neural Networks. In: Proceedings of the Tenth National Conference on Artificial Intelligence, pp. 177–182 (1992)
10. van Rijsbergen, C.J.: Information Retrieval, 2nd edn. Butterworths (1979)
11. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
12. Bi, Y., Anderson, T., McClean, S.: Combining Rules for Text Categorization Using Dempster's Rule of Combination. In: Yang, Z.R., Yin, H., Everson, R.M. (eds.) IDEAL 2004. LNCS, vol. 3177. Springer, Heidelberg (2004)
13. Bi, Y.: Combining Multiple Classifiers for Text Categorization using Dempster's rule of combination. PhD thesis, University of Ulster (2004)

Semantics of the Relative Belief of Singletons

Fabio Cuzzolin

INRIA Rhône-Alpes
655 avenue de l'Europe, 38334 SAINT ISMIER CEDEX, France
Fabio.Cuzzolin@inrialpes.fr

Summary. In this paper we introduce the relative belief of singletons as a novel Bayesian approximation of a belief function. We discuss its nature in terms of degrees of belief under several different angles, and its applicability to different classes of belief functions.

1 Introduction: A New Bayesian Approximation

The theory of evidence (ToE) [16] extends classical probability theory through the notion of *belief function* (b.f.), a mathematical entity which independently assigns probability values to *sets* of possibilities rather than single events. A belief function $b : 2^\Theta \rightarrow [0,1]$ on a finite set ("frame") Θ has the form $b(A) = \sum_{B \subseteq A} m_b(B)$ where $m_b : 2^\Theta \rightarrow [0,1]$, is called "basic probability assignment" (b.p.a.), and meets the positivity $m_b(A) \geq 0 \; \forall A \subseteq \Theta$ and normalization $\sum_{A \subseteq \Theta} m_b(A) = 1$ axioms. Events associated with non-zero basic probabilities are called "focal elements". A b.p.a. can be uniquely recovered from a belief function through Moebius inversion:

$$m_b(A) = \sum_{B \subseteq A} (-1)^{|A-B|} b(B). \tag{1}$$

1.1 Previous Work on Bayesian Approximation

As probability measures or *Bayesian* belief functions are just a special class of b.f.s (for which $m(A) = 0$ when $|A| > 1$), the relation between beliefs and probabilities plays a major role in the theory of evidence [9, 14, 23, 11, 12, 13, 2]. Tessem [21], for instance, incorporated only the highest-valued focal elements in his m_{klx} approximation; a similar approach inspired the summarization technique formulated by Lowrance *et al.* [15]. In Smets' "Transferable Belief Model" [17] beliefs are represented at credal level (as convex sets of probabilities), while decisions are made by resorting to a Bayesian belief function called pignistic transformation [19]. More recently, two new Bayesian approximations of belief functions have been derived from purely geometric considerations [7] in the context of the geometric approach to the ToE [6], in which belief and probability measures are represented as points of a Cartesian space.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 201–213, 2008.
springerlink.com

Another classical approximation is based on the plausibility function (pl.f.) $pl_b : 2^\Theta \to [0,1]$, where

$$pl_b(A) \doteq 1 - b(A^c) = \sum_{B \cap A \neq \emptyset} m_b(B)$$

represent of the evidence not against a proposition A. Voorbraak [22] proposed indeed to adopt the so-called *relative plausibility of singletons* (rel.plaus.) $\tilde{pl}_b$ as the unique probability that, given a belief function b with plausibility pl_b, assigns to each singleton $x \in \Theta$ its normalized plausibility (2). He proved that $\tilde{pl}_b$ is a perfect representative of b when combined with other probabilities p through Dempster's rule $\oplus$ [10],

$$\tilde{pl}_b(x) = \frac{pl_b(x)}{\sum_{y \in \Theta} pl_b(y)}, \qquad \tilde{pl}_b \oplus p = b \oplus p. \tag{2}$$

The properties of the relative plausibility of singletons have been later discussed by Cobb and Shenoy [3].

1.2 Relative Belief of Singletons

In this paper we introduce indeed a Bayesian approximation which is the *dual* of relative plausibility of singletons (2), as it is obtained by normalizing the *belief* (instead of plausibility) values of singletons:

$$\tilde{b}(x) \doteq \frac{b(x)}{\sum_{y \in \Theta} b(y)} = \frac{m_b(x)}{\sum_{y \in \Theta} m_b(y)}. \tag{3}$$

We call it *relative belief of singletons* $\tilde{b}$ (rel.bel.). Clearly $\tilde{b}$ exists iff b assigns some mass to singletons:

$$\sum_{x \in \Theta} m_b(x) \neq 0. \tag{4}$$

As it has been recently proven [4], both relative plausibility and belief of singletons commute with respect to Dempster's orthogonal sum, and $\tilde{b}$ meets the dual of Voorbraak's representation theorem (2).

Proposition 1. *The relative belief operator commutes with respect to Dempster's combination of plausibility functions, namely*

$$\tilde{b}[pl_1 \oplus pl_2] = \tilde{b}[pl_1] \oplus \tilde{b}[pl_2].$$

The relative belief of singletons $\tilde{b}$ represents perfectly the corresponding plausibility function pl_b when combined with any probability through (extended) Dempster's rule:

$$\tilde{b} \oplus p = pl_b \oplus p$$

for each Bayesian belief function $p \in \mathcal{P}$.

Moreover, $\tilde{b}$ meets a number of properties with respect to Dempster's rule which mirror the set of results proven by Cobb and Shenoy for the relative plausibility of singletons [3].

Proposition 2. *If pl_b is idempotent with respect to Dempster's rule, i.e. $pl_b \oplus pl_b = pl_b$, then $\tilde{b}[pl_b]$ is itself idempotent: $\tilde{b}[pl_b] \oplus \tilde{b}[pl_b] = \tilde{b}[pl_b]$. If $\exists x \in \Theta$ such that $b(x) > b(y)\ \forall y \neq x, y \in \Theta$, then $\tilde{b}[pl_b^\infty](x) = 1$, $\tilde{b}[pl_b^\infty](y) = 0\ \forall y \neq x$, where pl_b^∞ denotes the infinite limit of the combination of pl_b with itself.*

In this paper we focus instead on the *semantics* of rel.bel. in a comparative study with that of rel.plaus., in order to understand its meaning in terms of degrees of belief, the way it attributes a mass to singletons, the conditions under which it exists, and to which classes of belief function it can be applied.

1.3 Outline of the Paper

First (Section 2), we argue that rel.bel. gives a conservative estimate of the support b give to each singleton $x \in \Theta$, in opposition to the optimistic estimate provided by the relative plausibility of singletons. Interestingly (Section 3) the relative belief $\tilde{b}$ can indeed be interpreted as the relative plausibility of singletons *of the associated plausibility function*. In order to prove that, we need to extend the evidential formalism to functions whose Moebius inverse is not necessarily positive or *pseudo belief functions* (Section 3.1). Those two Bayesian approximations form then a couple which, besides having dual properties with respect to Dempster's sum, have dual semantics in terms of mass assignment.

In Section 4 we analyze the issue posed by the existence constraint (4), i.e. the fact that rel.bel. exists only when b assigns some mass to singletons. We will argue that situations in which the latter is not met are pathological, as all Bayesian approximations are forced to span a limited region of the probability simplex. Finally, we will prove that, as all those approximations converge for quasi-Bayesian b.f.s, rel.bel. can be seen as a low-cost proxy to pignistic transformation and relative plausibility, and discuss the applicability of $\tilde{b}$ to some important classes of b.f.s in order to shed more light on interpretation and application range of this Bayesian approximation.

2 A Conservative Estimate

A first insight on the meaning of $\tilde{b}$ comes from the original semantics of belief functions as constraints on the actual allocation of mass of an underlying unknown probability distribution. Accordingly, a focal element A with mass $m_b(A)$ indicates that this mass can "float" around in A and be distributed arbitrarily between the elements of A. In this framework, the relative plausibility of singletons $\tilde{pl}_b$ (2) can be interpreted as follows:

- for each singleton $x \in \Theta$ the most *optimistic* hypothesis in which the mass of all $A \supseteq \{x\}$ focuses on x is considered, yielding $\{pl_b(x), x \in \Theta\}$;

- this assumption, however, is contradictory as it is supposed to hold for all singletons (many of which belong to the same higher-size events);
- nevertheless, the obtained values are normalized to yield a Bayesian belief function.

$\tilde{pl}_b$ is associated with the less conservative (but incoherent) scenario in which all the mass that can be assigned to a singleton is actually assigned to it.

The relative belief of singletons (3) can then be naturally given the following interpretation in terms of mass assignments:

- for each singleton $x \in \Theta$ the most *pessimistic* hypothesis in which only the mass of $\{x\}$ itself actually focuses on x is considered, yielding $\{b(x) = m_b(x), x \in \Theta\}$;
- this assumption is also contradictory, as the mass of all higher-size events is not assigned to any singletons;
- the obtained values are again normalized to produce a Bayesian belief function.

Dually, $\tilde{b}$ reflects the most conservative (but still not coherent) choice of assigning to x only the mass that the b.f. b (seen as a constraint) assures it belong to x. The underlying mechanism, though, is exactly the same as the one supporting the rel.plaus. function.

3 Dual Interpretation as Relative Plausibility of a Plausibility

A different aspect of rel.bel. emerges when considering the dual representation of the evidence carried by b expressed by the plausibility function pl_b. We first need though to introduce the notion of "pseudo belief function".

3.1 Pseudo Belief Functions

A belief function is a function on 2^Θ whose Moebius inverse m_b (the basic probability assignment) meets the positivity axiom: $m_b(A) \geq 0 \; \forall A \subseteq \Theta$.

However, all functions $\varsigma : 2^\Theta \to \mathbb{R}$ admit Moebius inverse (11) $m_\varsigma : 2^\Theta \setminus \emptyset \to \mathbb{R}$ such that

$$\varsigma(A) = \sum_{B \subseteq A} m_\varsigma(B)$$

where $m_\varsigma(B) \not\geq 0 \; \forall B \subseteq \Theta$ [1].

Functions ς whose Moebius inverse meets the normalization constraint

$$\sum_{\emptyset \subsetneq A \subseteq \Theta} m_\varsigma(A) = 1$$

are then natural extensions of belief functions[1], and are called *pseudo belief functions* (p.b.f.s) [20].

[1] Geometrically, each p.b.f. can be thought of as a vector ς of $\mathbb{R}^N$, $N = 2^{|\Theta|} - 1$, while belief functions form a *simplex* in the same space [6].

As they meet the normalization constraint ($pl_b(\Theta) = 1$ for all b), plausibility functions are themselves pseudo belief functions. Their Moebius inverse [8]

$$\mu_b(A) \doteq \sum_{B \subseteq A} (-1)^{|A \setminus B|} pl_b(B) = (-1)^{|A|+1} \sum_{B \supseteq A} m_b(B), \quad A \neq \emptyset \tag{5}$$

is called *basic plausibility assignment* (b.pl.a.), with $\mu_b(\emptyset) = 0$.

3.2 Duality between Relative Belief and Plausibility

A useful property of μ_b is that

Theorem 1. $m_b(x) = \sum_{A \supseteq \{x\}} \mu_b(A)$.

Proof. By definition of b.pl.a. μ we have that

$$\sum_{A \supseteq \{x\}} \mu_b(A) = \sum_{A \supseteq \{x\}} (-1)^{|A|+1} \Big(\sum_{B \supseteq A} m_b(B) \Big) =$$
$$= - \sum_{B \supseteq \{x\}} m_b(B) \Big(\sum_{x \subseteq A \subseteq B} (-1)^{|A|} \Big)$$

where $\sum_{x \subseteq A \subseteq B} (-1)^{|A|} = 0$ if $B \neq \{x\}$, -1 if $B = \{x\}$ for Newton's binomial: $\sum_{k=0}^{n} 1^{n-k} (-1)^k = 0$.

If we write the plausibility of singletons as

$$pl_b(x) = \sum_{A \supseteq \{x\}} m_b(A)$$

we realize that Theorem 1 states that the belief of singletons $b(x)$ is nothing but *the plausibility of singletons of pl_b interpreted as a pseudo belief function*: $b(x) = pl_{pl_b}(x)$. Formally,

- there exists a class of pseudo b.f.s which correspond to the plausibility of some b.f. b: $\varsigma = pl_b$ for some $b \in \mathcal{B}$;
- each p.b.f. admits a (pseudo) plausibility function, analogous to the case of standard b.f.s: $pl_\varsigma(A) = \sum_{B \cap A \neq \emptyset} m_\varsigma(B)$;
- but for the above class of p.b.f. $\varsigma = pl_b$, so that the above equation reads as $pl_{pl_b}(A) = \sum_{B \cap A \neq \emptyset} \mu_b(B)$ (as μ_b is the Moebius inverse of pl_b);
- when applied to singletons this yields

$$pl_{pl_b}(x) = \sum_{B \ni x} \mu_b(B) = m_b(x) \tag{6}$$

  by Theorem 1, which implies $\tilde{pl}_{pl_b} = \tilde{b}$.

It is a bit paradoxical to point out that, as the basic plausibility assignment μ_b carries the same information as the basic probability assignment m_b, according to Equation (6) *all* the information carried by b is used to compute the relative belief of singletons, while its definition (3) seems to suggest that most of this information is discarded in the process.

4 On the Existence Constraint

The relative belief of singletons exists only for those belief function such that $\sum_x b(x) \neq 0$.

As $\tilde{b}$ is the relative plausibility of $\varsigma = pl_b$ (Section 3), and as relative plausibilities do not undergo any existence constraint (as $\sum_x pl_b(x) \neq 0$), one could argue that $\tilde{b}$ should always exist. However, the symmetry is broken by the fact that the b.pl.a. μ_b does not meet the non-negativity constraint ($\mu_b \not\geq 0$), and as a consequence $pl_{pl_b}(x)$ can actually be zero $\forall x \in \Theta$.

4.1 Example: The Binary Case

In the binary case $\Theta = \{x, y\}$, for instance, according to (4) the only b.f. which does not admit rel.bel. is the vacuous one b_Θ: $m_{b_\Theta}(\Theta) = 1$. For b_Θ, $m_{b_\Theta}(x) = m_{b_\Theta}(y) = 0$ so that $\sum_x m_{b_\Theta}(x) = 0$ and $\tilde{b}_\Theta$ does not exist. Symmetrically, the pseudo b.f. $\varsigma = pl_{b_\Theta}$ (for which $pl_{b_\Theta}(x) = pl_{b_\Theta}(y) = 1$) is such that $pl_{pl_{b_\Theta}} = b_\Theta$, so that $\tilde{pl}_{pl_{b_\Theta}}$ does not exist. In the binary frame each belief function is completely

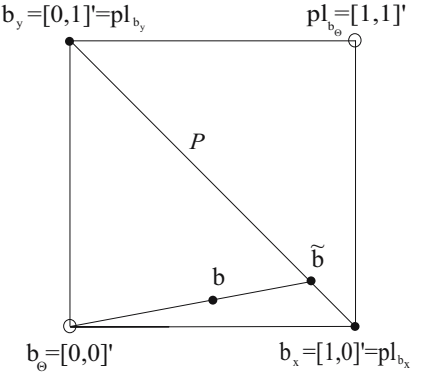


Fig. 1. B.f.s $b = [m_b(x), m_b(y)]'$ and pl.f.s $pl_b = [pl_b(x) = 1 - m_b(y), pl_b(y) = 1 - m_b(x)]'$ on $\Theta = \{x, y\}$ can be represented as points of $\mathbb{R}^2$ [6]. The locations of $\tilde{b} = [\frac{m_b(x)}{m_b(x)+m_b(y)}, \frac{m_b(y)}{m_b(x)+m_b(y)}]'$ and the singular points $b_\Theta = [0, 0]'$ and $pl_{b_\Theta} = [1, 1]'$ are shown.

determined by its belief values $b(x), b(y)$ (as $b(\emptyset) = 0$, $b(\Theta) = 1$ for all b) and can then be represented as a point of the plane $\mathbb{R}^2$:

$$b = [b(x), b(y)]'.$$

Figure 1 illustrates then the location of $\tilde{b}$ in the simple binary case and those of the dual singular points b_Θ, $\varsigma = pl_{b_\Theta}$.

4.2 Region Spanned by a Bayesian Approximation

One can argue that the existence of rel.bel. is subject to quite a strong condition (4). We can claim though that situations in which the constraint is not met are indeed rather pathological, in a very precise way.

To show this, let us compute the region spanned by the most common Bayesian approximations: rel.plaus. (2) and *pignistic function* [19]

$$BetP[b](x) \doteq \sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|}.$$

All Bayesian approximations can be seen as operators mapping belief functions to probabilities:

$$\begin{array}{rcl} \tilde{pl} : \mathcal{B} \to & \mathcal{P} & \qquad BetP : \mathcal{B} \to \quad \mathcal{P} \\ b \mapsto \tilde{pl}[b] = \tilde{pl}_b & & \qquad\qquad b \mapsto BetP[b] \end{array} \tag{7}$$

where $\mathcal{B}$ and $\mathcal{P}$ denote the set of all b.f.s and probability functions respectively. Now, it is well known [7] that the pignistic transformation (7)-right commutes with affine combination:

$$BetP\Big[\sum_i \alpha_i b_i\Big] = \sum_i \alpha_i BetP[b_i], \qquad \sum_i \alpha_i = 1. \tag{8}$$

If we then denote by Cl the convex closure operator

$$Cl(b_1, ..., b_k) = \left\{ b \in \mathcal{B} : b = \alpha_1 b_1 + \cdots + \alpha_k b_k, \sum_i \alpha_i = 1,\ \alpha_i \geq 0\ \forall i \right\} \tag{9}$$

(8) implies that $BetP$ commutes with Cl:

$$BetP[Cl(b_1, ..., b_k)] = Cl(BetP[b_i], i = 1, ..., k).$$

In the case of $\tilde{pl}_b$, even though the latter does not commute with affine combination (the relation being somehow more complex [5]) we can still prove that it commutes with convex closure (9).

Using this tools we can find the region of the probability simplex $\mathcal{P}$ spanned by the Bayesian transformation of a certain convex region $Cl(b_1, ..., b_k)$ of b.f.s. It suffices to compute in both cases the approximations of the vertices of the considered region.

4.3 Zero Mass to Singletons as a Pathological Situation

But the space of all belief functions $\mathcal{B} \doteq \{b : 2^\Theta \to [0, 1]\}$ defined on a frame Θ is indeed the convex closure [6]

$$\mathcal{B} = Cl(b_A, A \subseteq \Theta) \tag{10}$$

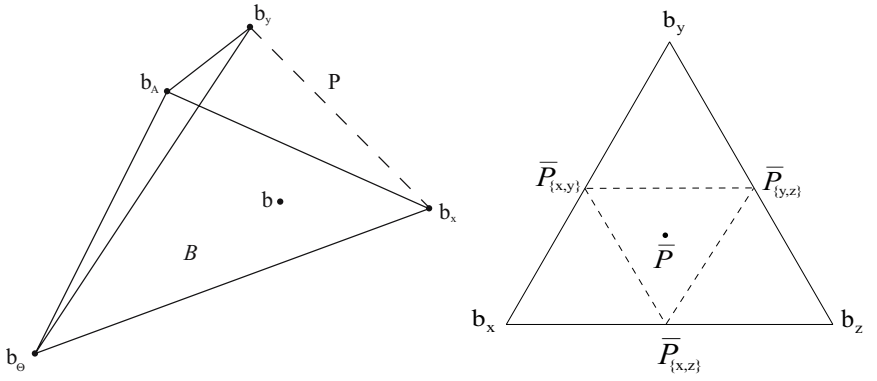


Fig. 2. Left: The space of all belief functions on a domain Θ is a polytope or "simplex" (10) in $\mathbb{R}^N$. The probability simplex $\mathcal{P}$ is a face of $\mathcal{B}$. Right: For the class of b.f.s $\{b : \sum_x m_b(x) = 0\}$, pignistic function and relative plausibility are allowed to span only a proper subset of the probability simplex (delimited by dashed lines in the ternary case $\Theta = \{x, y, z\}$). Otherwise $\tilde{b}$, $BetP[b]$, $\tilde{pl}_b$ can be located in any point of $\mathcal{P}$ for some values of b.

of all *basis* belief functions

$$b_A \doteq b \in \mathcal{B} \; s.t. \; m_b(A) = 1, \; m_b(B) = 0 \; \forall B \neq A \tag{11}$$

i.e. the belief functions focusing on a single event $A \subseteq \Theta$. Geometrically, they are the vertices of the polytope $\mathcal{B}$ of all belief functions (Figure 2-left).

The images of a basis b.f. b_A under the transformations (7) are

$$\tilde{pl}_{b_A}(x) = \frac{\sum_{B \supseteq \{x\}} m_{b_A}(B)}{\sum_{B \supseteq \{x\}} m_{b_A}(B)|B|} = \begin{cases} \frac{1}{|A|} & x \in A \\ 0 & else \end{cases} \doteq \overline{\mathcal{P}}_A$$
$$BetP[b_A](x) = \sum_{B \supseteq \{x\}} \frac{m_{b_A}(B)}{|B|} = \overline{\mathcal{P}}_A$$

so that

$$BetP[\mathcal{B}] = BetP[Cl(b_A, A \subseteq \Theta)] = Cl(BetP[b_A], A \subseteq \Theta) = \\ = Cl(\overline{\mathcal{P}}_A, A \subseteq \Theta) = \mathcal{P} = \tilde{pl}[\mathcal{B}].$$

In normal conditions *the whole* probability simplex $\mathcal{P}$ can host such approximations. On the other side, as they have the form $b = \sum_{|A|>1} m_b(A) b_A$ with $m_b(A) \geq 0$, $\sum_{|A|>1} m_b(A) = 1$, the set of (singular) b.f.s *not* meeting the constraint (4) is $Cl(b_A, |A| > 1)$ so that the region of $\mathcal{P}$ spanned by their Bayesian approximations is

$$\tilde{pl}[Cl(b_A, |A| > 1)] = Cl(\tilde{pl}_{b_A}, |A| > 1) = Cl(\overline{\mathcal{P}}_A, |A| > 1) = \\ = Cl(BetP[b_A], |A| > 1) = BetP[Cl(b_A, |A| > 1)].$$

The result is illustrated by Figure 2-right in the ternary case. If (4) is not met, all Bayesian approximations of b can span only a limited region

$$Cl(\overline{\mathcal{P}}_{\{x,y\}}, \overline{\mathcal{P}}_{\{x,z\}}, \overline{\mathcal{P}}_{\{y,z\}}, \overline{\mathcal{P}}_{\Theta}) = Cl(\overline{\mathcal{P}}_{\{x,y\}}, \overline{\mathcal{P}}_{\{x,z\}}, \overline{\mathcal{P}}_{\{y,z\}})$$

of the probability simplex (delimited by dashed lines).

The case in which $\tilde{b}$ does not exist is indeed pathological, as it excludes a great deal of belief and probability measures.

5 A Low-Cost Proxy for other Bayesian Approximations

A different angle on the utility of $\tilde{b}$ comes from a discussion of what classes of b.f.s are "suitable" to be approximated by means of (3). As it only makes use of the masses of singletons, working with $\tilde{b}$ requires storing n values to represent a belief function. As a consequence, the computational cost of combining new evidence through Dempster's rule or disjunctive combination [18] is reduced to $O(n)$ as only the mass of singletons has to be calculated.

When the actual values of $\tilde{b}(x)$ are close to those provided by, for instance, pignistic function or rel.plaus. is then more convenient to resort to the relative belief transformation.

5.1 Convergence under Quasi-bayesianity

A formal support to this argument is provided by the following result. Let us call *quasi-Bayesian* b.f.s the belief functions b for which the mass assigned to singletons is very close to one:

$$k_{m_b} \doteq \sum_{x \in \Theta} m_b(x) \to 1.$$

Theorem 2. *For quasi-Bayesian b.f.s all Bayesian approximations converge:*

$$\lim_{k_{m_b} \to 1} BetP[b] = \lim_{k_{m_b} \to 1} \tilde{pl}_b = \lim_{k_{m_b} \to 1} \tilde{b}.$$

Proof. If $k_{m_b} \to 1$ then $\sum_{|A|>1} m_b(A) \to 0$ which implies

$$m_b(A) \to 0 \quad \forall A : |A| > 1$$

(as $m_b(A) \geq 0 \; \forall A$). But by definition of $BetP$, $\tilde{b}$ $\tilde{pl}$, we have that

$$BetP[b](x) \doteq m_b(x) + \sum_{A \supsetneq \{x\}} m_b(A) \to m_b(x),$$

$$\tilde{b}(x) \doteq \frac{m_b(x)}{k_{m_b}} \to m_b(x),$$

$$\tilde{pl}_b(x) \doteq \frac{pl_b(x)}{\sum_x pl_b(x)} = \frac{m_b(x) + \sum_{A \supsetneq \{x\}} m_b(A)}{\sum_x (m_b(x) + \sum_{A \supsetneq \{x\}} m_b(A))} \to \frac{m_b(x)}{k_{m_b}} = m_b(x).$$

Theorem 2 highlights then the convenience of computing rel.bel. instead of other Bayesian approximations for quasi-Bayesian b.f.s defined on a large frame of discernment.

5.2 Convergence in the Ternary Case

Let us consider for instance the ternary case $\Theta = \{x, y, z\}$ in which

$$\begin{aligned}
\tilde{b}(x) &= \frac{m_b(x)}{m_b(x) + m_b(y) + m_b(z)}, \\
\tilde{pl}_b(x) &= \frac{(m_b(x) + m_b(\{x, y\}) + m_b(\{x, z\}) + m_b(\Theta))}{\sum_{w \in \Theta} pl_b(w)}, \\
BetP[b](x) &= m_b(x) + \frac{m_b(\{x, y\}) + m_b(\{x, z\})}{2} + \frac{m_b(\Theta)}{3}.
\end{aligned} \tag{12}$$

with $\sum_{w \in \Theta} pl_b(w) = (m_b(x) + m_b(y) + m_b(z)) + 2(m_b(\{x, y\}) + m_b(\{x, z\}) + m_b(\{y, z\})) + 3m_b(\Theta)$.

According to Theorem 2, if $k_{m_b} = \sum_w m_b(w) \to 1$ all the quantities (12) converge to $m_b(x)$. It is interesting to assess the velocity of this convergence with respect to the parameter k_{m_b}. For sake of comparison we consider two different mass allocations to higher-order events: a) $m_b(\Theta) = 1 - k_{m_b}$ and: b) $m_b(\{x, y\}) = 1 - k_{m_b}$. The above expressions (12) then yield

$$\tilde{b}(w) = \frac{m_b(w)}{k_{m_b}} \quad \forall w \in \{x, y, z\};$$

$$\tilde{pl}_b(w) = \begin{cases} \frac{m_b(w)+1-k_{m_b}}{k_{m_b}+3(1-k_{m_b})} \ \forall w \in \{x, y, z\} & m_b(\Theta) = 1 - k_{m_b}, \\ \frac{m_b(w)+1-k_{m_b}}{k_{m_b}+2(1-k_{m_b})} \ \forall w \in \{x, y\} & m_b(\{x, y\}) = 1 - k_{m_b}, \\ \frac{m_b(w)}{k_{m_b}+2(1-k_{m_b})} \ w = z & m_b(\{x, y\}) = 1 - k_{m_b}; \end{cases}$$

$$BetP[b](w) = \begin{cases} m_b(w) + \frac{1-k_{m_b}}{3} \ \forall w \in \{x, y, z\} & m_b(\Theta) = 1 - k_{m_b}, \\ m_b(w) + \frac{1-k_{m_b}}{2} \ \forall w \in \{x, y\} & m_b(\{x, y\}) = 1 - k_{m_b}, \\ m_b(w) \quad w = z & m_b(\{x, y\}) = 1 - k_{m_b}. \end{cases}$$

We evaluated the above expressions for $m_b(x) = k_{m_b}/3$, $m_b(y) = k_{m_b}/2$, $m_b(z) = k_{m_b}/6$ in order to maintain the same relative belief of singletons

$$\tilde{b}(x) = 1/3, \qquad \tilde{b}(y) = 1/2, \qquad \tilde{b}(z) = 1/6$$

in all three cases. Figure 3-top-left plots the L_2 distances in the probability simplex $\mathcal{P} = Cl(b_x, b_y, b_z)$

$$d(p, p') \doteq \sqrt{\sum_{w \in \Theta} |p(w) - p'(w)|^2}$$

of $BetP[b]$ and $\tilde{pl}_b$ from $\tilde{b}$ as a function of k_{m_b}, in both cases a) and b). As stated by Theorem 2, for quasi-Bayesian b.f.s ($k_{m_b} \to 1$) all approximations are the same. It is interesting to notice, however, that for the pignistic function the rate of convergence to $\tilde{b}$ is the same no matter how the mass is assigned to higher-size events, and is *constant*.

For $\tilde{pl}_b$, instead, the rate of convergence differs in the two cases and is actually slower for *discounted* belief functions, i.e. b.f.s which assign all the mass of non-singletons to the whole frame Θ (case b), a rather counterintuitive result.

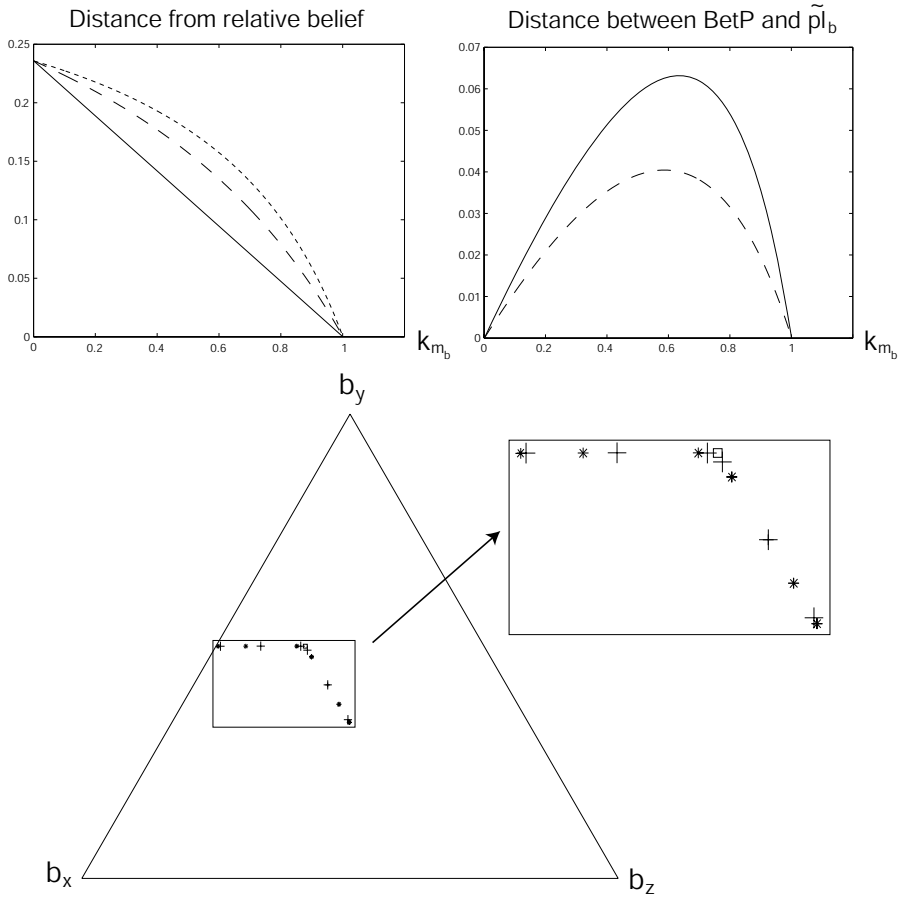


Fig. 3. Convergence of pignistic function and relative plausibility to the relative belief in the ternary frame $\Theta = \{x, y, z\}$. Top left: distance from $\tilde{b}$ of $BetP[b]$ (solid line) and $\tilde{pl}_b$ (dotted line: case a; dashed line: case b) as a function of k_{m_b}. Top right: Corresponding distance between $BetP$ and $\tilde{pl}_b$ (solid line: case a; dashed line: case b). Bottom: Sample locations in the probability simplex of $\tilde{b}$ (square), BetP[b] (stars) and $\tilde{pl}_b$ (crosses) for $k_{m_b} = 0.95$, $k_{m_b} = 0.5$, $k_{m_b} = 0.05$ in both case a) (towards the side b_x, b_y of the simplex) and b) (towards the barycenter of the simplex).

Figure 3-top-right plots by comparison the distance *between* $BetP[b]$ and $\tilde{pl}_b$ as a function of k_{m_b}, in the two cases (again: a - solid line, b - dashed line). The two Bayesian approximations turn out to be close for low values of k_{m_b} too (almost singular b.f.s) and their distance reaches a peak for intermediate values of the total mass of singletons. Such values are though different for the two functions, and the divergence is reduced in the case of asymmetric mass assignment ($m_b(\{x, y\}) = 1 - k_{m_b}$).

Finally, Figure 3-bottom shows the location of all considered Bayesian approximations on the probability simplex in both cases (a and b) for the three sample values $k_{m_b} = 0.95$, $k_{m_b} = 0.5$, $k_{m_b} = 0.05$ of the total mass of singletons.

6 Conclusions

In this paper we discussed interpretations and applicability of the relative belief of singletons as a novel Bayesian approximation of a belief function. It has recently been proven that relative belief and plausibility of singletons form a distinct family of Bayesian approximations related to Dempster's rule, as they both commute with $\oplus$, and meet dual representation and idempotence properties [4]. Here we focused in particular on the semantics of rel.bel. On one side we stressed the analogy between the mechanisms generating relative belief and plausibility, pointing out that they correspond to antithetical estimates of the evidence supporting each singleton. We proved that $\tilde{b}$ is in fact equivalent to the relative plausibility of a plausibility (seen as a pseudo belief function), but that this symmetry is broken by the existence constraint acting on $\tilde{b}$. We argued though that situations in which the latter is not met are pathological, as all Bayesian approximations are forced to span a limited region of the probability simplex. Finally, we proved that, as all those approximations converge for quasi-Bayesian b.f.s, rel.bel. can be seen as a low-cost proxy to pignistic transformation and relative plausibility. The analysis of this convergence for different classes of b.f.s has provided us with some insight on the relation between the probabilities associated with a belief function.

References

1. Aigner, M.: Combinatorial theory. Classics in Mathematics. Springer, New York (1979)
2. Bauer, M.: Approximation algorithms and decision making in the Dempster-Shafer theory of evidence–an empirical study. International Journal of Approximate Reasoning 17, 217–237 (1997)
3. Cobb, B.R., Shenoy, P.P.: A comparison of Bayesian and belief function reasoning. Information Systems Frontiers 5(4), 345–358 (2003)
4. Cuzzolin, F.: Dual properties of relative belief of singletons, IEEE Trans. Fuzzy Systems (submitted, 2007)
5. Cuzzolin, F.: Relative plausibility, affine combination, and Dempster's rule, Tech. report, INRIA Rhone-Alpes (2007)
6. Cuzzolin, F.: A geometric approach to the theory of evidence. IEEE Transactions on Systems, Man and Cybernetics part C (to appear, 2007)
7. Cuzzolin, F.: Two new Bayesian approximations of belief functions based on convex geometry. IEEE Transactions on Systems, Man, and Cybernetics - Part B 37(4) (August 2007)
8. Cuzzolin, F.: Geometry of upper probabilities. In: Proceedings of the 3rd Internation Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2003) (July 2003)

9. Daniel, M.: On transformations of belief functions to probabilities. International Journal of Intelligent Systems, special issue on Uncertainty Processing 21(3), 261–282 (2006)
10. Dempster, A.P.: Upper and lower probabilities generated by a random closed interval. Annals of Mathematical Statistics 39, 957–966 (1968)
11. Denoeux, T.: Inner and outer approximation of belief structures using a hierarchical clustering approach. Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(4), 437–460 (2001)
12. Denoeux, T., Ben Yaghlane, A.: Approximating the combination of belief functions using the Fast Moebius Transform in a coarsened frame. International Journal of Approximate Reasoning 31(1–2), 77–101 (2002)
13. Haenni, R., Lehmann, N.: Resource bounded and anytime approximation of belief function computations. International Journal of Approximate Reasoning 31(1–2), 103–154 (2002)
14. Kramosil, I.: Approximations of believeability functions under incomplete identification of sets of compatible states. Kybernetika 31, 425–450 (1995)
15. Lowrance, J.D., Garvey, T.D., Strat, T.M.: A framework for evidential-reasoning systems. In: Proceedings of the National Conference on Artificial Intelligence (American Association for Artificial Intelligence, ed.), pp. 896–903 (1986)
16. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
17. Smets, P.: Belief functions versus probability functions. In: Saitta, L., Bouchon, B., Yager, R. (eds.) Uncertainty and Intelligent Systems, pp. 17–24. Springer, Berlin (1988)
18. Smets, P.: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate reasoning 9, 1–35 (1993)
19. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. International Journal of Approximate Reasoning 38(2), 133–147 (2005)
20. Smets, P.: The canonical decomposition of a weighted belief. In: Proceedings of the International Joint Conference on AI, IJCAI 1995, Montréal, Canada, pp. 1896–1901 (1995)
21. Tessem, B.: Approximations for efficient computation in the theory of evidence. Artificial Intelligence 61(2), 315–329 (1993)
22. Voorbraak, F.: A computationally efficient approximation of Dempster-Shafer theory. International Journal on Man-Machine Studies 30, 525–536 (1989)
23. Ben Yaghlane, A., Denoeux, T., Mellouli, K.: Coarsening approximations of belief functions. In: Benferhat, S., Besnard, P. (eds.) ECSQARU 2001. LNCS (LNAI), vol. 2143, pp. 362–373. Springer, Heidelberg (2001)

A Lattice-Theoretic Interpretation of Independence of Frames

Fabio Cuzzolin

INRIA Rhône-Alpes, 655 avenue de l'Europe, 38334 SAINT ISMIER CEDEX, France
Fabio.Cuzzolin@inrialpes.fr

Summary. In this paper we discuss the nature of independence of sources in the theory of evidence from an algebraic point of view, starting from the analogy with the case of projective geometries. Independence in Dempster's rule is equivalent to independence of frames as Boolean algebras. Collection of frames, though, can be interpreted as semi-modular lattices on which independence can be defined in several different forms. We prove those forms to be distinct but related to Boolean independence, as a step towards a more general definition of this fundamental notion.

1 Introduction

The theory of evidence was born as a contribution to a mathematically rigorous description of subjective probability, where different observers (or "experts") of the same phenomenon possess in general different notions of what the decision space is. Mathematically, this translates into admitting the existence of several distinct representations of the decision space at different levels of refinement. Evidence available on those spaces or *frames* can then be "moved" to a common frame or "common refinement" to be fused. In the theory of evidence, information fusion takes place by combining evidence in the form of belief functions by means of Dempster's orthogonal sum [5].

Dempster's combination, however, is guaranteed to exist [4] only when the original frames are *independent* [15]. Combinability (in Dempster's approach) and independence of frames are strictly intertwined.

Evidence combination has indeed been widely studied [24, 23] in different mathematical frameworks [19, 7]: An exhaustive review would be impossible here [1, 11, 12, 2, 14]. In particular, a lot of work has been done on the issue of merging conflicting evidence [6, 8, 10, 22], while some attention has been given to situations in which the latter comes from dependent sources [3]. On the other hand not much work has been done on the properties of the families of compatible frames [17, 9, 4].

Here we build on the results obtained in [4] to complete an algebraic analysis of families of frames and conduct a comparative study of the notion of independence, so central in the theory of evidence, in an algebraic setup.

First, we recall the fundamental result on the equivalence between independence of sources in Dempster's combination (Section 2) and independence of

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 214–227, 2008.
springerlink.com

frames (Section 2.3). In this incarnation independence of sources can indeed be studied from an algebraic point of view, and compared with other classical forms of independence (Section 2.4)). In the core of the paper (Section 3) we prove in particular that families of compatible frames form *semi-modular lattices*, extending some recent preliminary results [4]. Independence can be defined on semi-modular lattices in several different forms: We can then study the relationship between evidential and lattice independence in all those different formulations (Section 4): they turn out to be distinct, but nevertheless strictly related.

As independence of frames is a direct consequence of independence of Boolean sub-algebras [18], the overall picture opens the way to a more comprehensive definition of this basilar concept.

2 Independence of Sources in Dempster's Combination

2.1 Dempster's Combination of Belief Functions

In the theory of evidence a *basic probability assignment* (b.p.a.) over a finite set or *frame* [15] Θ is a function $m : 2^\Theta \to [0, 1]$ on its power set $2^\Theta = \{A \subseteq \Theta\}$ such that $m(\emptyset) = 0$, $\sum_{A \subseteq \Theta} m(A) = 1$, $m(A) \geq 0 \; \forall A \subseteq \Theta$.

The *belief function* (b.f.) $b : 2^\Theta \to [0, 1]$ associated with a b.p.a. m on Θ is defined as $b(A) = \sum_{B \subseteq A} m(B)$. The *orthogonal sum* or *Dempster's sum* of two b.f.s b_1, b_2 is a new belief function $b_1 \oplus b_2$ with b.p.a.

$$m_{b_1 \oplus b_2}(A) = \frac{\sum_{B \cap C = A} m_{b_1}(B) m_{b_2}(C)}{\sum_{B \cap C \neq \emptyset} m_{b_1}(B) m_{b_2}(C)}. \tag{1}$$

When the denominator of (1) is zero the two b.f.s are *non-combinable*.

2.2 Independence of Sources

Independence plays a central role in Dempster's combination (1), as it is the fundamental assumption under which the combination of two belief functions can actually take place.

Consider a problem in which we have probabilities for a question Q_1 and we want to obtain degrees of belief for a related question Q_2 [16, 20], with Ω and Θ the sets of possible answers to Q_1 and Q_2 respectively. Formally, given a probability measure P on Ω we want to derive a degree of belief $b(A)$ that $A \subseteq \Theta$ contains the correct response to Q_2 (see Figure 1). Let us call $\Gamma(\omega)$ the subset of answers to Q_2 compatible with a given outcome $\omega \in \Omega$ of Q_1. The map $\Gamma : \Omega \to 2^\Theta$ is called a *multi-valued mapping*.

The degree of belief $b(A)$ of an event $A \subseteq \Theta$ is then the total probability of all answers ω that satisfy the above condition, namely $b(A) = P(\{\omega | \Gamma(\omega) \subseteq A\})$.

Consider now two multi-valued mappings Γ_1, Γ_2 inducing two b.f.s b_1, b_2 on the same frame Θ, Ω_1 and Ω_2 their domains and P_1, P_2 the associated probability measures on Ω_1 and Ω_2, respectively. If we suppose that the items of evidence

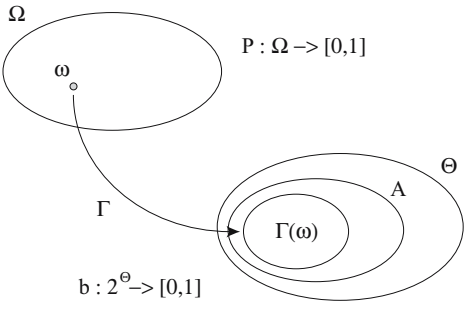


Fig. 1. A probability measure P on Ω induces a belief function b on Θ whose values on the events A of Θ are given by $b(A) = \sum_{\omega \in \Omega: \Gamma(\omega) \subseteq A} P(\omega)$

generating P_1 and P_2 are *independent*, we are allowed to build the product space $(\Omega_1 \times \Omega_2, P_1 \times P_2)$: the detection of two outcomes $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$ will then tell us that the answer to Q_2 is somewhere in $\Gamma(\omega_1, \omega_2) = \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$. We then need to condition the product measure $P_1 \times P_2$ over the set of pairs (ω_1, ω_2) whose images have non-empty intersection: $P = P_1 \times P_2|_\Omega$, with $\Omega = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 | \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) \neq \emptyset\}$. This new belief function b is precisely the orthogonal sum of b_1, b_2.

2.3 Independence of Sources and Independence of Frames

Families of compatible frames

Dempster's mechanism for evidence combination then assumes that the domains on which the evidence is present (in the form of a probability measure) are independent. This concept is mirrored by the notion of *independence of compatible frames* [15]. Given two frames Θ and Θ', a map $\rho : 2^\Theta \to 2^{\Theta'}$ is a *refining* if ρ maps the elements of Θ to a disjoint partition of Θ': $\rho(\{\theta\}) \cap \rho(\{\theta'\}) = \emptyset$ $\forall \theta, \theta' \in \Theta$, $\bigcup_{\theta \in \Theta} \rho(\{\theta\}) = \Theta'$, with $\rho(A) = \cup_{\theta \in A} \rho(\{\theta\})$ $\forall A \subset \Theta$. Θ' is called a *refinement* of Θ, Θ a *coarsening* of Θ'.

Shafer calls a structured collection of frames a *family of compatible frames of discernment* ([15], pages 121-125). In particular, in such a family every pair of frames has a common refinement, i.e. a frame which is a refinement of both. If $\Theta_1, ..., \Theta_n$ are elements of a family of compatible frames $\mathcal{F}$ then there exists a *unique* common refinement $\Theta \in \mathcal{F}$ of them such that $\forall \theta \in \Theta$ $\exists \theta_i \in \Theta_i$ for $i = 1, ..., n$ such that

$$\{\theta\} = \rho_1(\theta_1) \cap \cdots \cap \rho_n(\theta_n),$$

where ρ_i denotes the refining between Θ_i and Θ. This unique frame is called the *minimal refinement* $\Theta_1 \otimes \cdots \otimes \Theta_n$ of $\Theta_1, ..., \Theta_n$.

In the example of Figure 2 we want to find out the position of a target point in an image. We can pose the problem on a frame $\Theta_1 = \{c_1, ..., c_5\}$ obtained by partitioning the column range of the image into 5 intervals, but we can also partition it into 10 intervals, obtaining a different frame $\Theta_2 = \{c_{11}, c_{12}, ..., c_{51}, c_{52}\}$.

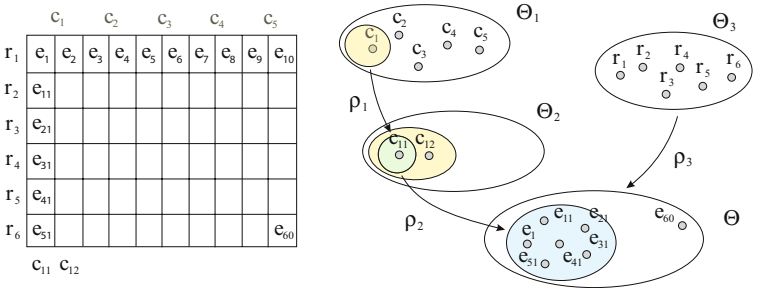


Fig. 2. An example of family of compatible frames. Different discrete quantizations of row and column ranges of an image have a common refinement, the set of cells shown on the left. The refinings ρ_1, ρ_2, ρ_3 between those frames appear to the right.

The row range can also be divided in, say, 6 intervals $\Theta_3 = \{r_1, ..., r_6\}$. All those frames belong to a family of compatible frames, with the collection of cells $\Theta = \{e_1, ..., e_{60}\}$ depicted in Figure 2-left as common refinement, and refinings shown in Figure 2-right. It is easy to verify that Θ is the minimal refinement of Θ_2, Θ_3 as, for example, $\{e_{41}\} = \rho_2(c_{11}) \cap \rho_3(r_4)$.

Independence of frames

Now, let $\Theta_1, ..., \Theta_n$ be elements of a family of compatible frames, and $\rho_i : \Theta_i \to 2^{\Theta_1 \otimes \cdots \otimes \Theta_n}$ the corresponding refinings to their minimal refinement. $\Theta_1, ..., \Theta_n$ are *independent* [15] ($\mathcal{IF}$) if, whenever $\emptyset \neq A_i \subseteq \Theta_i$ for $i = 1, ..., n$,

$$\rho_1(A_1) \cap \cdots \cap \rho_n(A_n) \neq \emptyset. \tag{2}$$

In particular, if $\exists j \in [1, .., n]$ s.t. Θ_j is a coarsening of some other frame Θ_i, $\Theta_1, ..., \Theta_n$ are *not* $\mathcal{IF}$. An equivalent condition is [4]

$$\Theta_1 \otimes \cdots \otimes \Theta_n = \Theta_1 \times \cdots \times \Theta_n \tag{3}$$

i.e. their minimal refinement is their Cartesian product.

Now, independence of frames and Dempster's rule are strictly related [4].

Proposition 1. *Let $\Theta_1, ..., \Theta_n$ elements of a family of compatible frames. Then they are independent iff all the possible collections of b.f.s $b_1, ..., b_n$ defined respectively on $\Theta_1, ..., \Theta_n$ are combinable on their minimal refinement $\Theta_1 \otimes \cdots \otimes \Theta_n$.*

Proposition 1 states that independence of frames and independence of sources (which is at the root of Dempster's combination) are in fact equivalent.

This is not at all surprising when we compare the condition under which Dempster's sum is well defined

$$\Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) \neq \emptyset, \qquad (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$$

with independence of frames expressed as

$$\rho_1(\theta_1) \cap \rho_2(\theta_2) \neq \emptyset, \qquad (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2.$$

2.4 An Algebraic Study of Independence

In its equivalent form of independence of frames (Proposition 1) independence of sources can be analyzed from an algebraic point of view.

A powerful intuition comes from the intriguing similarity between $\mathcal{IF}$ and independence of vector subspaces (recalling Equations 2 and 3):

$$\begin{array}{ccc} \rho_1(A_1) \cap \cdots \cap \rho_n(A_n) \neq \emptyset, \ \forall A_i \subseteq \Theta_i & \equiv & \Theta_1 \otimes \cdots \otimes \Theta_n = \Theta_1 \times \cdots \times \Theta_n \\ v_1 + \cdots + v_n \neq 0, \ \forall v_i \in V_i & \equiv & span\{V_1, ..., V_n\} = V_1 \times \cdots \times V_n. \end{array} \tag{4}$$

While a number of compatible frames $\Theta_1, ..., \Theta_n$ are $\mathcal{IF}$ iff each choice of their representatives $A_i \in 2^{\Theta_i}$ has non-empty intersection, a collection of vector subspaces $V_1, ..., V_n$ is "independent" iff for each choice of vectors $v_i \in V_i$ the sum of those vectors is non-zero. These relations, introduced in what seem very different contexts, can be formally obtained from each other under the following correspondence of quantities and operators:

$$v_i \leftrightarrow A_i, \quad V_i \leftrightarrow 2^{\Theta_i}, \quad + \leftrightarrow \cap, \quad 0 \leftrightarrow \emptyset, \quad \otimes \leftrightarrow span.$$

As we will see here, families of frames and collections of subspaces of a vector space or "projective geometries" share the algebraic structure of *semi-modular lattice*, which in turn admits a characteristic notion of independence. It is natural to wonder how $\mathcal{IF}$ is related to lattice-theoretic independence.

3 The Semi-modular Lattice of Frames

3.1 Lattices

A *partially ordered set* or *poset* is a set P endowed with a binary relation $\leq$ such that, for all x, y, z in P the following conditions hold: 1. $x \leq x$; 2. if $x \leq y$ and $y \leq x$ then $x = y$; 3. if $x \leq y$ and $y \leq z$ then $x \leq z$. In a poset we say that x "covers" y ($x \succ y$) if $x \geq y$ and there is no intermediate element in the chain linking them. A classical example is the power set 2^Θ of a set Θ together with the set-theoretic inclusion $\subset$. Given two elements $x, y \in P$ of a poset P their *least upper bound* $x \vee y$ is the smallest element of P that is bigger than both x and y, while their *greatest lower bound* $x \wedge y$ is the biggest element of P that is smaller than both x and y. Not every pair of elements of a poset, though, is guaranteed to admit inf and/or sup.

A *lattice* L is a poset in which each *pair* of elements admits both inf and sup.

When each *arbitrary* (even not finite) collection of elements of L admits both inf and sup, L is said *complete*. In this case there exist $\mathbf{0} \equiv \wedge L$, $\mathbf{1} \equiv \vee L$ called respectively *initial* and *final* element of L. 2^Θ is complete, with $\mathbf{0} = \emptyset$ and $\mathbf{1} = \{\Theta\}$. The *height* $h(x)$ of an element x in L is the length of a maximal chain from $\mathbf{0}$ to x. In the case of the power set 2^Θ, the height of a subset $A \in 2^\Theta$ is simply its cardinality $|A|$.

3.2 Semi-modularity of the Lattice of Frames

In a family of compatible frames one can define the following order relation:

$$\Theta_1 \leq \Theta_2 \Leftrightarrow \exists \rho : \Theta_2 \rightarrow 2^{\Theta_1} \; refining \tag{5}$$

i.e. Θ_1 is a refinement of Θ_2. The inverse relation $\Theta_1 \leq^* \Theta_2$ iff Θ_1 is a coarsening of Θ_2 is also a valid ordering. After introducing the notion of *maximal coarsening* as the largest cardinality common coarsening $\Theta_1 \oplus \cdots \oplus \Theta_n$ of a given collection of frames $\Theta_1, \cdots, \Theta_n$, we can prove that [4].

Proposition 2. *Both* $(\mathcal{F}, \leq)$ *and* $(\mathcal{F}, \leq^*)$ *where* $\mathcal{F}$ *is a family of compatible frames are lattices, with respectively* $\bigwedge_i \Theta_i = \bigotimes_i \Theta_i$, $\bigvee_i \Theta_i = \bigoplus_i \Theta_i$ *and* $\bigwedge_i^* \Theta_i = \bigoplus_i \Theta_i$, $\bigvee_i^* \Theta_i = \bigotimes_i \Theta_i$.

A special class of lattices arises from *projective geometries*, i.e. collections $L(V)$ of all subspaces of a given vector space V.

Definition 1. *A lattice* L *is* upper semi-modular *if for each pair* x, y *of elements of* L, $x \succ x \wedge y$ *implies* $x \vee y \succ y$. *A lattice* L *is* lower semi-modular *if for each pair* x, y *of elements of* L, $x \vee y \succ y$ *implies* $x \succ x \wedge y$.

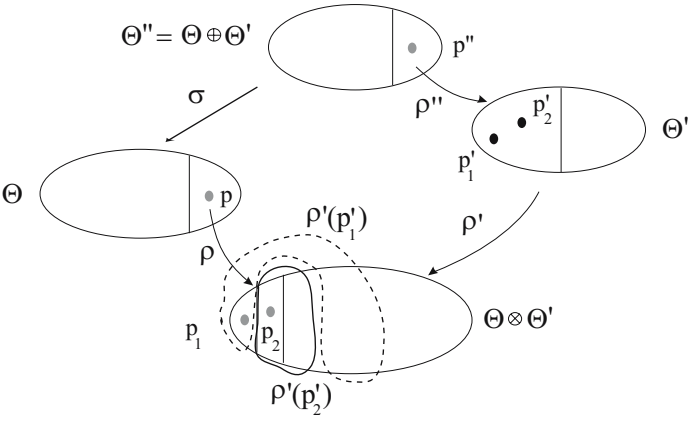


Fig. 3. Upper semi-modularity of $(\mathcal{F}, \leq)$

Clearly if L is upper semi-modular with respect to an order relation $\leq$, the corresponding dual lattice with order relation $\leq^*$ is lower semi-modular.

Theorem 1. $(\mathcal{F}, \leq)$ *is an upper semi-modular lattice;* $(\mathcal{F}, \leq^*)$ *is a lower semi-modular lattice.*

Proof. We just need to prove the upper semi-modularity of $\mathcal{F}$ with respect to $\leq$. Consider two compatible frames Θ, Θ', and suppose that Θ covers their minimal refinement $\Theta \otimes \Theta'$ (their inf with respect to $\leq$). The proof articulates into the following steps (see Figure 3):

- since Θ covers $\Theta \otimes \Theta'$ we have that $|\Theta| = |\Theta \otimes \Theta'| + 1$;
- hence there exists a single element $p \in \Theta$ which is refined into two elements p_1, p_2 of $\Theta \otimes \Theta'$: all other elements of Θ are left unchanged: $\{p_1, p_2\} = \rho(p)$;
- this in turn implies that p_1, p_2 belong each to the image of a different element of Θ' (otherwise Θ would itself be a refinement of Θ', and we would have $\Theta \otimes \Theta' = \Theta$): $p_1 \in \rho'(p'_1)$, $p_2 \in \rho'(p'_2)$;
- if we merge p'_1, p'_2 we have a coarsening Θ'' of Θ': $\{p'_1, p'_2\} = \rho''(p'')$;
- but Θ'' is a coarsening of Θ, too, as we can build the following refining

$$\sigma : \Theta'' \to 2^\Theta : \sigma(q) = \rho'(\rho''(q))$$

  as $\rho'(\rho''(q))$ is a subset of Θ $\forall q \in \Theta''$:
  - if $q = p''$, $\sigma(q)$ is $\{p\} \cup (\rho'(p'_1) \setminus \{p_1\}) \cup (\rho'(p'_2) \setminus \{p_2\})$;
  - if $q \neq p''$, $\rho'(\rho''(q))$ is also a set of elements of Θ, as all elements of Θ but p are left unchanged by ρ;
- as $|\Theta''| = |\Theta'| - 1$, Θ'' is the maximal coarsening of Θ, Θ': $\Theta'' = \Theta \oplus \Theta'$;
- hence Θ' covers $\Theta \oplus \Theta'$, which is the sup of Θ, Θ' in $(\mathcal{F}, \leq)$.

Theorem 1 strengthens the main result of [4], where we proved that finite families of frames are Birkhoff. A lattice is *Birkhoff* if $x \wedge y \prec x, y$ implies $x, y \prec x \vee y$. (Upper) semi-modularity implies the Birkhoff property, but not vice-versa.

3.3 Finite Lattice of Frames

We will here focus on *finite* families of frames. Given a set of compatible frames $\Theta_1, ..., \Theta_n$ consider the set $P(\Theta)$ of all partitions of their minimal refinement $\Theta = \Theta_1 \otimes \cdots \otimes \Theta_n$. As $\mathcal{IF}$ involves only partitions of $\Theta_1 \otimes \cdots \otimes \Theta_n$, we can conduct our analysis there. We denote by

$$L(\Theta) \doteq (P(\Theta), \leq), \qquad L^*(\Theta) \doteq (P(\Theta), \leq^*)$$

the two lattices associated with $P(\Theta)$. Consider for example the partition lattice associated with a frame of size 4: $\Theta = \{1, 2, 3, 4\}$, depicted in Figure 4. According to the ordering $\leq^*$ each edge indicates that the bottom element is bigger than the top one. If we pick the pair of partitions $y = \{1, 2/3, 4\}$ and $y' = \{1, 3/2, 4\}$, we can notice that both y, y' cover their inf $y \wedge^* y' = \{1, 2, 3, 4\}$ but their sup $y \vee^* y' = \{1/2/3/4\}$ does not cover any of them. Hence, $(P(\Theta), \leq^*)$ is not upper semi-modular but lower semi-modular.

3.4 A Lattice-Theoretic Interpretation of Independence

We can now reinterpret the analogy introduced in Section 2.4 between subspaces of a vector space V and elements of a family of compatible frames. Both are lattices: according to the chosen order relation we get an upper $L(\Theta)$ or lower $L^*(\Theta)$ semi-modular lattice (see table)

lattice	$L(V)$	$L^*(\Theta)$	$L(\Theta)$
initial element $\mathbf{0}$	$\{0\}$	$\mathbf{0}_{\mathcal{F}}$	Θ
$\sup l_1 \vee l_2$	$span(V_1, V_2)$	$\Theta_1 \otimes \Theta_2$	$\Theta_1 \oplus \Theta_2$
$\inf l_1 \wedge l_2$	$V_1 \cap V_2$	$\Theta_1 \oplus \Theta_2$	$\Theta_1 \otimes \Theta_2$
order relation $l_1 \leq l_2$	$V_1 \subseteq V_2$	Θ_1 coars. of Θ_2	Θ_1 refin. of Θ_2
height $h(l_1)$	$\dim(V_1)$	$\lvert\Theta_1\rvert - 1$	$\lvert\Theta\rvert - \lvert\Theta_1\rvert$

where $\mathbf{0}_{\mathcal{F}}$ denotes the unique frame of a family $\mathcal{F}$ with cardinality 1.

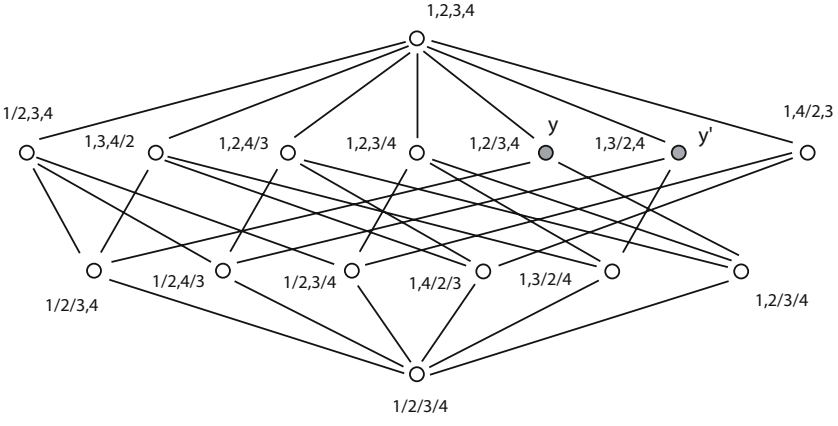


Fig. 4. The partition (lower) semi-modular lattice $L^*(\Theta)$ for a frame Θ of size 4. Partitions $A_1, ..., A_k$ of Θ are denoted by $A_1/.../A_k$. Partitions with the same number of elements are arranged on the same level.

4 Independence on Lattices and Independence of Frames

4.1 Independence on Lattices

As a matter of fact, abstract independence can be defined on the elements of a semi-modular lattice [21]. Consider again the classical example of linear independence of vectors. By definition $v_1, ..., v_n$ are *linearly independent* iff $\sum_i \alpha_i v_i = 0 \vdash \alpha_i = 0 \ \forall i$: Well known equivalent conditions are:

$$\begin{array}{lll} \mathcal{I}_1: & v_j \not\subset span(v_i, i \neq j) & \forall j = 1, ..., n; \\ \mathcal{I}_2: & v_j \cap span(v_1, ..., v_{j-1}) = 0 & \forall j = 2, ..., n; \\ \mathcal{I}_3: & \dim(span(v_1, ..., v_n)) = n. & \end{array} \tag{6}$$

As 1D subspaces are elements of a lattice $L(V)$ for which $span = \vee$, $\cap = \wedge$, $dim = h$ and $\mathbf{0} = 0$ we can generalize the relations (6) to collections $\{l_1, ..., l_n\}$ of non-zero elements of any semi-modular lattice with initial element $\mathbf{0}$ as

$$\begin{array}{ll} \mathcal{I}_1: & l_j \not\leq \bigvee_{i\neq j} l_i \qquad \forall j=1,...,n; \\ \mathcal{I}_2: & l_j \wedge \bigvee_{i<j} l_i = \mathbf{0} \qquad \forall j=2,...,n; \\ \mathcal{I}_3: & h(\bigvee_i l_i) = \sum_i h(l_i). \end{array} \tag{7}$$

4.2 Lattice-Theoretic Independence on the Lattice of Frames

Independence assumes then several different forms in lattice theory. As compatible frames form semi-modular lattices it is natural to suppose that some of those may indeed coincide with Shafer's independence of frames, or at least have some relations with it.

We analyze the relations (7) in the flag lower semi-modular case $L^*(\Theta)$:

$$\begin{aligned} &\Theta_1,...,\Theta_n \ \mathcal{I}_1^* \Leftrightarrow && \Theta_j \oplus \bigotimes_{\substack{i\neq j \\ j-1}} \Theta_i \neq \Theta_j && \forall\, j=1,...,n \\ &\Theta_1,...,\Theta_n \ \mathcal{I}_2^* \Leftrightarrow && \Theta_j \oplus \bigotimes_{i=1} \Theta_i = \mathbf{0}_{\mathcal{F}} && \forall\, j=2,...,n \\ &\Theta_1,...,\Theta_n \ \mathcal{I}_3^* \Leftrightarrow && \Big|\bigotimes_{i=1}^n \Theta_i\Big| - 1 = \sum_{i=1}^n (|\Theta_i|-1) \end{aligned} \tag{8}$$

as

$$\Theta_i \wedge \Theta_j = \Theta_i \oplus \Theta_j, \quad \Theta_i \vee \Theta_j = \Theta_i \otimes \Theta_j, \quad h^*(\Theta_i) = |\Theta_i| - 1.$$

The frames $\Theta_1,...,\Theta_n$ are $\mathcal{I}_1^*$ iff none of them is a coarsening of the minimal refinement of all the others; they are $\mathcal{I}_2^*$ iff $\forall j > 1$ Θ_j does not have a non-trivial common coarsening with the minimal refinement of its predecessors. $\mathcal{I}_3^*$ on its side has a very interesting semantics in terms of probability spaces: As the dimension of the polytope of probability measures definable on a domain of size k is $k-1$, $\mathcal{I}_3^*$ is equivalent to say that the dimension of the probability polytope for the minimal refinement is the sum of the dimensions of the polytopes associated with the individual frames.

4.3 Evidential Independence Is Stronger than $\mathcal{I}_1^*$, $\mathcal{I}_2^*$

To study the logical implications between these lattice-theoretic relations and independence of frames we first need to prove an interesting Lemma.

Lemma 1. $\Theta_1,...,\Theta_n \ \mathcal{IF},\ n>1 \vdash \bigoplus_{i=1}^n \Theta_i = \mathbf{0}_{\mathcal{F}}$.

Proof. We prove Lemma 1 by induction. For $n=2$, let us suppose that Θ_1, Θ_2 are $\mathcal{IF}$. Then $\rho_1(A_1) \cap \rho_2(A_2) \neq \emptyset\ \forall A_1 \subseteq \Theta_1, A_2 \subseteq \Theta_2,\ A_1, A_2 \neq \emptyset$ (ρ_i denotes as usual the refining from Θ_i to $\Theta_1 \otimes \Theta_2$). Suppose by absurd that their common coarsening has more than a single element, $\Theta_1 \oplus \Theta_2 = \{a, b\}$. But then $\rho_1(\rho^1(a)) \cap \rho_2(\rho^2(b)) = \emptyset$, where ρ^i denotes the refining between $\Theta_1 \oplus \Theta_2$ and Θ_i, which goes against the hypothesis.

Induction step. Suppose that the thesis holds for $n-1$. Then, since $\Theta_1,...,\Theta_n$ $\mathcal{IF}$ implies $\{\Theta_i, i \neq j\}$ $\mathcal{IF}$ $\forall j$, this implies by inductive hypothesis that

$$\bigoplus_{i \neq j} \Theta_i = \mathbf{0}_{\mathcal{F}} \quad \forall j = 1, ..., n.$$

Of course then, as $\mathbf{0}_{\mathcal{F}}$ is a coarsening of Θ_j $\forall j = 1, ..., n$,

$$\Theta_j \oplus \bigoplus_{i \neq j} \Theta_i = \Theta_j \oplus \mathbf{0}_{\mathcal{F}} = \mathbf{0}_{\mathcal{F}}.$$

We can use Lemma 1 to state that evidential independence of frames is indeed *stronger* than lattice-theoretic independence of frames in its first form.

Theorem 2. *$\Theta_1, ..., \Theta_n$ $\mathcal{IF}$ and $\Theta_j \neq \mathbf{0}_{\mathcal{F}}$ $\forall j$ then $\Theta_1, ..., \Theta_n$ $\mathcal{I}_1^*$.*

Proof. Let us suppose that $\Theta_1, ..., \Theta_n$ are $\mathcal{IF}$ but not $\mathcal{I}_1^*$, i.e. $\exists j : \Theta_j$ coarsening of $\bigotimes_{i \neq j} \Theta_i$.

We need to prove that $\exists A_1 \subseteq \Theta_1, ..., A_n \subseteq \Theta_n$ s.t. $\rho_1(A_1) \cap \cdots \cap \rho_n(A_n) = \emptyset$ where ρ_i denotes the refining from Θ_i to $\Theta_1 \otimes \cdots \otimes \Theta_n$ (Equation 2).

Since Θ_j is a coarsening of $\bigotimes_{i \neq j} \Theta_i$ then there exists a partition Π_j of $\bigotimes_{i \neq j} \Theta_i$ associated with Θ_j, and a refining ρ from Θ_j to $\bigotimes_{i \neq j} \Theta_i$.

As $\{\Theta_i, i \neq j\}$ are $\mathcal{IF}$, for all $\theta \in \bigotimes_{i \neq j} \Theta_i$ there exists $\theta_i \in \Theta_i$, $i \neq j$ s.t.

$$\{\theta\} = \bigcap_{i \neq j} \rho_i'(\Theta_i),$$

where ρ_i' is the refining to $\bigotimes_{i \neq j} \Theta_i$. Now, θ belongs to a certain element A of the partition Π_j. By hypothesis ($\Theta_j \neq \mathbf{0}_{\mathcal{F}}$ $\forall j$) Π_j contains at least two elements. But then we can choose $\theta_j = \rho^{-1}(B)$ with B another element of Π_j. In that case we obviously get

$$\rho_j(\theta_j) \cap \bigcap_{i \neq j} \rho_i(\theta_i) = \emptyset$$

which implies that $\{\Theta_i, i = 1, ..., n\} \neg \mathcal{IF}$ against the hypothesis.

However, the two notions are not equivalent: $\Theta_1, ..., \Theta_n$ $\mathcal{I}_1^* \not\vdash \Theta_1, ..., \Theta_n$ $\mathcal{IF}$. Consider as a counterexample two frames Θ_1 and Θ_2 in which Θ_1 is not a coarsening of Θ_2 (Θ_1, Θ_2 are $\mathcal{I}_1^*$). Then $\Theta_1, \Theta_2 \neq \Theta_1 \otimes \Theta_2$ but it easy to find a situation (see Figure 6-left) in which Θ_1, Θ_2 are not $\mathcal{IF}$.

More, it is easy to prove that $\mathcal{IF}$ is also *stronger than the second form* of lattice-theoretic independence.

Theorem 3. *$\Theta_1, ..., \Theta_n$ $\mathcal{IF} \vdash \Theta_1, ..., \Theta_n$ $\mathcal{I}_2^*$.*

Proof. We first need to show that $\Theta_1, ..., \Theta_n$ are $\mathcal{IF}$ iff the pair $\{\Theta_j, \otimes_{i \neq j} \Theta_i\}$ is $\mathcal{IF}$. As a matter of fact (3) can be written as

$$\Theta_j \otimes \bigotimes_{i \neq j} \Theta_i = \Theta_j \times (\times_{i \neq j} \Theta_i) \equiv \left\{ \Theta_j, \bigotimes_{i \neq j} \Theta_i \right\} \mathcal{IF}.$$

But then by Lemma 1 we get as desired.

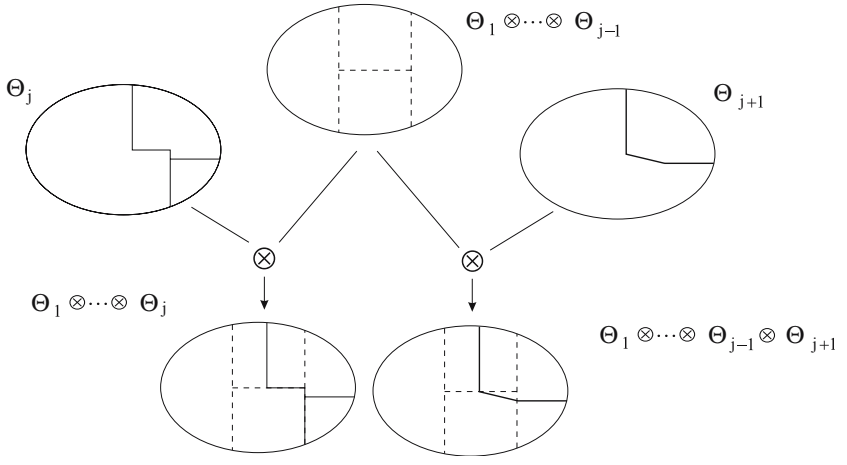


Fig. 5. A counterexample to $\mathcal{I}_2^* \vdash \mathcal{I}_1^*$

These two form of independence $\mathcal{I}_1^*$, $\mathcal{I}_2^*$ are not trivially related to each other: for instance, $\Theta_1, ..., \Theta_n$ $\mathcal{I}_2^*$ does not imply $\Theta_1, ..., \Theta_n$ $\mathcal{I}_1^*$. Figure 5 shows a counterexample: Given $\Theta_1 \otimes \cdots \otimes \Theta_{j-1}$ and Θ_j, one choice of Θ_{j+1} s.t. $\Theta_1, ..., \Theta_{j+1}$ are $\mathcal{I}_2^*$ but not $\mathcal{I}_1^*$ is shown.

It follows from Theorems 2 and 3 that, unless some frame is unitary, $\mathcal{IF} \vdash \mathcal{I}_1^* \wedge \mathcal{I}_2^*$. The converse is however false. Think of a pair of frames ($n = 2$), for which $\Theta_1 \oplus \Theta_2 \neq \Theta_1, \Theta_2$ ($\mathcal{I}_1^*$), $\Theta_1 \oplus \Theta_2 = \mathbf{0}_\mathcal{F}$ ($\mathcal{I}_2^*$). Now, those conditions are met by the counterexample of Figure 5, in which the two frames are not $\mathcal{IF}$.

4.4 Evidential Independence Is Opposed to $\mathcal{I}_3^*$

On its side, lattice independence in its third form $\mathcal{I}_3^*$ is actually *incompatible with evidential independence*.

Theorem 4. *If* $\Theta_1, ..., \Theta_n$ $\mathcal{IF}$, $n > 2$ *then* $\Theta_1, ..., \Theta_n \neg\mathcal{I}_3^*$. *If* Θ_1, Θ_2 $\mathcal{IF}$ *then* Θ_1, Θ_2 $\mathcal{I}_3^*$ *iff* $\exists \Theta_i = \mathbf{0}_\mathcal{F}$ $i \in \{1, 2\}$.

Proof. According to (3), $\Theta_1, ..., \Theta_n$ are $\mathcal{IF}$ iff $|\otimes \Theta_i| = \prod_i |\Theta_i|$, while according to (8) they are $\mathcal{I}_3$ iff $|\Theta_1 \otimes \cdots \otimes \Theta_n| - 1 = \sum_i (|\Theta_i| - 1)$. They are both met iff

$$\sum_i |\Theta_i| - \prod_i |\Theta_i| = n - 1,$$

which happens only if $n = 2$ and either $\Theta_1 = \mathbf{0}_\mathcal{F}$ or $\Theta_1 = \mathbf{0}_\mathcal{F}$.

Stronger results hold when considering only pairs of frames. For $n = 2$ the relations (8) read as

$$\Theta_1 \oplus \Theta_2 \neq \Theta_1, \Theta_2, \qquad \Theta_1 \oplus \Theta_2 = \mathbf{0}_\mathcal{F}, \qquad |\Theta_1 \otimes \Theta_2| = |\Theta_1| + |\Theta_2| - 1. \tag{9}$$

Theorem 5. *If* $\Theta_1, \Theta_2 \neq \mathbf{0}_\mathcal{F}$ *then* $\Theta_1, \Theta_2\ \mathcal{I}_2^*$ *implies* $\Theta_1, \Theta_2\ \mathcal{I}_1^*$. *If* $\exists \Theta_j = \mathbf{0}_\mathcal{F}$ $j \in \{1, 2\}$ *then* $\Theta_1, \Theta_2\ \mathcal{I}_2^*, \mathcal{I}_3^*, \mathcal{IF}, \neg\mathcal{I}_1^*$.

Proof. The first fact is obvious from (9). If instead $\Theta_2 = \mathbf{0}_\mathcal{F}$ then by (9) $\Theta_1 \oplus \mathbf{0}_\mathcal{F} = \mathbf{0}_\mathcal{F} = \Theta_2$ and Θ_1, Θ_2 are not $\mathcal{I}_1^*$ while they are $\mathcal{I}_2^*$. As $|\Theta_1 \otimes \Theta_2| = |\mathbf{0}_\mathcal{F}| \cdot |\Theta_1| = |\Theta_1|$, $|\Theta_2| = 1$ in that case Θ_1, Θ_2 are $\mathcal{I}_3^*$ again by (9). Finally, according to (3), they are $\mathcal{IF}$ as $|\Theta_1 \otimes \Theta_2| = |\Theta_1| = 1 \cdot |\Theta_1| = |\Theta_2||\Theta_1|$.

For the binary partitions of Θ, i.e. the *atoms* (elements covering $\mathbf{0}$) A^* of the lattice $L^*(\Theta)$, Theorem 4 implies that $\mathcal{IF}$ and $\mathcal{I}_3^*$ are incompatible.

Corollary 1. *If* $\Theta_1, ..., \Theta_n \in A^*$ *then* $\Theta_1, ..., \Theta_n\ \mathcal{IF}$ *implies* $\Theta_1, ..., \Theta_n\ \neg\mathcal{I}_3^*$.

On the other side, the other two relations are trivial for atoms of $L^*(\Theta)$.

Theorem 6. *If* $\Theta_1, ..., \Theta_n \in A^*$ *then* $\Theta_1, ..., \Theta_n$ *are both* $\mathcal{I}_1^*$ *and* $\mathcal{I}_2^*$.

Proof. If $\Theta_j \in A^*\ \forall j$ then $\Theta_j \oplus \bigotimes_{i \neq j} \Theta_i = \mathbf{0}_\mathcal{F} \neq \Theta_j\ \forall j$ and $\Theta_1, ..., \Theta_n$ are $\mathcal{I}_1^*$. But then by Equation 8 $\Theta_1, ..., \Theta_n$ are also $\mathcal{I}_2^*$.

As a matter of fact, evidential independence and $\mathcal{I}_3^*$ are in opposition for pairs of atoms of $L^*(\Theta)$.

Theorem 7. $\Theta_1, \Theta_2 \in A^*$ *are* $\mathcal{IF}$ *iff* $\Theta_1, \Theta_2 \neg\mathcal{I}_3^*$.

Proof. By Theorem 4 we have that $\mathcal{IF} \vdash \neg\mathcal{I}_3^*$. To prove the reverse implication $\mathcal{I}_3^* \vdash \neg\mathcal{IF}$ we just need to notice that $\Theta_1, \Theta_2 \in A^*$ are $\mathcal{I}_3^*$ iff $|\Theta_1 \otimes \Theta_2| = |\Theta_1| + |\Theta_2| - 1 = 2 + 2 - 1 = 3$ while for them to be $\mathcal{IF}$ it has to be $|\Theta_1 \otimes \Theta_2| = |\Theta_1||\Theta_2| = 2 \cdot 2 = 4$.

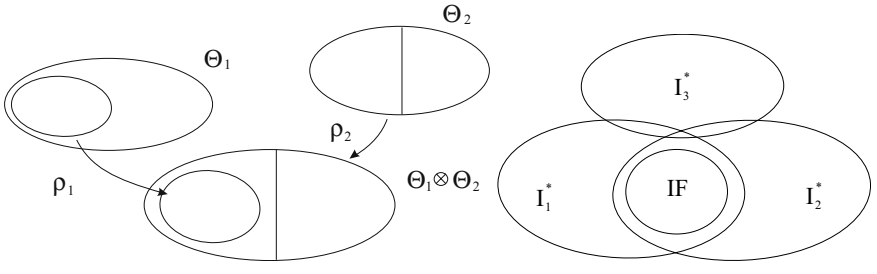


Fig. 6. Left: A counterexample to $\mathcal{I}_1^* \vdash \mathcal{IF}$. Right: Relations between independence of frames $\mathcal{IF}$ and all different forms of semi-modular independence on the lower semi-modular lattice of frames $L^*(\Theta)$.

5 Comments and Conclusions

Independence of sources in the theory of evidence can be reduced to independence of frames. This shows in turn intriguing formal analogies with linear independence. In this paper we proved that families of frames share indeed with

projective geometries the algebraic structure of semi-modular lattice (Theorem 1). Several forms of independence relations can be introduced on the elements of such lattices, and related with Shafer's independence of frames.

Figure 6-right illustrates what we have learned about how $\mathcal{IF}$ relates to the various forms of lattice-theoretic independence in the lower semi-modular lattice of frames, in the general case of a collection of more than two non-atomic frames (the case $\Theta_i = \mathbf{0}_\mathcal{F}$ is neglected). Evidential independence appears distinct from but related to lattice-theoretic independence.

This is even more interesting when we consider that condition (2) comes directly from the notion of independence of frames as Boolean sub-algebras [18]. Boolean independence $\mathcal{IF}$ is *a stronger condition* than both $\mathcal{I}_1^*$ and $\mathcal{I}_2^*$ (Theorems 2, 3) which are indeed trivial for binary partitions of Θ (Theorem 6). On the other side $\mathcal{IF}$ and $\mathcal{I}_3^*$ are mutually exclusive (Theorems 4 and 7, Corollary 1). As $\mathcal{I}_3^*$ is in turn a form of matroidal independence [13] this sheds new light on the relation between Boolean algebra and matroid theory.

The prosecution of this study, particularly in the context of matroid theory, could in the future shed some more light on both the nature of independence of sources in the theory of subjective probability, and the relationship between lattice, matroidal and Boolean independence in discrete mathematics, pointing out the necessity of a more general, comprehensive definition of this very useful and widespread notion.

References

1. Campos, F., de Souza, F.M.C.: Extending Dempster-Shafer theory to overcome counter intuitive results. In: Proceedings of IEEE NLP-KE 2005, vol. 3, pp. 729–734 (2005)
2. Carlson, J., Murphy, R.R.: Use of Dempster-Shafer conflict metric to adapt sensor allocation to unknown environments, Tech. report, Safety Security Rescue Research Center, University of South Florida (2005)
3. Cattaneo, M.E.G.V.: Combining belief functions issued from dependent sources. ISIPTA, 133–147 (2003)
4. Cuzzolin, F.: Algebraic structure of the families of compatible frames of discernment. AMAI 45(1-2), 241–274 (2005)
5. Dempster, A.P.: Upper and lower probabilities generated by a random closed interval. Annals of Mathematical Statistics 39, 957–966 (1968)
6. Deutsch-McLeish, M.: A study of probabilities and belief functions under conflicting evidence: Comparisons and new method. In: Bouchon-Meunier, B., Zadeh, L.A., Yager, R.R. (eds.) IPMU 1990. LNCS, vol. 521, pp. 41–49. Springer, Heidelberg (1991)
7. Dubois, D., Prade, H.: On the combination of evidence in various mathematical frameworks. In: flamm, J., Luisi, T. (eds.) Reliability Data Collection and Analysis, pp. 213–241 (1992)
8. Josang, A., Daniel, M., Vannoorenberghe, P.: Strategies for combining conflicting dogmatic beliefs. In: Proceedings of Fusion 2003, vol. 2, pp. 1133–1140 (2003)
9. Kohlas, J., Monney, P.-A.: A mathematical theory of hints - an approach to the dempster-shafer theory of evidence. Lecture Notes in Economics and Mathematical Systems. Springer, Heidelberg (1995)

10. Lefevre, E., Colot, O., Vannoorenberghe, P.: Belief functions combination and conflict management. Information Fusion Journal 3(2), 149–162 (2002)
11. Liu, W.: Analyzing the degree of conflict among belief functions. Artif. Intell. 170(11), 909–924 (2006)
12. Murphy, C.K.: Combining belief functions when evidence conflicts. Decision Support Systems 29, 1–9 (2000)
13. Oxley, J.G.: Matroid theory. Oxford University Press, Great Clarendon Street, Oxford (1992)
14. Sentz, K., Ferson, S.: Combination of evidence in Dempster-Shafer theory, Tech. report, SANDIA Tech. Report, SAND2002-0835 (April 2002)
15. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
16. Shafer, G.: Perspectives on the theory and practice of belief functions. International Journal of Approximate Reasoning 4, 323–362 (1990)
17. Shafer, G., Shenoy, P.P., Mellouli, K.: Propagating belief functions in qualitative Markov trees. IJAR 1(4), 349–400 (1987)
18. Sikorski, R.: Boolean algebras. Springer, Heidelberg (1964)
19. Smets, P.: The combination of evidence in the transferable belief models. IEEE Transactions on PAMI 12, 447–458 (1990)
20. Smets, P.: Upper and lower probability functions versus belief functions. In: Proceedings of the International Symposium on Fuzzy Systems and Knowledge Engineering, Guangzhou, China, pp. 17–21 (1987)
21. Szasz, G.: Introduction to lattice theory. Academic Press, New York (1963)
22. Wierman, M.J.: Measuring conflict in evidence theory. In: Proceedings of the Joint 9th IFSA World Congress, Vancouver, BC, Canada, vol. 3, pp. 1741–1745 (2001)
23. Yager, R.R.: On the Dempster-Shafer framework and new combination rules. Information Sciences 41, 93–138 (1987)
24. Zadeh, L.: A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. AI Magazine 7(2), 85–90 (1986)

Part V

Non-classical Logics

Completions of Ordered Algebraic Structures: A Survey

John Harding

Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003
jharding@nmsu.edu
Http://www.math.nmsu.edu/∼jharding

Summary. Ordered algebraic structures are encountered in many areas of mathematics. One frequently wishes to embed a given ordered algebraic structure into a complete ordered algebraic structure in a manner that preserves some aspects of the algebraic and order theoretic properties of the original. It is the purpose here to survey some recent results in this area.

1 Introduction

An ordered algebraic structure $\mathcal{A}$ consists of an algebra, in the sense commonly used in universal algebra [9], together with a partial ordering on the underlying set of the algebra. We require that the operations of the algebra are compatible with the partial ordering in that they preserve or reverse order in each coordinate. The partial orderings we consider here will almost always be lattice orderings.

Ordered algebraic structures occur in a wide variety of areas. Examples include partially ordered vector spaces, lattice ordered groups, Boolean algebras, Heyting algebras, modal algebras, cylindric algebras, relation algebras, orthomodular posets, and so forth. In applications, the existence of certain infinite joins and meets often play an important role. In analytic applications, certain infinite joins and meets are often related to limit processes; in logical applications, certain infinite joins and meets are often related to existential and universal quantification; and in quantum logic, countable orthogonal joins correspond to experiments built from countable families of mutually exclusive experiments. It is a common task to try to embed a given ordered algebraic structure into one where certain families of joins and meets exist.

Perhaps the best example is the earliest one. In 1858 (published in 1872 [12]) Dedekind used his methods of *cuts* to construct the real numbers $\mathbb{R}$ from the rationals $\mathbb{Q}$. He defined a real number to be a certain type of ordered pair (A, B) of subsets of the rationals called a cut. Each rational q yields such a cut, and this provides an embedding $\varphi : \mathbb{Q} \to \mathbb{R}$. Dedekind further defines an ordering $\leq$ and operations $+, -, \cdot$ on $\mathbb{R}$. He shows that with these operations $(\mathbb{R}, +, -\cdot, \leq)$ is an ordered field that is conditionally complete, meaning that

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 231–244, 2008.
springerlink.com

every non-empty subset of $\mathbb{R}$ that has an upper bound has a least upper bound and every non-empty subset that has a lower bound has a greatest lower bound. Having embedded the rationals into a conditionally complete ordered field, one might ask whether the rationals can even be embedded into a complete ordered field. This is trivially impossible as an ordered field can never have a largest or least element.

The embedding $\varphi : \mathbb{Q} \to \mathbb{R}$ produced above is more than just an order embedding that preserves algebraic structure. The map φ preserves all existing joins and meets in $\mathbb{Q}$, a property we call a regular embedding. Further, each element of $\mathbb{R}$ is both a join and meet of elements of the image of φ, properties called join and meet density. In many instances it may be desirable to find a completion that not only preserves some existing algebraic properties, but also preserves some existing joins and meets. Further, having some sort of density condition, to ensure the resulting completion is somewhat tightly tied to the original, is often desirable.

For a given type of algebraic structure, one can ask a variety of questions regarding the existence of an embedding into a complete ordered structure preserving certain aspects of the algebraic and order theoretic structure. There is a large, mostly scattered, literature on such questions for specific classes of structures. It is not our intent to review this literature in more than an incidental way. Rather, we concentrate on results, mostly in the past 20 years, that seem to form the beginnings of a general theory of such completions.

2 Preliminaries

In this section we review some basic definitions.

Definition 1. *For a poset P, an n-ary operation f on P is called monotone if it preserves or reverses order in each coordinate. An ordered algebraic structure $\mathcal{A} = (A, (f_i)_I, \leq)$ consists of an algebra $(A, (f_i)_I)$, together with a partial ordering $\leq$ on A, such that for each $i \in I$, the operation f_i is monotone.*

Note, this definition allows for Heyting implication $\longrightarrow$ that is order reversing in the first coordinate, and order preserving in the second.

Mostly we will consider here ordered algebraic structures where the underlying ordering is a lattice ordering. Such structures are also known under the name of monotone lattice expansions [14, 15].

Definition 2. *An embedding of an ordered algebraic structures $\mathcal{A}$ into $\mathcal{B}$ is a map $\varphi : A \to B$ that is both a homomorphism and an order embedding. A completion of $\mathcal{A}$ is an embedding $\varphi : \mathcal{A} \to \mathcal{B}$ where the underlying ordering of $\mathcal{B}$ is a complete lattice.*

For an ordered algebraic structure whose underlying ordering is a lattice, one might reasonably argue that an embedding should be required to be a lattice embedding. This is easily accomplished by keeping the current definition and adding the lattice operations as part of the basic algebraic structure.

Definition 3. *An embedding $\varphi : \mathcal{A} \to \mathcal{B}$ is called join dense if for each $b \in B$, we have $b = \bigvee\{\varphi(a) : a \in A \text{ and } \varphi(a) \leq b\}$. We say φ is join regular if for each $S \subseteq A$ that has a join in $\mathcal{A}$, the image $\varphi[S]$ has a join in $\mathcal{B}$, and $\varphi(\bigvee S) = \bigvee \varphi[S]$. Meet dense and meet regular are defined similarly. Finally, call φ regular if it is both join regular and meet regular.*

The following is well known, and easy to prove.

Proposition 1. *If φ is join dense, then it is meet regular, and if φ is meet dense, then it is join regular.*

We next describe various types of ideals an filters that play an important role. We recall that for a subset S of a poset P, that $U(S) = \{p \in P : s \leq p \text{ for all } s \in S\}$ is the set of upper bounds of S, and $L(S) = \{p \in P : p \leq s \text{ for all } s \in S\}$ is the set of lower bounds of S.

Definition 4. *For P a poset and $I \subseteq P$ we say*

(i) I is an order ideal if $b \in I$ and $a \leq b \Rightarrow a \in I$.
(ii) I is an ideal if I is an order ideal that is closed under existing finite joins.
(iii) I is a normal ideal if $I = LU(I)$.

Let $\mathcal{I}_O P$, $\mathcal{I}P$ and $\mathcal{I}_N P$ be the sets of order ideals, ideals, and normal ideals of P, considered as posets under the partial ordering of set inclusion.

Similarly an order filter is a subset $F \subseteq P$ where $a \in F$ and $a \leq b$ implies $b \in F$, a filter is an order filter closed under existing finite meets, and a normal filter is a set F with $F = UL(F)$. We let $\mathcal{F}_O P$, $\mathcal{F}P$ and $\mathcal{F}_N P$ be the sets of order filters, filters, and normal filters partially ordered by reverse set inclusion.

Proposition 2. *S is a normal ideal iff it is the intersection of principal ideals, and S is a normal filter iff it is the intersection of principal filters. Normal ideals are closed under all existing joins, and normal filters are closed under all existing meets.*

While normal ideals are closed under existing joins, in general, there are ideals of a lattice that are closed under existing joins but are not normal. However, for Heyting algebras an ideal is normal iff it is closed under existing joins, but the corresponding result does not hold for normal filters in a Heyting algebra [5].

Definition 5. *For $a \in P$ let $a\downarrow = \{p \in P : p \leq a\}$ and $a\uparrow = \{p \in P : a \leq p\}$. We call these the principal ideal and principal filter generated by a.*

3 Completion Methods

In this section we collect a number of common completion methods, as well as a general template into which these methods fit. We first discuss matters for posets and lattices, considering additional algebraic operations later.

Proposition 3. *For a poset P, the order ideals $\mathcal{I}_O P$ are a complete lattice and the map $\varphi : P \to \mathcal{I}_O P$ defined by $\varphi(a) = a{\downarrow}$ is a completion of P satisfying*

(i) φ is join dense.
(ii) For $a \in P, S \subseteq P$, if $\varphi(a) \leq \bigvee \varphi[S]$ then $a \leq s$ for some $s \in S$.

Further, if $\psi : P \to C$ is another completion satisfying these two properties, there is a unique isomorphism $\mu : \mathcal{I}_O P \to C$ with $\mu \circ \varphi = \psi$.

The order ideal completion preserves all existing meets as it is a join dense completion, but destroys all existing joins except those of subsets with a maximum element. Also of interest is that $\mathcal{I}_O P$ is a completely distributive lattice.

Proposition 4. *For a poset P, the ideals $\mathcal{I}P$ are a complete lattice and the map $\varphi : P \to \mathcal{I}P$ defined by $\varphi(a) = a{\downarrow}$ is a completion of P satisfying*

(i) φ is join dense.
(ii) For $a \in P, S \subseteq P$, if $\varphi(a) \leq \bigvee \varphi[S]$ then $\varphi(a) \leq \bigvee \varphi[S']$ for some finite $S' \subseteq S$.

Further, if $\psi : P \to C$ is another completion satisfying these two properties, there is a unique isomorphism $\mu : \mathcal{I}_O P \to C$ with $\mu \circ \varphi = \psi$.

The ideal completion preserves all existing meets, and all existing finite joins, however it destroys all existing joins that are not essentially finite. In the next section when we consider preservation of identities, we see some of the main advantages of the ideal completion.

Proposition 5. *For a poset P, the normal ideals $\mathcal{I}_N P$ are a complete lattice and the map $\varphi : P \to \mathcal{I}P$ defined by $\varphi(a) = a{\downarrow}$ is a completion of P satisfying*

(i) φ is join dense.
(ii) φ is meet dense.

Further, if $\psi : P \to C$ is another completion satisfying these two properties, there is a unique isomorphism $\mu : \mathcal{I}_O P \to C$ with $\mu \circ \varphi = \psi$.

The normal ideal completion is often called the MacNeille completion, or the completion by cuts. It was introduced by MacNeille [28] in the poset setting as an extension of the method used by Dedekind to construct the reals from the rationals. It preserves all existing joins and meets, so is a regular completion. The above characterization is due to Banaschewski and Schmidt [2, 31]. It provides a minimal completion of P in that for any completion $f : P \to C$ there is an order embedding $\mu : \mathcal{I}_N P \to C$ with $\mu \circ \varphi = f$, and this can be used to show that MacNeille completions provide strictly injective essential extensions in the category of posets [3].

One similarly obtains order filter, filter, and filter completions of a poset P using the embedding $\psi(a) = a{\uparrow}$. They have similar properties to the above, except that the roles of joins and meets are interchanged. As the roles of join and meet are symmetric for the normal ideal completion, $\mathcal{I}_N P$ and $\mathcal{F}_N P$ are isomorphic, and the maps U, L provide mutually inverse isomorphisms.

One can create additional types of completions by taking other families of order ideals or order filters that are closed under intersections. For instance, the family of all order ideals closed under existing countable joins provides a join dense completion that preserves all existing meets, and existing countable joins. We soon generalize this observation, but first we consider one more completion.

Proposition 6. *For a bounded lattice L, there is a completion $\varphi : L \to C$ satisfying*

(i) Each $c \in C$ is both a join of meets and a meet of joins of elements from the image of L.
(ii) For $S, T \subseteq L$, $\bigwedge \varphi[S] \leq \bigvee \varphi[T]$ iff $\bigwedge S' \leq \bigvee T'$ for some finite $S' \subseteq S, T' \subseteq T$.

Further, if $\varphi' : L \to C'$ is another completion satisfying these two properties, there is a unique isomorphism $\mu : C \to C'$ with $\mu \circ \varphi = \varphi'$.

Proof. We provide a sketch, details are found in [14]. Let $\mathcal{I}$ and $\mathcal{F}$ be the sets of ideals and filters of L and define a binary relation R from $\mathcal{F}$ to $\mathcal{I}$ by setting $F\,R\,I$ iff $F \cap I \neq \emptyset$. Then the polars of R [8] give a Galois connection between the power set of $\mathcal{F}$ and the power set of $\mathcal{I}$. The Galois closed elements of the power set of $\mathcal{F}$ form a complete lattice C, and the map $\varphi : L \to C$ defined by $\varphi(a) = \{F \in \mathcal{F} : a \in F\}$ is the required embedding. This gives existence, uniqueness is not difficult. ∎

This completion is called the canonical completion. The embedding φ preserves all finite joins and meets, so is a lattice embedding, but destroys all existing essentially infinite joins and meets. Canonical completions have their origins in Stone duality. For a Boolean algebra B the canonical completion is the natural embedding of B into the power set of its Stone space, for a distributive lattice it is given by the upsets of the Priestley space, and for general lattices it is given by the stable subsets of the Urquhart space [32]. An abstract characterization similar to that above was given in the Boolean case by Jónsson and Tarski [24], and in the distributive case by Gehrke and Jónsson [16]. The above abstract characterization in the lattice setting was given by Gehrke and Harding [14].

3.1 A General Template for Completions

The technique used in constructing the canonical completion can be adapted to create a range of completions. Let P be a poset, $\mathcal{I}$ be some set of order ideals of P containing all principal ideals, and $\mathcal{F}$ some set of order filters containing all principal filters. Define a relation R from $\mathcal{F}$ to $\mathcal{I}$ by $F\,R\,I$ iff $F \cap I \neq \emptyset$. Then the polars of R give a Galois connection, and the Galois closed subsets of $\mathcal{F}$ form a complete lattice $\mathcal{G}(\mathcal{F}, \mathcal{I})$. The map $\alpha : P \to \mathcal{G}(\mathcal{F}, \mathcal{I})$ defined by $\alpha(a) = \{F \in \mathcal{F} : a \in F\}$ is an embedding.

For sets $\mathcal{I}$ and $\mathcal{F}$ of order ideals and order filters, the completion $\alpha : P \to \mathcal{G}(\mathcal{F}, \mathcal{I})$ has the property that each element of $\mathcal{G}(\mathcal{F}, \mathcal{I})$ is both a join of meets and a meet of joins of elements of the image of P. Further, the embedding α

preserves all existing joins in P under which each member of $\mathcal{I}$ is closed, and all existing meets in P under which each member of $\mathcal{F}$ is closed. It destroys all other joins and meets.

A number of common completions arise this way. The order ideal completion arises by choosing $\mathcal{I}$ to be all order ideals of P and $\mathcal{F}$ to be all principal filters of P; the ideal completion by choosing $\mathcal{I}$ to be all ideals, $\mathcal{F}$ to be all principal filters; the MacNeille completion by choosing $\mathcal{I}$ to be all normal ideals, $\mathcal{F}$ to be all principal filters, or alternately, by choosing $\mathcal{I}$ to be all principal ideals and $\mathcal{F}$ all principal filters; and the canonical completion by choosing $\mathcal{I}$ to be all ideals, and $\mathcal{F}$ to be all filters. Clearly others are possible as well.

3.2 Extending Additional Operations

Suppose P is a poset, $\alpha : P \to C$ is a completion of P, and f is a monotone n-ary operation on P. We recall monotone means that f preserves or reverses order in each coordinate. For convenience we write $\overline{a}$ for an n-tuple of elements $(a_1, \ldots, a_n)$ of P, $\overline{c}$ for an n-tuple of elements $(c_1, \ldots, c_n)$ of C, and $\alpha(\overline{a})$ for $(\alpha(a_1), \ldots, \alpha(a_n))$.

Definition 6. *Let $\leq_f$ be the ordering on C^n defined by $\overline{c} \leq \overline{d}$ if $c_i \leq d_i$ for each i with f order preserving in the i^{th} coordinate, and $d_i \leq c_i$ for all other i.*

We now describe two ways to lift the operation f on P to an operation on C.

Definition 7. *Let f^- and f^+ be the n-ary operations on C defined by*

$$f^-(\overline{c}) = \bigvee\{\alpha(f(\overline{a})) : \alpha(\overline{a}) \leq_f \overline{c}\}.$$

$$f^+(\overline{c}) = \bigwedge\{\alpha(f(\overline{a})) : \overline{c} \leq_f \alpha(\overline{a})\}.$$

We call f^- and f^+ the lower an upper extensions of f.

Proposition 7. *Both f^- and f^+ are monotone maps and with repect to either extension, α is a homomorphism.*

For a join dense completion each $c \in C$ is given by $c = \bigvee\{\alpha(a) : \alpha(a) \leq c\}$. For f unary and order preserving, $f^-(c) = \bigvee\{\alpha(a) : \alpha(a) \leq c\}$. Clearly this is a natural choice of extension. Similarly, for a meet dense completion, f^+ is a natural choice. So for MacNeille completions, both are reasonable choices. In particular instances one may be preferable to the other. For Heyting algebras, extending the Heyting implication $\to$ using the upper extension yields a Heyting algebra, while the lower extension does not.

Canonical completions have neither join nor meet density, however, every element of is a join of meets and a meet of joins of elements of the image. We use this to define extensions of monotone maps suited to this type of completion. Let K be all elements that are meets of elements of the image and O be all elements that are joins of elements of the image. Then $c = \bigvee\{\bigwedge\{\alpha(a) : k \leq a\} : k \leq c \text{ and } k \in K\}$ for each c in the canonical completion, with a similar expression involving a meet of joins and all $c \leq o$ with $o \in O$.

Definition 8. *For f monotone and unary define f^σ and f^π by*

$$f^\sigma(c) = \bigvee\{\bigwedge\{\alpha(f(a)) : k \leq a\} : k \leq c \text{ and } k \in K\}.$$

$$f^\pi(c) = \bigwedge\{\bigvee\{\alpha(f(a)) : a \leq o\} : c \leq o \text{ and } o \in O\}.$$

We call f^σ and f^π the lower and upper canonical extensions of f.

This definition extends in a natural way to monotone n-ary operations, but one must use a mixture of open and closed elements depending on whether the coordinate of f preserves of reverses order. In this generality we have the following.

Proposition 8. *Both f^σ and f^π are monotone maps and with respect to either extension, α is a homomorphism.*

For a completion $\alpha : P \to C$ and a family of monotone operations $(f_i)_I$ on P, a map $\beta : I \to \{-, +, \sigma, \pi\}$ can be used to indicate which extension method to apply to each operation f_i.

Definition 9. *For an ordered structure $(A, (f_i)_I, \leq)$ and map $\beta : I \to \{-, +, \sigma, \pi\}$, define the β-ideal completion, β-MacNeille completion, and β-canonical completion to be the corresponding completion applied to the underlying ordered structure with operations extended in the indicated way.*

This by no means exhausts the range of possible completions, but it does include many of those commonly encountered. Generally one tends to use $-$ extensions for ideal completions, $+$ extensions for filter completions, $-, +$ for MacNeille completions, and σ, π for canonical completions to take advantage of various density properties.

4 Preservation of Identities

We consider the question of when an identity holding in an ordered structure $\mathcal{A}$ holds in a certain type of completion of $\mathcal{A}$. In the case of lattice ordered structures, a natural question becomes when a variety of lattice ordered structures is closed under a certain type of completion. There has been considerable progress in this area in the past twenty five years, but one of the more useful results is still one of the oldest.

Definition 10. *If $\mathcal{A}$ is a lattice with a family of operations that are order preserving in each coordinate, then every identity valid in $\mathcal{A}$ is valid in the ideal lattice completion of $\mathcal{A}$ where the operations are extended by the $-$ extension.*

Proof. As in [10] one shows that for a term $t(x_1, \ldots, x_n)$ and ideals $I_i, \ldots, I_n$ of $\mathcal{A}$, that $t(I_1, \ldots, I_n) = \{b \in A : b \leq t(a_1, \ldots, a_n)$ for some $a_1 \in I_1, \ldots, a_n \in I_n\}$. ∎

While ideal completions work very well with order preserving operations, they work very poorly when an operation has a coordinate where it is order reversing. For the basic case of Boolean algebras, the ideal lattice is hopeless as the ideal completion of a Boolean algebra is Boolean only if it is finite.

The preferred method to complete a Boolean algebra with additional operations is the canonical completion, usually using the σ extension of maps. Here there is an extensive literature, beginning with the work of Jónsson and Tarski [24, 25] in the 1950's, and continuing with the use of Kripke semantics in modal logic (see [7] for a complete account). Primary concern is Boolean algebras with additional operations that preserve finite joins in each coordinate. Such operations are called operators.

Theorem 1. *(Jónsson-Tarski) The canonical completion of a Boolean algebra with operators preserves all identities that do not use the Boolean negation.*

In the 1970's, Sahlqvist [30] generalized this result to apply to equations in which negation occurs, provided they are of a certain form. Usually, these equations are called Sahlqvist equations. While we don't describe the exact form here, we do remark that Sahlqvist terms are the ones that correspond to first order properties of the associated Kripke frame.

Theorem 2. *(Sahlqvist) The canonical completion of a Boolean algebra with operators preserves all Sahlqvist equations.*

Sahlqvist's result was set and proved via Kripke frames, which tied it to the Boolean algebra with operator setting. Jónsson [26] gave an algebraic proof that seems more portable. Gehrke, Nagahashi and Venema [17] used Jónsson's method to extend Sahlqvist's theorem to distributive modal logics, but it remains an open problem to see what portions of this result can be extended to canonical completions in more general settings. We remark that Jónsson and Tarski's original result extends nicely to this setting as described below.

Theorem 3. *(Gehrke-Harding) The canonical completion of a bounded lattice with additional monotone operations preserves all identities involving only operators.*

Note that join is an operator on any lattice, but meet being an operator is equivalent to distributivity. Of course, canonical completions of such lattices with operations also preserve some identities involving order inverting operations, such as those for orthocomplementations, and this provides an advantage for them over ideal completions in such settings. To illustrate the utility of canonical completions in the general setting we have the following.

Theorem 4. *(Gehrke-Harding) Let $\mathcal{K}$ be a class of bounded lattices with additional monotone operations. If $\mathcal{K}$ is closed under ultraproducts and β-canonical completions, where β uses only the extensions σ, π, then the variety generated by $\mathcal{K}$ is closed under β-canonical completions.*

In particular, the variety generated by a single finite lattice with monotone operations is closed under β-canonical completions. The proof of this result requires showing canonical completions work well with homomorphic images, subalgebras and Boolean products [14].

Turning to MacNeille completions, there are a good number of results from different areas stating that a particular variety of interest is closed under MacNeille completions. For instance, the varieties of lattices, Boolean algebras, Heyting algebras, ortholattices, closure algebras, and post algebras are closed under MacNeille completions, although one must be careful about choosing the $+, -$ extension of maps in certain cases.

The first general study of preservation of identities under MacNeille completions was conducted by Monk [29], who showed an analogous theorem to 1 holds for MacNeille completions of Boolean algebras with operators provided the operators preserve all existing joins in each coordinate. Givant and Venema [18] used Jónsson's technique to extend this result and obtain a type of Sahlqvist theorem for preservation of identities for MacNeille completions of Boolean algebras with operators. The key point in their work is the notion of a conjugated map, which plays the role for order preserving operations similar to that of residuation for order inverting ones. Among their results is the following which refers to the $-$ extensions of maps.

Theorem 5. *(Givant-Venema) The MacNeille completion of a Boolean algebra with a family of additional conjugated operators preserves Sahlqvist identities.*

I am not aware of a version of 4 for MacNeille completions, but in the setting of Boolean algebras with operators, a type of converse holds [15]. The following uses the idea from Kripke semantics that a set with a family of relations produces a Boolean algebra with operators consisting of the power set of the set and the operators defined using relational image.

Theorem 6. *(Gehrke-Harding-Venema) If a variety of Boolean algebras with operators is closed under MacNeille completions using the $-$ extensions of maps, then the variety is generated by an elementary class of relational structures.*

There is also an interesting connection between Boolean products and MacNeille completions. In [11] it was shown that if a lattice ordered algebraic structure has a well behaved Boolean product representation, then its MacNeille completion lies in the variety generated by the original. This result was used to show any variety of orthomodular lattices that is generated by its finite height members is closed under MacNeille completions. It can also be used to show that Post algebras are closed under MacNeille completions. We note that these results for orthomodular lattices have implications also for Boolean algebras with operators as every variety of ortholattices can be interpreted in a certain variety of modal algebras [21].

5 Comparing Completions

Here there is unfortunately little known. The main result is found in [15].

Theorem 7. *(Gehrke-Harding-Venema) For a bounded lattice L with additional monotone operations, the canonical completion of L is isomorphic to a sublattice of the MacNeille completion of an ultrapower of L. Here any mixture of σ and π extensions of maps can be used for the canonical completion provided the MacNeille completion uses the corresponding $-$ and $+$ extensions of these maps.*

Proof. The key point in the proof [15] is that every ideal of a sufficiently saturated ultrapower of L is a normal ideal. ∎

This is vaguely reminiscent of a result of Baker and Hales [1] showing that the ideal lattice of a lattice L is isomorphic to a subalgebra of an ultrapower of L. Perhaps other such relationships can be found among various completions.

6 Exploring the Boundaries

In this section, we look at a number of results that point to what may, and what may not, be possible. To begin, it is not the case that every ordered algebraic structure can be embedded into one that is complete and satisfies the same identities as the original. The rationals $\mathbb{Q}$ provide an example of a structure without such a completion as no lattice ordered group can have a largest or least element. For a simple example where even a conditional completion is impossible, consider the variety V of diagonalizable algebras [6]. These modal algebras have an order preserving unary operation f and for each member in V we have $x \leq f(x)$ implies $x = 0$. One then finds some $\mathcal{A} \in V$ with a family $a_1 \leq a_2 \leq \cdots$ where $a_n \leq f(a_{n+1})$. Then in any completion of $\mathcal{A}$ for $x = \bigvee a_n$ we have $x \leq f(x)$. Kowalski and Litak [27] provide a number of varieties sourced in logic that admit no completion.

Modular ortholattices provide another example of a variety admitting no completion, but the only known proof of this relies on Kaplansky's result that every complete modular ortholattice is a continuous geometry, and von Neumann's result that a continuous geometry has a dimension function, two of the deepest results in lattice theory. In contrast to this, it is known that every complemented modular lattice can be embedded into a complete complemented modular lattice, via a method known as the Frink embedding [10] which is a modification of the ideal lattice of the filter lattice. A related question, that remains open, is whether every orthomodular lattice admits a completion.

Moving to the topic of how specific completion methods behave, we first consider the canonical completion. Here it had long been conjectured that every variety of Boolean algebras with operators that is closed under canonical completions is generated by an elementary class of frames. Hodkinson and Venema [19] showed this is not the case, but their counterexample is not finitely based. The question remains open in the finitely based setting. On a related note, the matter of determining whether a finitely based variety of Boolean algebras with operators is closed under canonical completions is an undecidable problem [33].

A simpler problem, but also open, is to determine which varieties of lattices are closed under canonical completions. Here we know every finitely generated variety of lattices is closed, but the variety of modular lattices is not closed under canonical completions.

There are a number of results showing that subvarieties of familiar varieties are not closed under MacNeille completions. The only varieties of lattices closed under MacNeille completions are the trivial variety and the variety of all lattices [20]; for Heyting algebras only the trivial variety, the variety of Boolean algebras, and the variety of all Heyting algebras are closed [5]; and [6] describes the situation for some varieties of closure algebras and derivative algebras. Belardinelli, Jipsen, and Ono have shown [4] that in a certain setting, cut elimination for a logic implies the closure of a corresponding variety under MacNeille completions. So the above result for Heyting algebras explains why so few superintuitionistic logics have cut elimination.

It can be a non-trivial task to determine when the variety generated by a given finite ordered structure $\mathcal{A}$ is closed under β-MacNeille completions for some given method β of extending the operations. Indeed, this is not trivial even for $\mathcal{A}$ being the two-element lattice, or the three-element Heyting algebra. It would be desirable to have a decision process for this problem, if indeed it is even decidable.

One might also ask whether a variety is closed under MacNeille completions in the sense that for each $\mathcal{A} \in V$, the operations on $\mathcal{A}$ can be extended in some manner to the MacNeille completion of $\mathcal{A}$ to produce an algebra in V. Here there are further complications. An example in [6] gives a variety V generated by a four-element modal algebra that is not closed under MacNeille completions using either the $-, +$ extensions of maps, but whose closure under MacNeille completions in this more general sense is equivalent to some weak form of the axiom of choice.

Rather than focusing on MacNeille completions, one may ask more generally whether a variety admits some type of regular completion. In many applications it is the regularity that is of primary interest anyway. In some instances, such as for orthomodular lattices, a variety admitting a regular completion is equivalent to closure under MacNeille completions as any regular completion factors through the MacNeille completion [21]. However, this is not generally the case. The variety generated by the three-element Heyting algebra is not closed under MacNeille completions, but does admit a regular completion [22]. This is the only example of this phenomenon that I know. It would be natural to see if there are other varieties of lattices or Heyting algebras that admit regular completions.

7 Conclusions and Discussion

The topic of completing ordered structures is a broad one with a long history. The results mentioned here focus on one area of this topic, preservation of identities,

and deal only with a fragment of the array of completion methods available. Still, there is reason to believe these results form a basis around which a unified theory can be built, and that this theory addresses questions of concern in many areas of mathematics.

We have a fairly general template for completions, the $\mathcal{I}, \mathcal{F}$-method, that includes many of the completions commonly encountered. This method also points a way to tailor completions to specific need. We have also various methods of extending operations to such completions, and Jónsson's approach to Sahlqvist's theorem may give a portable tool to address preservation of certain types of identities by various types of completions.

We have also seen several techniques occurring repeatedly in our work. These include the use of ultraproducts, which occurs when considering the preservation of identities by canonical extensions, and also in the proof that closure of a variety under MacNeille completions implies closure under canonical extensions. Ultraproducts are also related to the formation of the ideal lattice. Boolean products are another recurring theme. They occur in connection with preservation of identities by canonical extensions, when considering preservation of identities by MacNeille completions, and are also key in constructing a variety that admits a regular completion but is not closed under MacNeille completions.

The use of relational structures or Kripke frames has also been closely tied to our considerations of completions of Boolean algebras with operators, both in the case of canonical completions and MacNeille completions. It would desirable to extend this to the more general setting of bounded lattices with additional operations. There are more general notions of frames in this setting, see for example Gehrke [13]. Also, Harding [23] has a notion of frames involving a set X with additional binary relation P and family of relations R_i satisfying certain conditions, where the relations R_i are used to form operations on the Galois closed subsets of X under the polarity induced by P. No matter which method one uses to create frames, it would seem worthwhile to see the extent to which results can be lifted from the Boolean setting to completions of more general structures.

In sum, it seems an exciting time to be working in this area of completions. A sufficient groundwork has been laid to map out a direction of research, and a number of the tools have been identified. To be sure, there remains much work to be done, with likely more than a few surprises, but one hopes to see considerable progress in the near future.

References

1. Baker, K.A., Hales, A.W.: From a lattice to its ideal lattice. Algebra Universalis 4, 250–258 (1974)
2. Banaschewski, B.: Hüllensysteme und Erweiterungen von Quasi-Ordnungen. Z. Math. Logik Grund. Math. 2, 35–46 (1956)
3. Banaschewski, B., Bruns, G.: Categorical characterization of the MacNeille completion. Arch. Math (Basel) 18, 369–377 (1967)

4. Belardinelli, F., Jipsen, P., Ono, H.: Algebraic aspects of cut elimination. Studia Logica 77(2), 209–240 (2004)
5. Bezhanishvili, G., Harding, J.: MacNeille completions of Heyting algebras. The Houston J. of Math. 30(4), 937–952 (2004)
6. Bezhanishvili, G., Harding, J.: MacNeille completions of modal algebras. The Houston J. of Math. 33(2), 355–384 (2007)
7. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, Cambridge (2001)
8. Birkhoff, G.: Lattice Theory. In: Amer. Math. Soc. Coll. Publ., 3rd edn., vol. 25, Amer. Math. Soc., Providence (1967)
9. Burris, S., Sankappanavar, H.P.: A Course in Universal Algebra. Springer, Heidelberg (1981)
10. Crawley, P., Dilworth, R.P.: Algebraic Theory of Lattices. Prentice-Hall, New Jersey (1973)
11. Crown, G.D., Harding, J., Janowitz, M.F.: Boolean products of lattices. Order 13(2), 175–205 (1996)
12. Dedekind, R.: Essays on the Theory of Numbers. Dover, New York (1963)
13. Gehrke, M.: Generalized Kripke frames. Studia Logica 84(2), 241–275 (2006)
14. Gehrke, M., Harding, J.: Bounded lattice expansions. J. Algebra 238, 345–371 (2001)
15. Gehrke, M., Harding, J., Venema, Y.: MacNeille completions and canonical extensions. Trans. Amer. Math. Soc. 358(2), 573–590 (2005)
16. Gehrke, M., Jónsson, B.: Bounded distributive lattice expansions. Math. Scand. 94, 13–45 (2004)
17. Gehrke, M., Nagahashi, H., Venema, Y.: A Sahlqvist theorem for distributive modal logic. Ann. Pure Applied Logic 131(1–3), 65–102 (2005)
18. Givant, S., Venema, Y.: The preservation of Sahlqvist equations in completions of Boolean algebras with operators. Algebra Universalis 41, 47–84 (1999)
19. Goldblatt, R., Hodkinson, I., Venema, Y.: Erdös graphs resolve Fine's canonicity problem. Bull. of Symbolic Logic 10, 186–208 (2004)
20. Harding, J.: Any lattice can be regularly embedded into the MacNeille completion of a distributive lattice. The Houston J. of Math. 19, 39–44 (1993)
21. Harding, J.: Canonical completions of lattices and ortholattices. Tatra Mountains Math. Publ. 15, 85–96 (1998)
22. Harding, J.: A regular completion for the variety generated by the three-element Heyting algebra. The Houston J. of Math (to appear)
23. Harding, J.: Unpublished manuscript
24. Jónsson, B., Tarski, A.: Boolean algebras with operators I. Amer. J. of Math. 73, 891–939 (1951)
25. Jónsson, B., Tarski, A.: Boolean algebras with operators II. Amer. J. of Math. 74, 127–162 (1952)
26. Jónsson, B.: On the canonicity of Sahlqvist identities. Studia Logica 53(4), 473–491 (1994)
27. Kowalski, T., Litak, T.: Completions of GBL algebras: Negative results. Algebra Universalis (to appear)
28. MacNeille, H.M.: Partially ordered sets. Trans. Amer. Math. Soc. 42, 416–460 (1937)
29. Monk, J.D.: Completions of Boolean algebras with operators. Mathematische Nachrichten 46, 47–55 (1970)

30. Sahlqvist, H.: Completeness and correspondence in the first and second order semantics for modal logic. In: Proceedings of the Third Scandinavian Logic Symposium, Univ. Uppsala, Uppsala, pp. 110–143 (1973), Stud. Logic Found. Math. 82, North-Holland, Amsterdam (1975)
31. Schmidt, J.: Zur Kennzeichnung der Dedekind-MacNeilleschen Hülle einer geordneten Menge. Archiv. d. Math. 7, 241–249 (1956)
32. Urquhart, A.: A topological representation theory for lattices. Algebra Universalis 8, 45–58 (1978)
33. Venema, Y.: Personal communication

The Algebra of Truth Values of Type-2 Fuzzy Sets: A Survey

Carol L. Walker and Elbert A. Walker

Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003
hardy@nmsu.edu, elbert@nmsu.edu

Summary. Type-2 fuzzy sets have come to play an increasingly important role in both applications and in the general theory of fuzzy sets. The basis of type-2 fuzzy sets is a certain algebra of truth values, as set forth by Zadeh. This paper is a survey of results about this algebra, along with some new material.

1 Introduction

Type-2 fuzzy sets were introduced by Zadeh in 1975 [26]. The basic idea is to generalize the notion of ordinary fuzzy sets (type-1), affording wider applicability. The fundamental object for a "fuzzy set theory" is an algebra of truth values. For ordinary fuzzy sets, or fuzzy sets of type-1, that algebra, which we denote by $\mathbf{I}$, is the unit interval with the usual max, min, and negation, and constants 0 and 1. A fuzzy subset of a set S is then a mapping from S into this algebra, and operations of these mappings come pointwise from operations on $[0, 1]$, yielding the algebra of fuzzy subsets of S. A common generalization of the algebra $\mathbf{I}$ is an algebra $\mathbf{I}^{[2]}$ consisting of pairs (a, b) with $0 \leq a \leq b \leq 1$, and with appropriate coordinate operations. Mappings $f : S \to \mathbf{I}^{[2]}$ are interval-valued fuzzy sets, which have come to play a big role in applications. Type-2 fuzzy sets are mappings into the set $Map([0, 1], [0, 1])$ of fuzzy subsets of the unit interval. The usual operations put on these mappings to make it into an appropriate algebra are certain convolutions of the operations of $\mathbf{I}$, as proposed by Zadeh, resulting in an algebra $\mathbf{M}$ of truth values [26]. Mappings $f : S \to \mathbf{M}$ are fuzzy sets of S of type-2.

The algebra $\mathbf{M}$ is a rather complicated object. However, it is entirely appropriate for generalizations of type-1 and interval-valued fuzzy sets, and has been investigated in many papers. This paper is a survey of some of the results of these investigations, along with a few new results. The material here is drawn heavily from the papers [2], [3], [4], [5], [19], [20], [21], [22].

2 Type-1 Fuzzy Sets

The fundamental object for a "fuzzy set theory" is a algebra of truth values. For ordinary fuzzy sets, or fuzzy sets of type-1, that algebra is

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 245–255, 2008.
springerlink.com

$$\mathbf{I} = ([0,1], \vee, \wedge, ', 0, 1)$$

with operations

$$\begin{aligned} x \vee y &= \max\{x, y\} \\ x \wedge y &= \min\{x, y\} \\ x' &= 1 - x \end{aligned}$$

and the nullary operations 0 and 1. The algebra $([0,1], \vee, \wedge, ', 0, 1)$ is a bounded distributive lattice with an involution $'$ that satisfies De Morgan's laws and the Kleene inequality $x \wedge x' \leq y \vee y'$, that is, is a **Kleene algebra**.

3 Interval-Valued Fuzzy Sets

For interval-valued fuzzy sets, the algebra of truth values is

$$\mathbf{I}^{[2]} = ([0,1]^{[2]}, \vee, \wedge, ', (0,0), (1,1))$$

with

$$\begin{aligned} [0,1]^{[2]} &= \{(a,b) : a, b \in [0,1], a \leq b\} \\ (a,b) \vee (c,d) &= (a \vee c, b \vee d) \\ (a,b) \wedge (c,d) &= (a \wedge c, b \wedge d) \\ (a,b)' &= (b', a') \end{aligned}$$

and the nullary operations $(0,0)$ and $(1,1)$. This algebra is a bounded distributive lattice with an involution $'$ that satisfies De Morgan's laws. It is not a Kleene algebra.

4 Type-2 Fuzzy Sets

The algebra of truth values for fuzzy sets of type-2 is much more complicated than those for type-1 and interval-valued ones. The basic set is that of all mappings of $[0,1]$ into $[0,1]$, and the operations are certain convolutions of operations on $[0,1]$, as in the following definition.

Definition 1. *On* $[0,1]^{[0,1]}$*, let*

$$\begin{aligned} (f \sqcup g)(x) &= \sup\{f(y) \wedge g(z) : y \vee z = x\} \\ (f \sqcap g)(x) &= \sup\{f(y) \wedge g(z) : y \wedge z = x\} \\ f^*(x) &= \sup\{f(y) : y' = x\} = f(x') \\ \bar{1}(x) &= \begin{cases} 1 \text{ if } x = 1 \\ 0 \text{ if } x \neq 1 \end{cases} \quad \bar{0}(x) = \begin{cases} 1 \text{ if } x = 0 \\ 0 \text{ if } x \neq 0 \end{cases} \end{aligned}$$

The **algebra of truth values for type-2 fuzzy sets** is

$$\mathbf{M} = (\,[0,1]^{[0,1]}\,, \sqcup, \sqcap, ^*\,, \bar{0}, \bar{1})$$

A **fuzzy subset of type-2 of a set** S is a mapping $f : S \to [0,1]^{[0,1]}$, and operations on the set $\mathcal{F}_2(S)$ of all such fuzzy subsets are given pointwise from the operations in $\mathbf{M}$. Thus we have the **algebra**

$$\mathcal{F}_2(S) = (Map(S, [0,1]^{[0,1]}), \sqcup, \sqcap, ^*\,, \bar{0}, \bar{1})$$

of fuzzy subsets of type-2 of the set S**.**

From the above, it is not clear at all exactly how type-2 generalizes type-1 and interval-valued fuzzy sets. At least, the algebras of truth values of type-1 and of interval-valued fuzzy sets should be subalgebras of **M.** Actually, more is true as we shall see.

Determining the properties of the algebra $\mathbf{M}$ is a bit tedious, but is helped by introducing the auxiliary operations in the definition following.

Definition 2. *For $f \in \mathbf{M}$, let f^L and f^R be the elements of* $\mathbf{M}$ *defined by*

$$f^L(x) = \sup\{f(y) : y \leq x\}$$
$$f^R(x) = \sup\{f(y) : y \geq x\}$$

Note that f^L is pointwise the smallest monotone increasing function larger than f, and similarly, that f^R is pointwise the smallest monotone decreasing function larger than f. The point of this definition is that the operations $\sqcup$ and $\sqcap$ in $\mathbf{M}$ can be expressed in terms these auxiliary operations and of the pointwise max and min of functions, as follows.

Theorem 1. *The following hold for all $f, g \in \mathbf{M}$.*

$$\begin{aligned} f \sqcup g &= \left(f \wedge g^L\right) \vee \left(f^L \wedge g\right) \\ &= (f \vee g) \wedge \left(f^L \wedge g^L\right) \end{aligned}$$

$$\begin{aligned} f \sqcap g &= \left(f \wedge g^R\right) \vee \left(f^R \wedge g\right) \\ &= (f \vee g) \wedge \left(f^R \wedge g^R\right) \end{aligned}$$

Using these auxiliary operations, it is fairly routine to verify the following properties of the algebra $\mathbf{M}$.

Corollary 1. *Let f, g, $h \in \mathbf{M}$. The basic properties of* $\mathbf{M}$ *follow.*

(i) $f \sqcup f = f$; $f \sqcap f = f$
(ii) $f \sqcup g = g \sqcup f$; $f \sqcap g = g \sqcap f$
(iii) $\bar{1} \sqcap f = f$; $\bar{0} \sqcup f = f$
(iv) $f \sqcup (g \sqcup h) = (f \sqcup g) \sqcup h$; $f \sqcap (g \sqcap h) = (f \sqcap g) \sqcap h$

(v) $f \sqcup (f \sqcap g) = f \sqcap (f \sqcup g)$
(vi) $f^{**} = f$
(vii) $(f \sqcup g)^* = f^* \sqcap g^*$; $(f \sqcap g)^* = f^* \sqcup g^*$

Notice that this list does not include the absorption laws or distributive laws. In particular, the algebra $\mathbf{M}$ is not a lattice. The properties of this algebra are investigated in detail in [19].

The following problems seem of some interest.

Problem 1. Does $\mathbf{M}$ satisfy any equation not a consequence of these equations? That is, are these equations an equational base for the variety generated by $\mathbf{M}$?

Problem 2. Is the variety generated by $\mathbf{M}$ generated by a finite algebra?

The answers to this last problem are "yes" for $\mathbf{I}$ and $\mathbf{I}^{[2]}$.

The algebra $\mathbf{M}$ contains copies of $\mathbf{I}$ and $\mathbf{I}^{[2]}$, which are Kleene and De Morgan algebras, respectively. We indicate that now. For $a \in [0,1]$, let $\overline{a}$ denote the characteristic function of $\{a\}$, that is, $\overline{a}\,(a) = 1$ and $\overline{a}\,(b) = 0$ for $b \neq a$. The map

$$\mathbf{I} \to \mathbf{M} : a \mapsto \overline{a}$$

is a monomorphism. Observe, for example, that $\overline{a} \sqcap \overline{b} = \overline{a \wedge b}$, and $\overline{a} \sqcup \overline{b} = \overline{a \vee b}$. Thus the image of this map is a subalgebra of $\mathbf{M}$ isomorphic to $\mathbf{I}$. We will identify $\mathbf{I}$ with its image and say $\mathbf{I} \subset \mathbf{M}$. Note also that $\mathbf{I} \subset \mathbf{I}^{[2]}$. The map

$$\mathbf{I}^{[2]} \to \mathbf{M} : (a,b) \mapsto \overline{a}^L \wedge \overline{b}^R$$

is also a monomorphism, and identifying both $\mathbf{I}$ and $\mathbf{I}^{[2]}$ with their images, we have

$$\mathbf{I} \subset \mathbf{I}^{[2]} \subset \mathbf{M}$$

We are saying more than that fuzzy sets are special cases of interval-valued sets and of type-2 sets, and that interval-valued ones are special cases of type-2 fuzzy sets. The inclusions are as subalgebras. But significantly more is true, as will be pointed out in the next section.

5 Automorphisms

The algebras $\mathbf{I}$ and $\mathbf{I}^{[2]}$ are subalgebras of $\mathbf{M}$. One aim of this section is to point out that they are **characteristic subalgebras** of $\mathbf{M}$. This means that automorphisms of $\mathbf{M}$ induce automorphisms of these subalgebras, so that they are very special subalgebras. Intuitively, $\mathbf{M}$ contains no subalgebra isomorphic to $\mathbf{I}$ sitting in $\mathbf{M}$ in the same way, and similarly for $\mathbf{I}^{[2]}$.

Definition 3. *An automorphism of an algebra is a one-to-one map of the algebra onto itself that preserves the operations.*

Here we will limit ourselves to automorphisms of the algebra $\mathbb{M} = \left([0,1]^{[0,1]}, \sqcup, \sqcap, \bar{0}, \bar{1}\right)$, that is, to the algebra $\mathbf{M}$ without the negation operation *. Similarly, $\mathbb{I}$ and $\mathbb{I}^{[2]}$ will denote the algebras $\mathbf{I}$ and $\mathbf{I}^{[2]}$ without the negations. It turns out to be technically advantageous to leave off the negation operation in the initial investigation of the automorphisms of $\mathbf{M}$. Thus the automorphisms of $\mathbb{M}$ are the one-to-one maps φ of $[0,1]^{[0,1]}$ onto itself such that

(i) $\varphi(f \sqcup g) = \varphi(f) \sqcup \varphi(g)$; $\varphi(f \sqcap g) = \varphi(f) \sqcap \varphi(g)$
(ii) $\varphi(\overline{0}) = \overline{0}$; $\varphi(\overline{1}) = \overline{1}$

The automorphisms of any algebra $\mathbb{A}$ form a group $Aut(\mathbb{A})$ under composition of maps. It is easy to see that the automorphisms of $\mathbb{I}$ are the strictly monotone increasing maps of $[0,1]$ onto itself. In [3], it is shown that the automorphisms of $\mathbb{I}^{[2]}$ are of the form $(a,b) \to (\varphi(a), \varphi(b))$, where φ is an automorphism of $\mathbb{I}$.

We want to determine the group of automorphisms of $\mathbb{M}$. Two basic automorphisms of $\mathbb{M}$ arise from an automorphism of $\mathbb{I}$. For $\alpha \in Aut(\mathbb{I})$, and $f \in [0,1]^{[0,1]}$, let

$$\alpha_L(f) = \alpha f$$
$$\alpha_R(f) = f\alpha$$

Proposition 1. *For $\alpha \in Aut(\mathbb{I})$, both α_L and α_R are automorphisms of*

$$\mathbb{M} = ([0,1]^{[0,1]}, \sqcup, \sqcap, \overline{0}, \overline{1}).$$

It turns out [21] that all the automorphisms of $\mathbb{M}$ are of the form $\alpha_L\beta_R$. and that $Aut(\mathbb{M}) \approx Aut(\mathbb{I}) \times Aut(\mathbb{I})$. We also get the following theorems.

Theorem 2. *The algebras $\mathbb{I}$ and $\mathbb{I}^{[2]}$ are characteristic subalgebras of $\mathbb{M}$ [20].*

Since $\mathbf{I}$ and $\mathbf{I}^{[2]}$ are subalgebras of $\mathbf{M}$, it also follows that

Theorem 3. *The algebras $\mathbf{I}$ and $\mathbf{I}^{[2]}$ are characteristic subalgebras of $\mathbf{M}$.*

These facts have been determined by identifying the join irreducibles and the meet irreducibles of $\mathbb{M}$ and of some of its subalgebras [20, 21]. Join irreducibles are those elements f such that $f = g \sqcup h$ implies that $f = g$ or $f = h$. Meet irreducibles are those f such that $f = g \sqcap h$ implies that $f = g$ or that $f = h$. Irreducible elements are those that are both join and meet irreducible. Identifying these various irreducibles requires quite detailed calculations. They are important in the study of automorphisms since they are carried onto themselves by every automorphism.

The upshot is that type-2 fuzzy sets, as defined by Zadeh in [26], are, in a strong mathematical sense, an appropriate generalization of type-1 and of interval-valued fuzzy sets. The algebra of fuzzy truth values of type-2 fuzzy sets contains as characteristic subalgebras the truth values algebras of type-1 and of interval-valued fuzzy sets.

6 Some Subalgebras of M

The algebra **M** has a number of quite special subalgebras besides the copies of **I** and $\mathbf{I}^{[2]}$ it contains. Some of these subalgebras could possibly serve as the algebra of truth values for fuzzy set theories of practical importance. In any case, they are of some mathematical interest.

6.1 The Subalgebra of Convex Normal Functions

One of the most important subalgebras is the subalgebra of normal convex functions.

Definition 4. *A function* $f:[0,1]\to[0,1]$ *is* normal *if* $\sup\{f(x): x\in[0,1]\}=1$. *A function* f *is* convex *if for every* $x\le y\le z$, $f(y)\ge f(x)\wedge f(z)$.

A equivalent condition for f being convex is that $f = f^L \wedge f^R$. The normal functions form a subalgebra of **M,** as do the convex functions.

Theorem 4. *[13, 19] The subalgebra* **L** *of functions that are both normal and convex is a De Morgan algebra. It contains* **I** *and* $\mathbf{I}^{[2]}$*, and as a lattice, it is maximal in* **M**.

Proposition 2. *The subalgebra* **L** *is characteristic in* **M,** *as are the subalgebras of normal and of convex functions.*

It is natural to wonder whether or not **L** is complete as a lattice. Recently, it has been shown that this it is indeed complete, and a proof is in [8]. There are also some closely related subalgebras of **M** which are complete as lattices, for example, the normal convex upper semicontinuous functions. The latter algebra is a very special type of complete distributive lattice known as a *continuous* distributive lattice. In particular, it is a complete Heyting algebra. Again, the details are in [8].

6.2 The Subalgebra of Subsets

The functions f in $[0,1]^{[0,1]}$ such that $f(x)=0$ or $f(x)=1$ are in natural one-to-one correspondence with the subsets of $[0,1]$. The set S of such functions forms a subalgebra $\mathbf{S}=(S,\sqcup,\sqcap,^*,\overline{0},\overline{1})$ of **M**. However, the operations $\sqcup$ and $\sqcap$ do not correspond to ordinary union and intersection, so the algebra is a bit mysterious. The finite subsets of S, that is, those functions that are 1 at only finitely many places is also a subalgebra. An automorphism of **M** takes a finite set into a finite set with the same number of elements. Both S and this subalgebra are characteristic subalgebras. The algebra S is studied in detail in [23].

6.3 The Subalgebra of Points

This subalgebra consists of those functions that are non-zero at exactly one element of their domain, but can have any value in $(0,1]$ at that element. This

subalgebra **P** is the subject of the paper [24]. It is a generalization of the truth value algebra of type-1 fuzzy sets—the subalgebra of points with value 1, and seems to be a reasonable candidate for applications. For a function in **P**, its support could be viewed as degree of membership, and its value as level of confidence. This generalizes type-1 fuzzy sets, where the "level of confidence" is always 1.

One feature of this subalgebra is that it can be realized as an algebra of pairs of elements from $[0,1]$ with simple pointwise operations, making its study much simpler, avoiding computations with convolutions of mappings. In [24], it is shown that **P** is a characteristic subalgebra of **M** and its automorphism group is computed.

6.4 The Subalgebra of Intervals of Constant Height

This subalgebra consists of those functions from the unit interval into itself whose support is a nonempty closed interval and that are constant on that interval. It is the subject of the paper [25]. The elements of this subalgebra can be represented by triples of points from the unit interval—the two end points of the support interval and the constant value on that interval. The relevant operations turn out to be simple—not requiring computations with convolutions of mappings. This subalgebra is characteristic in **M,** and contains the subalgebras **I**, $\mathbf{I}^{[2]}$, and the subalgebra **P** of points. These facts and many more are detailed in [25].

7 T-Norms on M

In type-1 fuzzy theory, t-norms play an important role. They are associative binary operations on $[0,1]$ that are monotone increasing in each variable and have an identity, and they have been thoroughly studied and used. These can be extended to the algebra **M** by taking the convolution ▲ of a t-norm $\triangle$ on [0,1], that is, by defining

$$(f \blacktriangle g)(x) = \sup\{f(y) \wedge g(z) : y \triangle z = x\}$$

Such an operation has many desirable properties, and t-norms of this type are studied in some detail in [19]. We mention a few facts that seem to be of some interest. Suppose that the t-norm $\triangle$ is continuous.

Proposition 3. *The mapping* $a \to \overline{a}$ *is an isomorphism from the algebra* $([0,1], \vee, \wedge, \triangle, ', 0, 1)$ *into the algebra* $([0,1]^{[0,1]}, \sqcup, \sqcap, \blacktriangle, ^*, \bar{0}, \bar{1})$*, that is, from the algebra* $(\mathbf{I}, \triangle)$ *into the algebra* $(\mathbf{M}, \blacktriangle)$*.*

Another fact is that ▲ restricted to the copy of $\mathbf{I}^{[2]}$ in **M** induces a t-norm on $\mathbf{I}^{[2]}$ as we defined t-norms there [3], in particular, they distribute over the meet and join of $\mathbf{I}^{[2]}$. This is not at all obvious, and was a bit of a surprise.

Proposition 4. *The mapping* $(a,b) \to \overline{a}^L \wedge \overline{b}^R$ *is an isomorphism from the algebra* $([0,1]^{[2]}, \vee, \wedge, \triangle, ', 0, 1)$ *into the algebra* $([0,1]^{[0,1]}, \sqcup, \sqcap, \blacktriangle, ^*, \bar{0}, \bar{1})$*.*

The subalgebra **L** of normal convex functions is turning out to be of some importance, and the following proposition could be useful. It proof was furnished to us by Professor Vladik Kreinovich, and appears in [18].

Proposition 5. *If f and g are in* **L** *and the t-norm $\triangle$ is continuous, then $f \blacktriangle g$ is in* **L.**

8 Finite Type-2 Fuzzy Sets

The basic algebraic properties of **M** depend principally on the fact that $[0,1]$ is a complete chain, so this algebra lends itself to various generalizations. One special case is where each of the two copies of $[0,1]$ is replaced by a finite chain, say of lengths m and n. This yields a finite algebra $\mathbf{F}(\mathbf{m^n})$ with basically the same algebraic properties (possibly generating the same variety) as **M**. Again, the normal convex functions form a De Morgan algebra, and a basic question is where do these special De Morgan algebras fit into the world of all finite De Morgan algebras? These De Morgan algebras are characterized as those whose poset of join irreducible elements has a particularly simple structure. This leads to the determination of the automorphism groups of these algebras. Our basic tool is an alternate representation of these algebras, making their operations much more intuitive, and avoiding technical computations with convolutions. The details are in [22]. Type-2 fuzzy sets with $\mathbf{F}(\mathbf{m^n})$ as the algebra of truth values are called the *grid method of discretisation* in [6], where they give efficient algorithms for computing type-2 joins and meets based essentially on the formulas in Theorem 1.

The algebra $\mathbf{F}(\mathbf{m^n})$ can actually be realized as a subalgebra of **M,** and its subalgebra of convex normal functions as a subalgebra of **L**. For example, take $X = \left\{0, \frac{1}{n-1}, \frac{2}{n-1}, \ldots, \frac{n-2}{n-1}, 1\right\}$ and $Y = \left\{0, \frac{1}{m-1}, \frac{2}{m-1}, \ldots, \frac{m-2}{m-1}, 1\right\}$. (The even spacing of the elements of X and Y is needed only for preserving the negation.) Then $\mathbf{F}(\mathbf{m^n})$ can be identified with the set of functions $f \in \mathbf{M}$ such that $f(a) \in Y$ for all a, and $f(a) = 0$ if $a \notin X$, and this correspondence preserves the operations. These algebras are not characteristic subalgebras of **M**. The finite character is preserved under automorphisms, but the domain and range will not stay fixed. Each $\mathbf{F}(\mathbf{m^n})$ is a subalgebra of the characteristic subalgebra of **M** consisting of all functions with finite support.

9 Miscellany

As usual, $Map(X, Y)$ denotes the set of all functions from the set X into the set Y. For a set S, the set of its fuzzy subsets of type-2 is the set $Map(S, Map([0,1],[0,1])$. There are many forms in which this can be written. That is, there are natural one-to-one correspondences between the set $Map(S, Map([0,1],[0,1])$ and many similar expressions. We list a number of them below. Since they are not peculiar to the set $[0,1]$, we consider $Map(S, Map(A, B))$ for any sets S, A, and B. Technically,

the following are isomorphic in the category of sets. The one-to-one correspondences between them are easy to see and are standard mathematical facts.

$$\begin{aligned} Map(S, Map(A,B)) &\approx Map(\bigcup_{s\in S}\{s\}, Map(A,B)) \\ &\approx \prod_{s\in S} Map(\{s\}, Map(A,B)) \approx \prod_{s\in S} Map(A,B) \end{aligned}$$

$$\begin{aligned} Map(S, Map(A,B)) &\approx Map(S\times A, B) \approx Map(A, Map(S,B)) \\ &\approx Map(\bigcup_{a\in A}\{a\}, Map(S,B)) \\ &\approx \prod_{a\in A} Map(\{a\}, Map(S,B)) \approx \prod_{a\in A} Map(S,B) \end{aligned}$$

The isomorphism $Map(S, Map(A,B)) \approx Map\,(S\times A, B)$ says that fuzzy subsets of S of type-2 may be viewed as type-1 fuzzy subsets of $S\times A$, i.e., every type-2 fuzzy set can be viewed as a type-1 fuzzy set by changing the universe. Similar interpretations can be made for the other expressions. It should be noted that $\prod_{s\in S} Map(\{s\}, Map(A,B))$ is *not* equivalent to $\bigcup_{s\in S} Map(\{s\}, Map(A,B))$.

This latter expression is equivalent to $\bigcup_{s\in S} Map(A,B)$, which, in general, is much smaller than the product $\prod_{s\in S} Map(\{s\}, Map(A,B))$. Technically, the coproduct (in this case, disjoint union) $\bigcup_{s\in S}\{s\}$ comes outside as a product.

Now let $f \in Map(S\times A, B)$, and for $a\in A$, let f_a be the restriction of f to $S\times\{a\}$. Then $f_a = \{((s,a), f(s,a)) : s\in S\}$ is a type-1 set, and, viewing $f\in Map(S, Map(A,B))$, the maps f and f_a have essentially the same universe S. But $f = \bigcup_{a\in A} f_a$, so a type-2 fuzzy set is the union of type-1 fuzzy sets with essentially the same universe.

In [12] Mendel and John state a representation theorem for finite type-2 fuzzy sets in terms of type-1 fuzzy sets. They make the following definition.

Definition 5. *Let $\tilde{A} : X \to [0,1]^J$ where $J\subseteq[0,1]$. Let J_x denote the support of $\tilde{A}(x)$. An **embedded type-2 set of** $\tilde{A}$ is a type-1 function $\tilde{A}_e : X\to[0,1]$ defined by $\tilde{A}_e(x) = \tilde{A}(x)(f(x))$, where $f : X\to J$ such that $f(x)\in J_x$ for all $x\in X$.*

Theorem 5 (Representation Theorem). *A type-2 fuzzy set $\tilde{A}$ is the union of all of its embedded type-2 sets $\tilde{A}_e$.*

Their proof relies heavily on their assumption that the domains of all functions involved are finite. However, the theorem is true in complete generality and stated in this generality the proof is transparent. The type-2 fuzzy set $\tilde{A} : X\to[0,1]^J$ can be identified with the set of triples $\tilde{A} = \left\{\left(x, u, \tilde{A}(x)(u)\right) : x\in X,\, u\in J_x\right\}$. Let $E = \{f : X\to J \mid f(x)\in J_x\}$. Then

each embedded type-2 set of $\tilde{A}$ can be identified with a set of triples $\tilde{A}_e^f = \left\{\left(x, f(x), \tilde{A}(x)(f(x))\right) : x \in X\right\}$ for some $f \in E$. To see that

$$\begin{aligned}\tilde{A} &= \left\{\left(x, u, \tilde{A}(x)(u)\right) : x \in X,\, u \in J_x\right\} \\ &= \bigcup_{f\in E}\left\{\left(x, f(x), \tilde{A}(x)(f(x))\right) : x \in X\right\} = \bigcup_{f\in E}\tilde{A}_e^f\end{aligned}$$

just observe that both sets contain the same triples. The fact that the same triple may appear multiple times on the right side causes no problem, by the very nature of set unions.

Note that under the isomorphism

$$Map(X, Map(J, [0,1])) \approx \prod_{u\in J} Map(X, [0,1])$$

in the previous paragraphs, a type-2 fuzzy set $\tilde{A} \in Map(X, Map(J, [0,1]))$ corresponds to the family of maps $\{A_u : X \to [0,1]\}_{u\in J}$ given by $A_u(x) = \tilde{A}(x)(u)$. This is the precisely the essence of the Representation Theorem above.

10 Conclusions

The notion of type-2 fuzzy sets leads to many interesting theoretical results. Many of these same results are also found in some form in applications of type-2 fuzzy sets. But some of these results are difficult to ferret out, and we have endeavored in various papers to give a precise mathematical treatment of them, using standard mathematical notation. We hope this will be of benefit to both theoreticians and those applying fuzzy set theory. In the meantime, we will continue our effort to understand the mathematical underpinnings of the various applications.

References

1. Dubois, D., Prade, H.: Operations in a fuzzy-valued logic. Inform. and Control 43, 224–240 (1979)
2. Gehrke, M., Walker, C., Walker, E.: De Morgan Systems on the Unit Interval. International Journal of Intelligent Systems 11, 733–750 (1996)
3. Gehrke, M., Walker, C., Walker, E.: Some Comments on Interval Valued Fuzzy Sets. International Journal of Intelligent Systems 11, 751–759 (1996)
4. Gehrke, M., Walker, C., Walker, E.: A Mathematical Setting for Fuzzy Logics. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 5, 223–238 (1997)
5. Gehrke, M., Walker, C., Walker, E.: Algebraic aspects of fuzzy sets and fuzzy logic. In: Proceedings of the Workshop on Current Trends and Developments in Fuzzy Logic, Thessaloniki, Greece, October 16–20, 1998, pp. 101–170 (1999)
6. Greenfield, S., John, R.: Optimised Generalised Type-2 Join and Meet Operations. In: Proceedings of 2007 IEEE International Conference on Fuzzy systems, London, UK, July 23–26, 2007, pp. 141–146 (2007)

7. John, R.: Type-2 fuzzy sets: An appraisal of theory and applications. Int. J. Uncertainty, Fuzziness Knowledge-Based Systems 6(6), 563–576 (1998)
8. Harding, J., Walker, C., Walker, E.: Lattices of Convex Normal Functions. Fuzzy Sets and Systems (to appear)
9. Karnik, N., Mendel, J.: Operations on type-2 fuzzy sets. Fuzzy Sets and Systems 122, 327–348 (2001)
10. Klement, P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, Dordrecht (2000)
11. Mendel, J.M.: Uncertain Rule-Based Fuzzy Logic Systems. Prentice Hall PTR, Upper Saddle River (2001)
12. Mendel, J., John, R.: Type-2 Fuzzy Sets Made Simple. IEEE Transactions on Fuzzy Systems 10(2), 117–127 (2002)
13. Mizumoto, M., Tanaka, K.: Some Properties of Fuzzy Sets of Type-2. Information and Control 31, 312–340 (1976)
14. Mizumoto, M., Tanaka, K.: Fuzzy sets of type-2 under algebraic product and algebraic sum. Fuzzy Sets and Systems 5, 277–290 (1981)
15. Mukaidono, M.: Algebraic Structures of Truth Values in Fuzzy Logic. In: Driankov, D., L. Ralescu, A., Eklund, P.W. (eds.) IJCAI-WS 1991. LNCS, vol. 833, pp. 15–21. Springer, Heidelberg (1994)
16. Emoto, M., Mukaidono, M.: Necessary and Sufficient Conditions for Fuzzy Truth Values to Form a De Morgan Algebra. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 7(4), 309–316 (1999)
17. Nieminen, J.: On the Algebraic Structure of Fuzzy Sets of Type 2. Kybernetika 13, 261–273 (1977)
18. Nguyen, H., Walker, E.: A First Course in Fuzzy Logic. Chapman & Hall/CRC, Boca Raton (2000)
19. Walker, C., Walker, E.: The Algebra of Fuzzy Truth Values. Fuzzy Sets and Systems 149(2), 309–347 (2005)
20. Walker, C., Walker, E.: Automorphisms of the Algebra of Fuzzy Truth Values. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 11, 711–732 (2006)
21. Walker, C., Walker, E.: Automorphisms of the Algebra of Fuzzy Truth Values II (preprint)
22. Walker, C., Walker, E.: Finite Algebras of Fuzzy Truth Values (in preparation)
23. Walker, C., Walker, E.: Sets With Type-2 Operations. International Journal of Approximate Reasoning (to appear)
24. Walker, C., Walker, E.: Points With Type-2 Operations. In: Melin, P., et al. (eds.) IFSA 2007. LNCS (LNAI), vol. 4529, Springer, Heidelberg (2007)
25. Walker, C., Walker, E.: Type-2 Intervals of Constant Height. In: Proceedings of the NAFIPS Conference, June 2007, San Diego (2007)
26. Zadeh, L.: The concept of a linguistic variable and its application to approximate reasoning. Inform Sci. 8, 199–249 (1975)

Some Properties of Logic Functions over Multi-interval Truth Values

Noboru Takagi

Department of Intelligent Systems Design Engineering
Toyama Prefectural University
takagi@pu-toyama.ac.jp

Summary. Many types of fuzzy truth values have been proposed, such as numerical truth values, interval truth values, triangular truth values and trapezoid truth values and so on. Recently, different type of fuzzy truth values (which we call multi-interval truth values) have been proposed and discussed. A characteristic feature of multi-interval truth values is that some of the truth values are not convex. This fact makes us difficult to understand algebraic properties on multi-interval truth values. This paper first shows that a set of multi-interval truth values is de Morgan bi-lattices when the conventional logical operations are introduced. Next, this paper focuses on functions over the set of the simplest multi-interval truth values, i.e., we define a multi-interval truth value as a non-empty subset of $\{0, 1, 2\}$. Then, this paper discusses mathematical properties of functions over multi-interval truth values.

1 Introduction

It is well known that binary logic, which focuses on only propositions whose truth values are either one of true or false, is strongly related to Boolean algebras. Then, fuzzy logic was introduced in order to deal with propositions whose truth values are not determined one of true or false. In fuzzy logic, many types of fuzzy truth values have been proposed and studied on their algebraic structures [6, 7]. The simplest and traditional fuzzy truth values are numerical truth values. Then, numerical truth values are expanded into interval truth values, triangular truth values, and trapezoid truth values. It is a characteristic feature of numerical, interval, triangular and trapezoid fuzzy truth values that all of them are convex. It is known that a set of all convex fuzzy truth values together with operations is a de Morgan algebra [3].

In this paper, we focus on a new type of fuzzy truth values, multi-interval truth values. A multi-interval truth value is defined as a collection of interval truth values. Multi-interval truth values were introduced simultaneously by Emoto [10] and Kikuchi [9]. Figure 2 is an illustration of a multi-interval truth value. As you can understand by looking the figure, it is a characteristic feature of multi-interval truth values that some of them are not convex any more. Operations over multi-interval truth values are introduced. They are given by applying Zadeh's

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 256–267, 2008.
springerlink.com

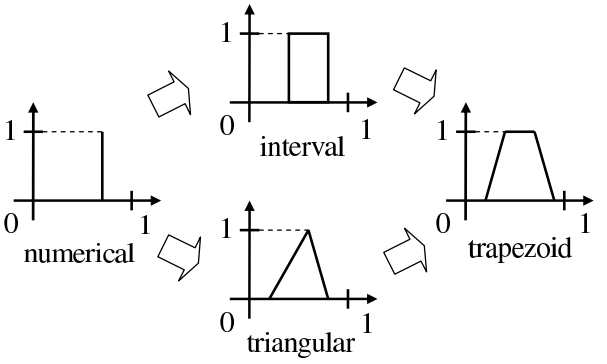


Fig. 1. Relationships over Numerical, Interval, Triangular and Trapezoid Fuzzy Truth Values

extension principle to the conventional logical operations min, max and $x \mapsto 1-x$ over the unit interval $[0, 1]$. This paper then clarifies algebraic properties of multi-interval truth values when the three operations are introduced.

This paper is organized below. In Section 2, multi-interval truth values are introduced, and three operations over multi-interval truth values are defined by using Zadeh's extension principle. Then, it is shown a set of multi-interval truth values is a de Morgan bi-lattices, which is defined as an algebraic system such that there are two distributive lattices and de Morgan's laws connect them to each other. Section 3 proposes functions over multi-interval truth values, which we call multi-interval logic functions. Then, algebraic properties of multi-interval logic functions is discussed. To prove properties, some of binary relations play important roles. Lastly, Section 4 concludes the paper.

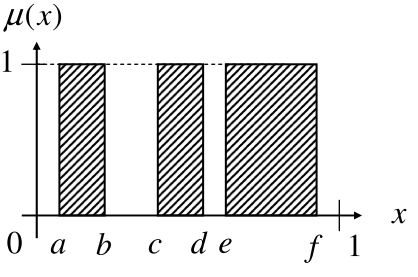


Fig. 2. Illustration of A Multi-Interval Truth Value

2 Multi-interval Truth Values and Basic Properties

Let E_r be the set $\{0, 1, \ldots, r-1\}$ $(r \geq 2)$, and let P_r be a set of all subsets of E_r, but except the empty set, i.e., $P_r = 2^{E_r} - \{\emptyset\}$, where 2^{E_r} is a power set of E_r. We call an element of P_r a multi-interval truth value. When $r =$

3, $P_3 = \{\{0\},\{1\},\{2\},\{0,1\},\{0,2\},\{1,2\},\{0,1,2\}\}$. For simplicity, each multi-interval truth value is expressed by listing its all elements with underline. P_3 is represented as $\{\underline{0},\underline{1},\underline{2},\underline{01},\underline{02},\underline{12},\underline{012}\}$, for example.

The following three logical operations play an important role in fuzzy logic.

$$a \cdot b = \min(a,b) \tag{1}$$
$$a + b = \max(a,b) \tag{2}$$
$$\sim a = (r-1) - a, \tag{3}$$

where a and b are elements of E_r. The logical operations (1)$\sim$(3) are expanded into logical operations over the set of multi-interval truth values P_r by applying Zadeh's extension principle.

Definition 1. Let A and B be any elements of P_r. The logical operations (1)$\sim$(3) are expanded into logical operations $\wedge$, $\sqcup$ and $\overline{\ }$ over multi-interval truth values in the following manner.

$$\lambda_{A\wedge B}(z) = \max_{\substack{x,y\in E_r \\ x\cdot y=z}} \min(\lambda_A(x), \lambda_B(y)), \tag{4}$$
$$\lambda_{A\sqcup B}(z) = \max_{\substack{x,y\in E_r \\ x+y=z}} \min(\lambda_A(x), \lambda_B(y)), \tag{5}$$
$$\lambda_{\bar{A}}(z) = \max_{\substack{x\in E_r \\ \sim x=z}} \lambda_A(x), \tag{6}$$

where λ_A and λ_B are characteristic functions of A and B, respectively. That is, $\lambda_A(x) = 1$ when $x \in A$, $\lambda_A(x) = 0$ otherwise. λ_B is the same with λ_A.

Since characteristic function for every multi-interval truth value takes only 0 and 1, definitions (4)$\sim$(6) can be reformulated into the following formulas.

$$A \wedge B = \{\min(a,b) \mid a \in A \text{ and } b \in B\} \tag{7}$$
$$A \sqcup B = \{\max(a,b) \mid a \in A \text{ and } b \in B\} \tag{8}$$
$$\bar{A} = \{(r-1) - a \mid a \in A\} \tag{9}$$

Tables 1, 2 and 3 are truth tables of the operations $\wedge$, $\sqcup$ and $\overline{\ }$, respectively.

A set of multi-interval truth values P_r with the logical operations $\wedge$, $\sqcup$, $\overline{\ }$ is an algebra [11, 12]. It is easy to show that the following equations are always true in the algebra $(P_r, \wedge, \sqcup, \overline{\ }, \mathbf{0}, \mathbf{1})$, where $\mathbf{0}$ and $\mathbf{1}$ stand for $\underline{0}$ and $\underline{r-1}$, respectively.

S1 $A \wedge A = A$
S2 $A \wedge B = B \wedge A$
S3 $A \wedge (B \wedge C) = (A \wedge B) \wedge C$
S4 $\mathbf{1} \wedge A = A$
S5 $\mathbf{0} \wedge A = \mathbf{0}$
S6 $\overline{(A \wedge B)} = \overline{A} \sqcup \overline{B}$
S7 $\overline{(\overline{A})} = A$

S1' $A \sqcup A = A$
S2' $A \sqcup B = B \sqcup A$
S3' $A \sqcup (B \sqcup C) = (A \sqcup B) \sqcup C$
S4' $\mathbf{1} \sqcup A = \mathbf{1}$
S5' $\mathbf{0} \sqcup A = A$
S6' $\overline{(A \sqcup B)} = \overline{A} \wedge \overline{B}$

The algebra $(P_r, \wedge, \sqcup, \overline{\ }, \mathbf{0}, \mathbf{1})$ is not a lattice when $r \geq 3$ because it does not satisfy the absorption laws

Table 1. Truth Table of $X \wedge Y$

$X\backslash Y$	0	1	2	01	02	12	012
0	0	0	0	0	0	0	0
1	0	1	1	01	01	1	01
2	0	1	2	01	02	12	012
01	0	01	01	01	01	01	01
02	0	01	02	01	02	012	012
12	0	1	12	01	012	12	012
012	0	01	012	01	012	012	012

Table 2. Truth Table of $X \sqcup Y$

$X\backslash Y$	0	1	2	01	02	12	012
0	0	1	2	01	02	12	012
1	1	1	2	1	12	12	12
2	2	2	2	2	2	2	2
01	01	1	2	01	012	12	012
02	02	12	2	012	02	12	012
12	12	12	2	12	12	12	12
012	012	12	2	012	012	12	012

Table 3. Truth Table of $\bar{X}$

X	0	1	2	01	02	12	012
$\bar{X}$	2	1	0	12	02	01	012

$$A \wedge (A \sqcup B) = A \text{ and } A \sqcup (A \wedge B) = A.$$

For example, let $A = \underline{02}$ and $B = \underline{1}$, then $A \wedge (A \sqcup B) = \underline{012} \neq A$ and $A \sqcup (A \wedge B) = \underline{012} \neq A$. Similarly, $(P_r, \wedge, \sqcup, \neg, \mathbf{0}, \mathbf{1})$ does not also satisfy all of the following equalities.

$$\begin{aligned}
&A \wedge (B \sqcup C) = (A \wedge B) \sqcup (A \wedge C)\\
&A \sqcup (B \wedge C) = (A \sqcup B) \wedge (A \sqcup C)\\
&A \wedge \overline{A} = \mathbf{0}\\
&A \sqcup \overline{A} = \mathbf{1}\\
&(A \wedge \overline{A}) \wedge (B \sqcup \overline{B}) = A \wedge \overline{A}\\
&(A \wedge \overline{A}) \sqcup (B \sqcup \overline{B}) = B \sqcup \overline{B}
\end{aligned}$$

Since $\wedge$ is idempotent (S1), commutative (S2) and associative (S3), the algebra $(P_r, \wedge)$ is a lower semilattice. Furthermore, $(P_r, \wedge)$ is both 0- and 1-bounded (i.e., bounded) because $\mathbf{0} \wedge A = \mathbf{0}$ and $\mathbf{1} \wedge A = A$ for all $A \in P_r$. To emphasize that the lower semilattice $(P_r, \wedge)$ has the greatest element $\mathbf{1}$ and the least element $\mathbf{0}$, it is denoted by $(P_r, \wedge, \mathbf{0}, \mathbf{1})$. Similarly, the algebra $(P_r, \sqcup, \mathbf{0}, \mathbf{1})$ is a bounded upper semilattice. Therefore, $(P_r, \wedge, \sqcup, \bar{\ }, \mathbf{0}, \mathbf{1})$ is a de Morgan bisemilattice [8]. Two partial orders are associated with P_r, as follows. For all $X, Y \in P_r$, $X \leq_\wedge Y$ is hold if and only if $X \wedge Y = X$, and $X \leq_\sqcup Y$ is hold if and only if $X \sqcup Y = Y$.

In the following, we discuss some properties of the de Morgan bisemilattice $(P_r, \wedge, \sqcup, \bar{\ }, \mathbf{0}, \mathbf{1})$. As mentioned above, the two partial orders $\leq_\wedge$ and $\leq_\sqcup$ are associated with P_r.

Let A and B be any elements of P_r. Then, $A \wedge B$ and $A \sqcup B$ are calculated by the equations (7) and (8). $A \wedge B$ is also equal to $\{a \in A \mid a \leq b^u\} \cup \{b \in B \mid b \leq a^u\}$, where a^u and b^u are the greatest elements of A and B, respectively. This is because $\{\min(a, b) \mid a \in A\} \subseteq \{\min(a, b^u) \mid a \in A\}$ is hold for every $b \in B$, and $\{\min(a, b) \mid b \in B\} \subseteq \{\min(a^u, b) \mid b \in B\}$ is also hold for every $a \in A$. Therefore, we have

$$\begin{aligned} A \wedge B &= \{\min(a,b) \mid a \in A \text{ and } b \in B\} \\ &= \{a \in A \mid a \leq b^u\} \cup \{b \in B \mid b \leq a^u\}. \end{aligned} \tag{10}$$

Similarly, we also have

$$\begin{aligned} A \sqcup B &= \{\max(a,b) \mid a \in A \text{ and } b \in B\} \\ &= \{a \in A \mid b^\ell \leq a\} \cup \{b \in B \mid a^\ell \leq b\}, \end{aligned} \tag{11}$$

where a^ℓ and b^ℓ are the least elements of A and B, respectively.

Now, we describe on algebraic structures of the lattices $(P_r, \wedge, \vee, \mathbf{0}, \mathbf{1})$ and $(P_r, \sqcup, \sqcap, \mathbf{0}, \mathbf{1})$.

Theorem 1. *Let A and B be elements of P_r. Then, the least upper bound of A and B with respect to $\leq_\wedge$, which is expressed by $A \vee B$, is*

$$A \vee B = (A \cap B) \cup \{a \in A \mid b^u \leq a\} \cup \{b \in B \mid a^u \leq b\},$$

where a^u and b^u are the greatest elements of A and B, respectively. Further, the greatest lower bound of A and B with respect to $\leq_\sqcup$, which is expressed by $A \sqcap B$, is

$$A \sqcap B = (A \cap B) \cup \{a \in A \mid a \leq b^\ell\} \cup \{b \in B \mid b \leq a^\ell\},$$

where a^ℓ and b^ℓ are the least elements of A and B, respectively.

(Proof) We prove only the first equation because the second one can be proved similarly.

Suppose $b^u \leq a^u$, then it is sufficient to show that the least upper bound of A and B is equal to $(A \cap B) \cup \{a \in A \mid b^u \leq a\}$. Let C be $(A \cap B) \cup \{a \in A \mid b^u \leq a\}$. We then have $C \subseteq A$, and further it follows by $b^u \leq a^u$ that $c^u = a^u$. So, we have the following equations.

$$\begin{aligned} A \wedge C &= \{a \in A \mid a \leq c^u\} \cup \{c \in C \mid a \leq a^u\} \\ &= A \cup C \\ &= A \end{aligned}$$

Therefore, we have $A \leq_\wedge C$. Next, let us consider $B \wedge C$. Since $b^u \leq a^u$, $c^u = a^u$, and $C = (A \cap B) \cup \{a \in A \mid b^u \leq a\}$, we have the following equations.

$$\begin{aligned} B \wedge C &= \{b \in B \mid b \leq c^u\} \cup \{c \in C \mid c \leq b^u\} \\ &= B \cup \{c \in C \mid c \leq b^u\} \\ &= B \cup \{a \in A \cap B \mid a \leq b^u\} \cup \{a \in A \mid b^u \leq a \leq b^u\} \\ &= B \cup \{a \in A \cap B \mid a \leq b^u\} \\ &= B \end{aligned}$$

Therefore, we have $B \leq_\wedge C$. So, we have proved that C is an upper bound of A and B with respect to $\leq_\wedge$.

Finally, we will show that C is the least element of all the upper bounds of A and B with respect to $\leq_\wedge$, that is, it is proved that for any element D of P_r

such that $A \leq_\wedge D$ and $B \leq_\wedge D$, $D \leq_\wedge C$ implies $D = C$. It follows by $A \leq_\wedge D$ and $D \leq_\wedge C$ that we have $a^u \leq d^u$ and $d^u \leq c^u$. Thus, the equation $a^u = d^u$ is hold since $a^u = c^u$. So, we have the following equations.

$$\begin{aligned} A \wedge D &= \{a \in A \mid a \leq d^u\} \cup \{d \in D \mid d \leq a^u\} \\ &= A \cup D \\ &= A \end{aligned}$$

Therefore, we have $D \subseteq A$. Similarly, since $b^u \leq a^u$, $a^u = d^u$, and $B \leq_\wedge D$, we have the following equations.

$$\begin{aligned} B \wedge D &= \{b \in B \mid b \leq d^u\} \cup \{d \in D \mid d \leq b^u\} \\ &= B \cup \{d \in D \mid d \leq b^u\} \\ &= B \end{aligned}$$

Therefore, we have $\{d \in D \mid d \leq b^u\} \subseteq B$. Then, it follows by $D \subseteq A$ and $D \subseteq \{d \in D \mid d \leq b^u\} \subseteq B$ that we have $D \subseteq (A \cap B) \cup \{a \in A \mid b^u \leq a\}$, that is, $D \subseteq C$. Furthermore, it also follows by $c^u = d^u$ and $D \leq_\wedge C$ that we have the following equations.

$$\begin{aligned} D \wedge C &= \{d \in D \mid d \leq c^u\} \cup \{c \in C \mid c \leq d^u\} \\ &= D \cup C \\ &= D \end{aligned}$$

This implies that $C \subseteq D$. Therefore, we have proved that $D = C$, which means that C is the least upper bound of A and B with respect to $\leq_\wedge$ when $b^u \leq a^u$. We can prove in a similar way that the first equation of the theorem is hold when $a^u \leq b^u$. □

Theorem 1 can prove the following theorem.

Theorem 2. *For any A, B and C of P_r, $(P_r, \wedge, \vee, \mathbf{0}, \mathbf{1})$ and $(P_r, \sqcap, \sqcup, \mathbf{0}, \mathbf{1})$ satisfy the following equalities S8 ∼ S11 and S8' ∼ S11'.*

S8	$A \wedge (A \vee B) = A$	*S8'*	$A \sqcap (A \sqcup B) = A$
S9	$A \vee (A \wedge B) = A$	*S9'*	$A \sqcup (A \sqcap B) = A$
S10	$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$	*S10'*	$A \sqcap (B \sqcup C) = (A \sqcap B) \sqcup (A \sqcap C)$
S11	$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$	*S11'*	$A \sqcup (B \sqcap C) = (A \sqcup B) \sqcap (A \sqcup C)$

(Proof is omitted.)

By Theorem 2, $(P_r, \wedge, \vee, \mathbf{0}, \mathbf{1})$ and $(P_r, \sqcap, \sqcup, \mathbf{0}, \mathbf{1})$ are distributive lattices. Further, these two distributive lattices are related to each other by de Morganfs laws. So, we define such two distributive lattices with de Morgan's laws are de Morgan bi-lattices. Figure 3 shows Hasse diagrams of the de Morgan bi-lattices $(P_r, \wedge, \vee, \sqcap, \sqcup, \overline{\ }, \underline{0}, \underline{2})$.

Several extensions of fuzzy truth values in fuzzy logic have been discussed [7]. As well known that the set of binary truth values $\{0, 1\}$ forms a Boolean algebra.

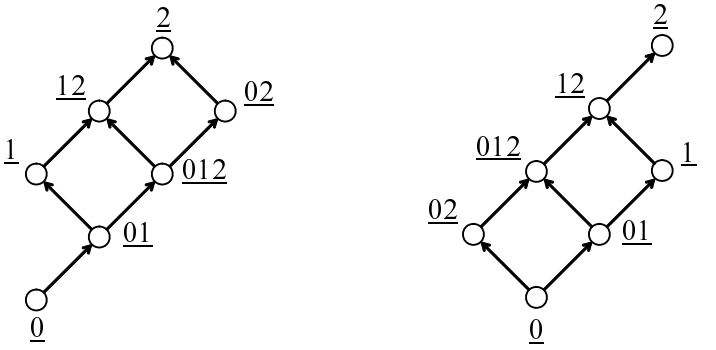


Fig. 3. De Morgan Bi-lattices $(P_3, \wedge, \vee, \sqcap, \sqcup, \bar{}, \underline{0}, \underline{2})$

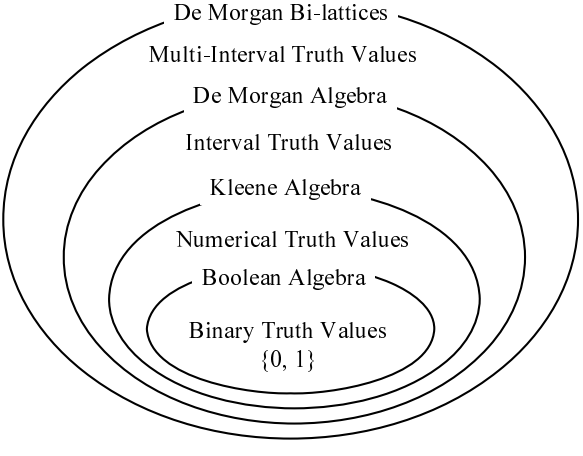


Fig. 4. Relationship of Algebras in Fuzzy Logic

Then, the Boolean algebra is extended into an algebraic system over the unit interval $[0, 1]$. This algebra is known as a Kleene algebra [1, 4, 6]. Interval truth values are one of the extensions of numerical truth values over $[0, 1]$. It is also known that the set of interval truth values over $[0, 1]$ forms a de Morgan algebra [5]. Multi-interval truth values are introduced as a collection of interval truth values, and the set of multi-interval truth values forms a de Morgan bi-lattices. Figure 4 shows the relationship of the algebras.

3 3-Valued Multi-interval Logic Functions

In this section, we discuss properties of logic functions over multi-interval truth values. First, let us define logic formulas.

Definition 2. *Logic formulas are defined inductively in the following manner.*

(1) Constants $\underline{0}$ *and* $\underline{r-1}$ *and variables* $X_1, \ldots, X_n$ *are logic formulas.*
(2) If g *is a logic formula, then* $\bar{g}$ *is also a logic formula.*
(3) If g *and* h *are logic formulas, then* $(g \wedge h)$ *and* $(g \sqcup h)$ *are also logic formulas.*
(4) It is a logic formula if and only if we get it from (1), (2) and (3) in a finite number of steps.

Definition 3. *If a function* $f : P_r^n \to P_r$ *is expressed by a logic formula, then* f *is called an* n*-variable* r*-valued multi-interval logic function.*

First of all, the next theorem can be proved by definitions of logical operations (7)∼(9).

Theorem 3. *Let* f *be a* r*-valued multi-interval logic function. Then,* f *is* $\subseteq$*-monotonic, i.e., for any elements* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *of* P_r^n, $\boldsymbol{A} \subseteq \boldsymbol{B}$ *implies* $f(\boldsymbol{A}) \subseteq f(\boldsymbol{B})$.
(Proof) This theorem is proved by the induction of the number of logical operations. It is evident that constants $\underline{0}$ and $\underline{r-1}$, and variables $X_1, \ldots, X_n$ satisfy the theorem. Suppose g and h satisfy the theorem, that is, for $\boldsymbol{A}$ and $\boldsymbol{B}$ of P_r such that $\boldsymbol{A} \subseteq \boldsymbol{B}$, $g(\boldsymbol{A}) \subseteq g(\boldsymbol{B})$ and $h(\boldsymbol{A}) \subseteq h(\boldsymbol{B})$. Then, we have $\{\min(a,b) \mid a \in g(\boldsymbol{A}) \text{ and } b \in h(\boldsymbol{A})\} \subseteq \{\min(a,b) \mid a \in g(\boldsymbol{B}) \text{ and } b \in h(\boldsymbol{B})\}$. Therefore, $(g \wedge h)(\boldsymbol{A}) \subseteq (g \wedge h)(\boldsymbol{B})$. Similarly, we can prove $(g \sqcup h)(\boldsymbol{A}) \subseteq (g \sqcup h)(\boldsymbol{B})$ and $(\overline{g})(\boldsymbol{A}) \subseteq (\overline{g})(\boldsymbol{B})$. □

In the following, we discuss properties on 3-valued multi-interval logic functions. First, two partial orders are introduced into P_3 below.

Definition 4. *Two binary relations* $\preccurlyeq_1$ *and* $\preccurlyeq_2$ *are defined in* P_3.

(1) $\underline{1} \preccurlyeq_1 \underline{01}$, $\underline{1} \preccurlyeq_1 \underline{12}$, $\underline{1} \preccurlyeq_1 \underline{012}$, $\underline{01} \preccurlyeq_1 \underline{012}$, $\underline{12} \preccurlyeq_1 \underline{012}$, *and* $A \preccurlyeq_1 A$ *for any* $A \in P_3$
(2) $\underline{0} \preccurlyeq_2 \underline{01}$, $\underline{2} \preccurlyeq_2 \underline{12}$, $\underline{02} \preccurlyeq_2 \underline{012}$, *and* $A \preccurlyeq_2 A$ *for any* $A \in P_3$

It is easy to show that two binary relations $\preccurlyeq_1$ and $\preccurlyeq_2$ are partial order relations on P_3. Fig. 6 shows Hasse diagrams of the partial order relations $\preccurlyeq_1$ and $\preccurlyeq_2$.

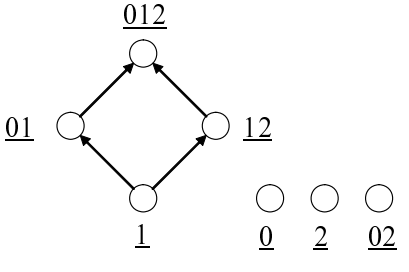


Fig. 5. Partial Order Relation $\preccurlyeq_1$

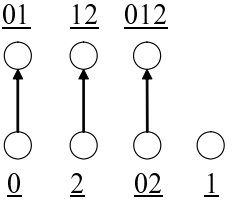


Fig. 6. Partial Order Relation $\preccurlyeq_2$

Theorem 4. *Let f be a 3-valued multi-interval logic function. Then, f is $\preccurlyeq_1$-monotonic and $\preccurlyeq_2$-monotonic, i.e., for any elements $\boldsymbol{A}$ and $\boldsymbol{B}$ of P_3^n, $\boldsymbol{A} \preccurlyeq_1 \boldsymbol{B}$ implies $f(\boldsymbol{A}) \preccurlyeq_1 f(\boldsymbol{B})$ and $\boldsymbol{A} \preccurlyeq_2 \boldsymbol{B}$ implies $f(\boldsymbol{A}) \preccurlyeq_2 f(\boldsymbol{B})$.*

Proof. The theorem can be proved by the induction of the number of operations. □

Let $Q = \{\underline{0}, \underline{1}, \underline{2}, \underline{02}\}$, and let $\boldsymbol{A}$ be an element of P_3^n. Then, a subset $S(\boldsymbol{A})$ is defined as the following subset of Q^n, which depends on the element $\boldsymbol{A}$.

$$S(\boldsymbol{A}) = \{\boldsymbol{B} : \boldsymbol{B} \subseteq \boldsymbol{A} \text{ and } \boldsymbol{B} \in Q^n\}$$

Theorem 5. *Let f be a 3-valued multi-interval logic function. Then, the following equation is true for any element $\boldsymbol{A}$ of P_3^n.*

$$f(\boldsymbol{A}) = \bigcup_{\boldsymbol{B} \in S(\boldsymbol{A})} f(\boldsymbol{B})$$

Proof. First, consider the case where $f(\boldsymbol{A}) = \underline{0}$. Then since it follows by the $\subseteq$-monotonicity of f that $f(\boldsymbol{B}) = \underline{0}$ holds for every element $\boldsymbol{B}$ of $S(\boldsymbol{A})$, the theorem is true when $f(\boldsymbol{A}) = \underline{0}$. It can be proved in a similar way that $f(\boldsymbol{A}) = \bigcup_{\boldsymbol{B} \in S(\boldsymbol{A})} f(\boldsymbol{B})$ when $f(\boldsymbol{A}) = \underline{1}$ or $f(\boldsymbol{A}) = \underline{2}$.

Next, consider the case where $f(\boldsymbol{A}) = \underline{01}$. Then, it follows by the $\preccurlyeq_1$-monotonicity of f that there is an element $\boldsymbol{B}$ of $S(\boldsymbol{A})$ such that $f(\boldsymbol{B}) = \underline{1}$ or $\underline{01}$. It also follows by the $\preccurlyeq_2$-monotonicity of f that there is an element $\boldsymbol{B}$ of $S(\boldsymbol{A})$ such that $f(\boldsymbol{B}) = \underline{0}$ or $\underline{01}$. By the $\subseteq$-monotonicity of f, $f(\boldsymbol{B}) = \underline{0}$, $\underline{1}$ or $\underline{01}$ holds for every element $\boldsymbol{B}$ of $S(\boldsymbol{A})$. Furthermore, it is not true by the $\preccurlyeq_1$-monotonicity and the $\preccurlyeq_2$-monotonicity that $f(\boldsymbol{B}) = \underline{0}$ holds for every element $\boldsymbol{B}$ of $S(\boldsymbol{A})$ and $f(\boldsymbol{B}) = \underline{1}$ holds for every element $\boldsymbol{B}$ of $S(\boldsymbol{A})$. This completes the theorem when $f(\boldsymbol{A}) = \underline{01}$.

Since it can be proved in a similar way that the theorem is satisfied when $f(\boldsymbol{A}) = \underline{02}$, $\underline{12}$, or $\underline{012}$, the theorem is always true. □

The next theorem is easy to prove by Theorem 5.

Theorem 6. *Let g and h be any 3-valued multi-interval logic functions. Then, $g(\boldsymbol{A}) = h(\boldsymbol{A})$ holds for every $\boldsymbol{A}$ of P_3^n if and only if $g(\boldsymbol{A}) = h(\boldsymbol{A})$ holds for every $\boldsymbol{A}$ of Q^n.*

In the remainder of this section, some of necessary conditions for a function over P_3 to be a 3-valued multi-interval logic function.

Let T_1, T_2 and T_3 be $\{\underline{0}, \underline{1}, \underline{2}\}$, $\{\underline{0}, \underline{2}, \underline{02}\}$ and $\{\underline{01}, \underline{12}, \underline{012}\}$, respectively. Then, define binary relations over T_1, T_2 and T_3 below.

Definition 5. *Define binary relations $\preceq_1$, $\preceq_2$ and $\preceq_3$ in the sets T_1, T_2 and T_3, respectively, in the following manner.*

(1) $\underline{0} \preceq_1 \underline{1}$, $\underline{2} \preceq_1 \underline{1}$, *and* $a \preceq_1 a$ *for every* $a \in T_1$.

(2) $\underline{0} \preceq_2 \underline{02}$, $\underline{2} \preceq_2 \underline{02}$, *and* $a \preceq_2 a$ *for every* $a \in T_2$.
(3) $\underline{01} \preceq_3 \underline{012}$, $\underline{12} \preceq_3 \underline{012}$, *and* $a \preceq_3 a$ *for every* $a \in T_3$.

It is easy to show that binary relations $\preceq_1$, $\preceq_2$ and $\preceq_3$ are partial order relations. Hasse diagrams of the partial order relations are given in Fig. 7.

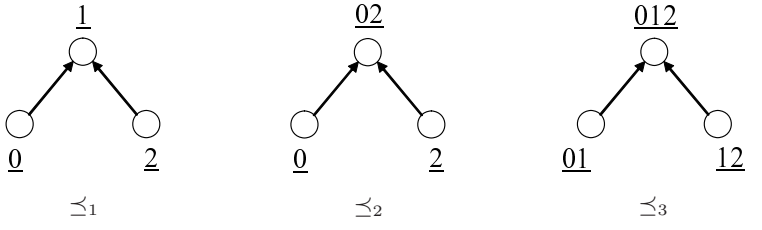


Fig. 7. Partial Order Relations $\preceq_1$, $\preceq_2$ and $\preceq_3$

Theorem 7. *Any 3-valued multi-interval logic function* f *is monotonic in* $\preceq_1$, $\preceq_2$ *and* $\preceq_3$, *i.e.,*

(1) for any elements $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *of* T_1^n, $\boldsymbol{A} \preceq_1 \boldsymbol{B}$ *implies* $f(\boldsymbol{A}) \preceq_1 f(\boldsymbol{B})$,
(2) for any elements $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *of* T_2^n, $\boldsymbol{A} \preceq_2 \boldsymbol{B}$ *implies* $f(\boldsymbol{A}) \preceq_2 f(\boldsymbol{B})$,
(3) for any elements $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *of* T_3^n, *and* $\boldsymbol{A} \preceq_3 \boldsymbol{B}$ *implies* $f(\boldsymbol{A}) \preceq_3 f(\boldsymbol{B})$.

(Proof is omitted.)

Let A be any element of T_1. Then, define elements $\dot{A}$ and $\ddot{A}$ of T_2 and T_3, respectively, in the following manner.

$$\dot{A} = \begin{cases} \underline{02} \text{ if } A = \underline{1} \\ A \text{ otherwise} \end{cases}$$

$$\ddot{A} = \begin{cases} \underline{01} \text{ if } A = \underline{0} \\ \underline{12} \text{ if } A = \underline{2} \\ \underline{012} \text{ if } A = \underline{1} \end{cases}$$

Theorem 8. *Let* f *be any 3-valued multi-interval logic function, and let* $(A_1, \ldots, A_n)$ *be any element of* T_1^n. *Then, we have the following equations.*

(1) $f(\dot{A}_1, \ldots, \dot{A}_n) = \dot{f}(A_1, \ldots, A_n)$
(2) $f(\ddot{A}_1, \ldots, \ddot{A}_n) = \ddot{f}(A_1, \ldots, A_n)$

(Proof is omitted.)

Definition 6. *Let* A *be any element of* Q. *Then, define an element* A' *from* A *in the following manner.*

$$A' = \begin{cases} \underline{02} \textit{ if } A = \underline{1} \\ \underline{1} \textit{ if } A = \underline{02} \\ A \textit{ otherwise} \end{cases}$$

Theorem 9. *Let f be any 3-valued multi-interval logic function, and let $\boldsymbol{A}$ be any element of Q^n. Then, we have the following conditions.*

(1) If $f(\boldsymbol{A}) = \underline{0}, \underline{2}, \underline{01}$, or $\underline{12}$ holds, then $f(\boldsymbol{A}') = f(\boldsymbol{A})$ holds.
(2) If $f(\boldsymbol{A}) = \underline{1}$ holds, then $f(\boldsymbol{A}') = \underline{02}$ or $\underline{012}$ holds.
(3) If $f(\boldsymbol{A}) = \underline{02}$ or $\underline{012}$ holds, then $f(\boldsymbol{A}') = \underline{1}$ holds.

(Proof is omitted.)

4 Conclusion

This paper focused on multi-interval truth values, in which some of them are not convex. First of all, we showed that the algebra $(P_r, \wedge, \sqcup, \overline{\ }, \mathbf{0}, \mathbf{1})$ is a de Morgan bi-lattices. A de Morgan bi-lattices has two distributive lattices, and de Morgan laws connects the two distributive lattices. Next, we discussed multi-interval logic functions when $r = 3$, which are defined as functions over P_3 expressed by logic formulas. One of the main theorems of this paper is that any multi-interval logic function f is determined uniquely if we know all the values $f(\boldsymbol{A})$ for every element $\boldsymbol{A} \in Q^n$. Moreover, some of necessary conditions are also proved.

This is a early trial to study mathematical properties of non-convex fuzzy truth values. So, many open problems still remain; a necessary and sufficient condition for a function over P_3 to be a multi-interval logic functions, canonical forms of multi-interval logic functions, etc. All of them are our future works.

References

1. Balbes, R., Dwinger, P.: Distributive Lattices. University of Missouri Press (1974)
2. Mukaidono, M.: On Some Properties of Fuzzy Logic. Transactions on IECE Japan 58–D(3), 150–157 (1975)
3. Mizumoto, M., Tanaka, K.: Some Properties of Fuzzy Sets of Type 2. Information and Control 31(4), 341–363 (1976)
4. Mukaidono, M.: A set of independent and complete axioms of a fuzzy algebra (Kleene algebra). In: Proc. of 11th International Symposium on Multiple-Valued Logic, p. 81 (1981)
5. Mukaidono, M.: Algebraic structures of truth values in fuzzy logic. In: Driankov, D., L. Ralescu, A., Eklund, P.W. (eds.) IJCAI-WS 1991. LNCS, vol. 833, pp. 1880–1883. Springer, Heidelberg (1994)
6. Mukaidono, M.: Kleene Algebras in Fuzzy Truth Values. In: Proc. of the 4th International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery, pp. 61–78 (1996)
7. Mukaidono, M.: Several extensions of the truth values in fuzzy logic. In: Raś, Z.W., Skowron, A. (eds.) ISMIS 1999. LNCS, vol. 1609, pp. 15–21. Springer, Heidelberg (1999)
8. Brzozowski, J.A.: De Morgan Bisemilattices. In: Proc. of 30th IEEE International Symposium on Multiple-Valued Logic, pp. 173–178 (May 2000)

9. Kikuchi, H., Nakanishi, S.: Closures of fuzzy linguistic truth value with regards to the extension principle. In: Proc. of the 16th Fuzzy System Symposium, pp. 203–206 (in Japanese, 2000)
10. Emoto, M., Mukaidono, M.: The difference between properties of convex and concave fuzzy truth values. In: Proc. of the 16th Fuzzy System Symposium, pp. 193–196 (in Japanese, 2000)
11. Takagi, N., Kikuchi, H., Nakashima, K.: Multi-Interval Truth Valued Logic. In: Proceedings of the IEEE 10th International Conference on Fuzzy Systems, vol. 2, pp. 1119–1122 (2001)
12. Takagi, N., Nakashima, K.: Multi-Interval Truth Valued Logic Functions and Their Properties. Note on Multiple-Valued Logic in Japan 27(9) (in Japanese, 2004)

Possible Semantics for a Common Framework of Probabilistic Logics

Rolf Haenni[1], Jan-Willem Romeijn[2], Gregory Wheeler[3], and Jon Williamson[4]

[1] Bern University of Applied Sciences, Engineering and Information Technology, 2501 Biel, Switzerland
rolf.haenni@bfh.ch
[2] University of Groningen, Faculty of Philosophy, 9712 GL Groningen, The Netherlands
J.W.Romeijn@rug.nl
[3] Universidade Nova de Lisboa, Department of Computer Science, 2829-516 Caparica, Portugal
grw@fct.unl.pt
[4] University of Kent, SECL, Philosophy, Canterbury, CT2 7NF, United Kingdom
j.williamson@kent.ac.uk

Summary. This paper proposes a common framework for various probabilistic logics. It consists of a set of uncertain premises with probabilities attached to them. This raises the question of the strength of a conclusion, but without imposing a particular semantics, no general solution is possible. The paper discusses several possible semantics by looking at it from the perspective of probabilistic argumentation.

1 Introduction

If the premises of a valid logical inference are not entirely certain, how certain is its conclusion? To find an answer to this is an important question, it is necessary to overcome the restrictions and limits of the classical fields of logical and probabilistic inference. This simple observations is not entirely new [3, 4, 6, 9, 22, 30, 26], but attempts of building such unifying *probabilistic logics* (or *logics of probability*) are rather sparse, especially in comparison with the long traditions of logic and probability theory as independent disciplines both in philosophy and in science.[1] Nevertheless, probabilistic logic is nowadays a rapidly developing interdisciplinary research topic with contributions from philosophical logic [1, 8, 17, 19, 31] and Artificial Intelligence [7, 10, 18, 25, 24, 27], but also from mathematics, linguistics, statistics, and decision theory [2, 20]. While it is clear that logic and probability theory are intimately related, the exact shape of this relationship is still the subject of an ongoing debate.

In principle, there are at least two different ways of constructing a combined theory of logical and probabilistic inference, depending on whether logic or

[1] For more information about the historical account of probabilistic logics, we refer to the excellent survey in [17].

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 268–279, 2008.
springerlink.com

probability theory is at its center. The majority of approaches in the literature is logic-centered, either by defining a probability function on the sentences of the logic [9, 30, 25] or by incorporating probabilities into the syntax of the logic [7, 24]. In the theory of probabilistic argumentation [11, 15, 21], where the available knowledge is partly encoded as a set of logical premises and partly as a fully specified probability space, the starting point is neither biased towards logic, nor is it biased towards probability. This setting gets particularly interesting when some of the logical premises include variables that are not contained in the probability space. The two classical questions of the probability and the logical deducibility of a hypothesis can then be replaced by the more general question of the probability of a hypothesis being logically deducible from the premises.

In Section 2, we first propose a neutral common framework for a variety of different probabilistic logics. The framework as such has no particular semantics, but we will shortly discuss what most people would probably consider its "standard semantics". In Section 3, we first give a short summary of the theory of probabilistic argumentation, which then allows us to discuss various semantics for the common framework. Hence, the goal of this paper is to establish a link between probabilistic argumentation and other probabilistic logics via the common framework.

2 Probabilistic Logics

The principal goal of any *probabilistic logic* (sometimes called *probability logic* [1, 16, 31], or *progic* for short) is to combine the capacity of probability theory to handle uncertainty with the capacity of deductive logic to cope with qualitative and structural knowledge such as logical relationships. As most probabilistic logics are constructed on top of an existing logic (propositional logic in the simplest case), probabilities are usually treated as an addendum rather than as an integral part of the theory. In this section, we propose such a simple addendum, in which probabilities (or sets of probabilities) are attached to premises to represent their respective uncertainties. This then raises the question of the extent to which a possible conclusion follows from the uncertain premises. Given the simplicity and generality of the proposed extension, which allows it to be taken as a common unifying umbrella for many existing probabilistic logics, we will refer to as the *progic framework*.

2.1 The Progic Framework

In a classical logic, the fundamental question of interest is whether a conclusion ψ is logically entailed by a given set of premises $\Phi = \{\varphi_1, \ldots, \varphi_n\}$. Logical inference is thus essentially a problem of verifying the entailment relation $\models$ between Φ and ψ. The entailment relation itself is usually defined in terms of a subset relation $\subseteq$ of corresponding sets of truth assignments (models) in the respective logic.

To augment the fundamental question of classical logic towards probabilistic logic, we will now consider a set of premises with probabilities attached to them. In the simplest case, this means that each premise φ_i has an attached probability $x_i \in [0,1]$, but to be as general as possible, we may also allow the case where a set of probabilities $X_i \subseteq [0,1]$ is attached to each premise φ_i. In this augmented setting, which includes the special case of sharp probabilities by $X_i = \{x_i\}$, the traditional question of classical logic turns into a more general question of the form

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^Y, \tag{1}$$

where the set $Y \subseteq [0,1]$ is intended to represent the extent to which the conclusion ψ follows from the premises.[2] This is a very general question, which covers a multitude of frameworks of existing probabilistic logics. We will thus refer to is as the *general progic framework* (or *progic framework* for short). Note that the problem is the determination of the set Y itself, not the verification of the entailment relation for a given Y. Needless to say that the determination of Y is heavily dependent on the semantics imposed by the chosen framework. In the next subsection, we will discuss one of the most straightforward semantics for the progic framework.

2.2 The Standard Semantics

In the so-called *standard semantics* of the progic framework, we consider each attached probability set X_i as a constraint for the probability $P(\varphi_i)$ in a corresponding probability space. For the sake of simplicity, we will restrict the premises to be propositional sentences. Formally, we write $V = \{Y_1, \ldots, Y_r\}$ to denote the set of involved Boolean variables Y_i, each with a set $\Omega_i = \{0,1\}$ of possible values. In the corresponding propositional language $\mathcal{L}_V$, we use propositional symbols y_i as placeholders for $Y_i = 1$. The Cartesian product $\Omega_V = \Omega_1 \times \cdots \times \Omega_r = \{0,1\}^r$ then contains the set of all possible truth assignments of the propositional language, each of which representing a possible (state of the) world. For a given propositional sentence $\varphi \in \mathcal{L}_V$, we write $[\![\varphi]\!] \subseteq \Omega_V$ to denote the set of truth assignments for which φ evaluates to 1 (according to the usual semantics of propositional logic), and we say that φ entails ψ, or that $\varphi \models \psi$ holds, iff $[\![\varphi]\!] \subseteq [\![\psi]\!]$.

To make a connection to probability theory, let Ω_V play the role of a finite sample space. The finiteness of Ω_V allows us to work with the σ-algebra 2^{Ω_V} of all subsets of Ω_V, i.e. we obtain a probability space $(\Omega_V, 2^{\Omega_V}, P)$ for any measure $P : 2^{\Omega_V} \to [0,1]$ that satisfies the Kolmogorov's probability axioms. With $\mathbb{P}$ we denote the set of all such probability measures for a given set of variable V. Note that we adopt the usual notational convention of writing $P(\varphi)$ rather than $P([\![\varphi]\!])$ for the probability of the event $[\![\varphi]\!]$.

According to the above-mentioned general idea of the standard semantics, we consider each set X_i as a constraint $P(\varphi_i) \in X_i$ for the unknown probability

[2] For $X_i = \{1\}$, this general setting degenerates into the classical problem of logical inference.

measure P. Formally, let $\mathbb{P}_i = \{P \in \mathbb{P} : P(\varphi_i) \in X_i\}$ denote the set of all probability measures satisfying the constraint for the i-th premise. The intersection of all these sets, $\mathbb{P}_* = \mathbb{P}_1 \cap \cdots \cap \mathbb{P}_n$, defines then the set of probability measures satisfying all constraints. From this, we obtain with $Y = \{P(\psi) : P \in \mathbb{P}_*\}$ a simple solution for the generalized inference problem of the progic framework. Note that inference according to the standard semantics can be seen as a generalization of classical logical inference, which is concerned with a continuum of truth assignments in form of all possible probability measures.

An important special case of the above setting arises when the attached probability sets X_i are all functionally unrelated intervals, i.e. $X_i = [\ell_i, u_i]$. This means that all sets $\mathbb{P}_i$ are convex, which implies that $\mathbb{P}_*$ is also convex and that Y is again an interval with a lower and an upper bound.[3] The lower and upper bounds of Y are usually denoted by $\underline{P}(\psi) = \min\{P(\psi) : P \in \mathbb{P}_*\}$ and $\overline{P}(\psi) = \max\{P(\psi) : P \in \mathbb{P}_*\}$, respectively. Note that the convexity of $\mathbb{P}_*$ guarantees that $\underline{P}$ and $\overline{P}$ are among the extremal points of $\mathbb{P}_*$. Interestingly, we may obtain an interval for Y even if all sets X_i are singletons. From a computational point of view, we can translate the problem of finding Y according to the standard semantics into a (very large) linear optimization problem, e.g. with three constraints $P(\varphi_i) \geq \ell_i$, $P(\varphi_i) \leq u_i$, and $P(\varphi_i) = \sum_{\omega \in [\![\varphi_i]\!]} P(\{\omega\})$ for all premises [1, 25].

Example 1. To illustrate the standard semantics, consider two premises $(a \wedge b)^{[0,0.25]}$ and $(a \vee \neg b)^{\{1\}}$. For the specification of a probability measure with respect to the corresponding 2-dimensional sample space $\{0,1\}^2$ at least three parameters are needed (the size of the sample space minus 1). This means that the set of all possible probability measures $\mathbb{P}$ can be nicely depicted by a tetrahedron (3-simplex) with maximal probabilities for the state descriptions $a \wedge b$, $a \wedge \neg b$, $\neg a \wedge b$, and $\neg a \wedge \neg b$ at each of its four extremities. This tetrahedron is depicted in Fig. 1, together with the convex sets $\mathbb{P}_1$, $\mathbb{P}_2$, and $\mathbb{P}_*$. The picture also shows that $Y = [0,1]$ is the result for the conclusion a, whereas $Y = [0, 0.25]$ is the result for the conclusion b.

3 Probabilistic Argumentation

The theory of *probabilistic argumentation* [11, 13, 15, 21] is first of all driven by the general idea of putting forward the pros and cons of a hypothesis in question, from which it derives its name. The weights of the resulting logical arguments and counter-arguments are measured by probabilities, which are then turned into (sub-additive) *degrees of support* and (super-additive) *degrees of possibility*. Intuitively, degrees of support measure the presence of evidence supporting the hypothesis, whereas degrees of possibility measure the absence of evidence refuting the hypothesis. For this, probabilistic argumentation is concerned with probabilities of a particular type of event of the form "*the hypothesis is a logical consequence*" rather than "*the hypothesis is true*", i.e. very much like Ruspini's

[3] Convex set of probability measures are sometimes called *credal sets* [5, 23].

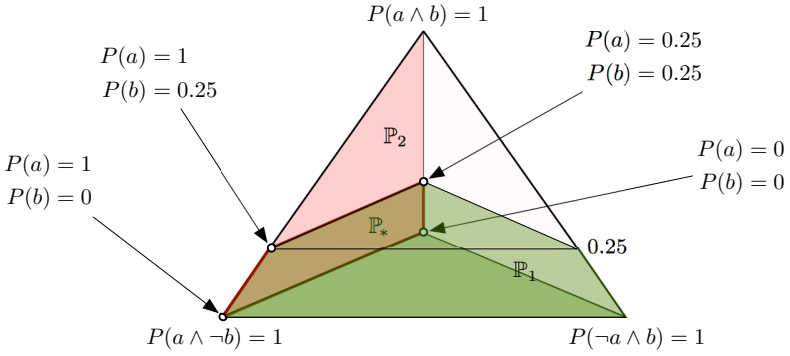


Fig. 1. The set $\mathbb{P}$ of all possible probability measures for the sample space $\{0,1\}^2$, depicted as a tetrahedron, together with the convex sets $\mathbb{P}_1$, $\mathbb{P}_2$, and $\mathbb{P}_*$ of Example 1

epistemic probabilities [28, 29]. Apart from that, they are classical additive probabilities in the sense of Kolmogorov's axioms.

3.1 Degrees of Support and Possibility

Probabilistic argumentation requires the available evidence to be encoded by a finite set $\Phi = \{\varphi_1, \ldots, \varphi_n\} \subset \mathcal{L}_V$ of sentences in a logical language $\mathcal{L}_V$ (over a set of discrete variables V) and a fully specified probability measure $P : 2^{\Theta_W} \to [0,1]$, where Θ_W denotes the discrete sample space generated by a subset $W \subseteq V$ of so-called *probabilistic variables*. These are the theory's basic ingredients. There are no further assumptions regarding the specification of the probability measure P (we may for example use a Bayesian network) or the language $\mathcal{L}_V$.

Definition 1. *A* probabilistic argumentation system *is a quintuple*

$$\mathcal{A} = (V, \mathcal{L}_V, \Phi, W, P), \tag{2}$$

where V, $\mathcal{L}_V$, Φ, W, and P are as defined above [13].

For a given probabilistic argumentation system $\mathcal{A}$, let another logical sentence $\psi \in \mathcal{L}_V$ represent the hypothesis in question. For the formal definition of degrees of support and possibility, consider the subset of Θ_W, whose elements, if assumed to be true, are each sufficient to make ψ a logical consequence of Φ. Formally, this set of so-called *arguments* is denoted and defined by

$$Args_{\mathcal{A}}(\psi) = \{\omega \in \Omega_W : \Phi_\omega \models \psi\} = \Omega_W \setminus [\![\Phi \cup \{\neg\psi\}]\!]^{\downarrow W}, \tag{3}$$

where Φ_ω is obtained from Ω by instantiating all the variables from W according to the partial truth assignment ω [13]. The elements of $Args_{\mathcal{A}}(\neg\psi)$ are sometimes called *counter-arguments* of ψ, see Fig. 2 for an illustration. Note that the elements of $Args_{\mathcal{A}}(\bot)$ are inconsistent with the available evidence Φ, which is why they are sometimes called *conflicts*. The complement of the set of conflicts,

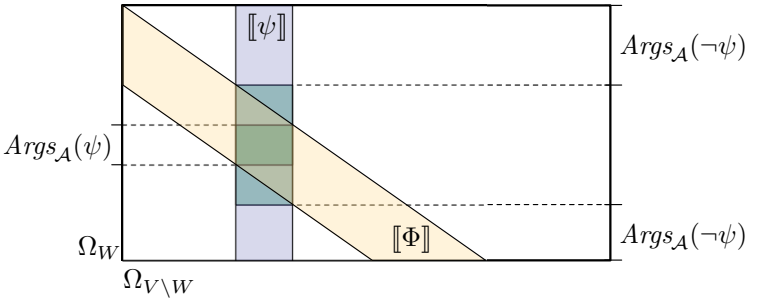


Fig. 2. The sets of arguments and counter-arguments of a hypothesis ψ obtained from the given premises Φ. The sample space Ω_W is a sub-space of the entire space $\Omega_V = \Omega_W \times \Omega_{V\setminus W}$.

$$E_{\mathcal{A}} = \Omega_W \setminus Args_{\mathcal{A}}(\bot) = [\![\Phi]\!]^{\downarrow W}, \tag{4}$$

can thus be interpreted as the available *evidence* in the sample space Ω_W induced by Φ. We will use $E_{\mathcal{A}}$ in its typical role to condition P.

Definition 2. *The* degree of support *of ψ, denoted by $dsp_{\mathcal{A}}(\psi)$, is the conditional probability of the event $Args(\psi)$ given the evidence $E_{\mathcal{A}}$,*

$$dsp_{\mathcal{A}}(\psi) = P(Args_{\mathcal{A}}(\psi)|E_{\mathcal{A}}) = \frac{P(Args_{\mathcal{A}}(\psi)) - P(Args_{\mathcal{A}}(\bot))}{1 - P(Args_{\mathcal{A}}(\bot))}. \tag{5}$$

Definition 3. *The* degree of possibility *of ψ, denoted by $dps_{\mathcal{A}}(\psi)$, is defined by*

$$dps_{\mathcal{A}}(\psi) = 1 - dsp_{\mathcal{A}}(\neg\psi). \tag{6}$$

Note that these formal definitions imply $dsp_{\mathcal{A}}(\psi) \leq dps_{\mathcal{A}}(\psi)$ for all hypotheses $\psi \in \mathcal{L}_V$ and $dsp_{\mathcal{A}}(\psi) = dps_{\mathcal{A}}(\psi)$ for $W = V$. An important property of degree of support is its consistency with pure logical and pure probabilistic inference. By looking at the extreme cases of $W = \emptyset$ and $W = V$, it turns out that degrees of support naturally degenerate into logical entailment $\Phi \models \psi$ and into ordinary posterior probabilities $P(\psi|\Phi)$, respectively. This underlines the theory's pretense of being a unified formal theory of logical and probabilistic reasoning [11].

When it comes to quantitatively evaluate the truth of a hypothesis ψ, it is possible to interpret degrees of support and possibility as respective lower and upper bounds of an interval. The fact that such bounds are obtained without effectively dealing with probability sets or probability intervals distinguishes the theory from most other approaches to probabilistic logic.

3.2 Possible Semantics for the Progic Framework

Now let's turn our attention to the question of interpreting an instance of the progic framework in form of Equation (1) as a probabilistic argumentation

system. For this, we will first generalize in various ways the idea of the standard semantics as exposed in Subsection 2.2 to degrees of support and possibility (Semantics 1 to 4). Then we will explore the perspective obtained by considering each attached probability set as an indicator of the premise's reliability (Semantics 5–7). In all cases we will end up with lower and upper bounds for the target interval Y in Equation (1).

Semantics 1: The Generalized Standard Semantics

As in the standard semantics, let each attached probability set X_i be interpreted as a constraint for the possible probability measures, except that we will now restrict the sample space to be a sub-space Ω_W of Ω_V for some fixed set $W \subseteq V$ of probabilistic variables. We use again $\mathbb{P}$ to denote the set of all possible probability measures. Since each premise φ_i defines an event $[\![\varphi_i]\!]^{\downarrow W}$ in Ω_W, we can interpret the set X_i as a constraint $P([\![\varphi_i]\!]^{\downarrow W}) \in X_i$. As before, we use $\mathbb{P}_i = \{P \in \mathbb{P} : P(\varphi_i^{\downarrow W}) \in X_i\}$ to denote[4] the set of all probability measures satisfying the constraint for the i-th premise, and $\mathbb{P}_* = \mathbb{P}_1 \cap \cdots \cap \mathbb{P}_n$ for the combination of all constraints. This leads then to a whole family $\mathbb{A} = \{(V, \mathcal{L}_V, \Phi, W, P) : P \in \mathbb{P}_*\}$ of probabilistic argumentation systems, each of which with its own degree of support (and degree of possibility) function.

To use this interpretation to produce an answer to our main question regarding the extent of the set Y for a conclusion ψ, there are different ways to go. By considering all possible degrees of support, i.e. by defining $Y_1 = \{dsp_{\mathcal{A}}(\psi) : \mathcal{A} \in \mathbb{A}\}$, the first option focuses on degrees of support. As a second option, we may consider the counterpart of the first one with degrees of possibility in its center, from which we get $Y_2 = \{dps_{\mathcal{A}}(\psi) : \mathcal{A} \in \mathbb{A}\}$. As a third alternative, we may consider the minimal degree of support, $\underline{dsp}(\psi) = \min\{dsp_{\mathcal{A}}(\psi) : \mathcal{A} \in \mathbb{A}\}$, and the maximal degree of possibility, $\overline{dps}(\psi) = \max\{dps_{\mathcal{A}}(\psi) : \mathcal{A} \in \mathbb{A}\}$, and use them as respective lower and upper bounds for the target interval $Y_3 = [\underline{dsp}(\psi), \overline{dps}(\psi)]$. Note that in the special case of $W = V$, all three options coincide with the standard semantics as described in Subsection 2.2.

Semantics 2: The Standard Semantics Applied to Degrees of Support

A similar semantics arises, if we consider each set X_i to be a constraint for the degree of support of φ_i. Again, we need to fix a set $W \subseteq V$ of probabilistic variables to get started. Consider then the set $\mathbb{S} = \{dsp_{\mathcal{A}} : \mathcal{A} = (V, \mathcal{L}_V, \Phi, W, P),\ P \in \mathbb{P}\}$ of all possible degree of support functions, the corresponding constraints $\mathbb{S}_i = \{dsp_{\mathcal{A}} \in \mathbb{S} : dsp_{\mathcal{A}}(\varphi_i) \in X_i\}$ for each premise, and the combined constraint $\mathbb{S}_* = \mathbb{S}_1 \cap \cdots \cap \mathbb{S}_n$. As before, we obtain a whole family $\mathbb{A} = \{\mathcal{A} : dsp_{\mathcal{A}} \in \mathbb{S}_*\}$ of probabilistic argumentation systems.

For the determination of the target set Y, we may now consider the same three options as in the first semantics. The story is exactly the same, except that it starts from a different set $\mathbb{A}$. As before, $W = V$ leads in all three cases back to the standard semantics.

[4] We prefer to use the simplified notation $P(\varphi_i^{\downarrow W})$ as an abbreviation for $P([\![\varphi_i]\!]^{\downarrow W})$.

Semantics 3: The Standard Semantics Applied to Degrees of Possibility

By considering each sets X_i as a constraint for the degree of support of φ_i, we obtain another possible semantics for the progic framework. Due to its perfect symmetry to the previous semantics, we will not not discuss it explicitly. Note that we may "simulate" this option by applying the second semantics to the negated premises $\neg\varphi_1^{Y_1}, \ldots, \neg\varphi_n^{Y_n}$, where $Y_i = \{1 - x : x \in X_i\}$ denotes the corresponding set of "negated" probabilities, and vice versa. This string relationships is a simple consequence of the relationship between degrees of support and possibility.

Semantics 4: The Standard Semantics Applied Symmetrically

To obtain a more symmetrical semantics, in which degrees of support and degrees of possibility are equally important, we consider the restricted case where each set $X_i = [\ell_i, u_i]$ is an interval. We may then interpret the lower bound ℓ_i as a sharp constraint for the degree of support and the upper bound u_i as a sharp constraint for the degree of possibility of φ_i. For this, we need again a fixed set $W \subseteq V$ of probabilistic variables to get started. Note that we can use the relationship $dps_{\mathcal{A}}(\psi) = 1 - dsp_{\mathcal{A}}(\neg\psi)$ to turn the two constraints $dsp_{\mathcal{A}}(\psi_i) = \ell_i$ and $dps_{\mathcal{A}}(\psi_i) = u_i$ into two constraints for respective degrees of support or into two constraints for respective degrees of possibility. To obtain a target interval Y for a conclusion ψ, we may then proceed in the same way as in Semantics 2 and 3, the results however will be quite different for all possible options for Y.

Semantics 5: Unreliable Premises (Incompetent Sources)

A very simple, but quite different semantics exists when each premise has a sharp probability $X_i = \{x_i\}$ attached to it. We can then think of x_i to represent the *evidential uncertainty* of the premise φ_i in the sense that φ_i belongs to Φ with probability x_i. Formally, we could express this idea by $P(\varphi_i \in \Phi) = x_i$ and thus interpret Φ as a fuzzy set whose membership function is determined by the attached probabilities.

To make this setting compatible with a probabilistic argumentation system, let us first redirect each attached probability x_i to an auxiliary propositional variable rel_i. The intuitive idea of this is to consider each premise φ_i as a piece of evidence from a possibly unreliable source S_i. The *reliability* of S_i is thus modeled by the proposition rel_i, and with $P(rel_i) = x_i$ we measure its degree of reliability. The subsequent discussion will be restricted to the case of *independent*[5] sources, which allows us to multiply the marginal probabilities $P(rel_i)$ to obtain a fully specified probability measure P over all auxiliary variables.

[5] This assumption may appear to be overly idealized, but there are many practical situations in which this is approximately correct [12, 14]. Relaxing the independence assumption would certainly allow us to cover a broader class of problems, but it would also make the analysis more complicated.

On the purely logical side, we should expect that any statement from a reliable source is indeed true. This allows us to write $rel_i \rightarrow \varphi_i$ to connect the auxiliary variable rel_i with φ_i. With

$$\Phi^+ = \{rel_1 \rightarrow \varphi_1, \ldots, rel_n \rightarrow \varphi_n\}$$

we denote the set of all such material implications, from which we obtain a probabilistic argumentation system $\mathcal{A}^+ = (V \cup W, \mathcal{L}_{V \cup W}, \Phi^+, W, P)$ with $W = \{rel_1, \ldots, rel_n\}$ and P as defined above. This allows us then to compute the degrees of support and possibility for the conclusion ψ and to use them as lower and upper bounds for the target interval Y.

In the proposed setting, only the positive case of a reliable source is modeled, but nothing is said about the behaviour of an unreliable source. For this, it is possible to distinguish between *incompetent* or *dishonest* (but competent) sources. In the case of an incompetent source, from which no meaningful evidence should be expected, we may model the negative behaviour by auxiliary implications of the form $\neg rel_i \rightarrow \top$. Note that these implications are all irrelevant tautologies, i.e. we get back to the same set Φ^+ from above. In this semantics, the values $P(rel_i) = x_i$ should therefore be interpreted as *degrees of competence* rather than degrees of reliability.

Semantics 6: Unreliable Premises (Dishonest Sources)

As before, we suppose that all attached probabilities are sharp values x_i, but now we consider the possibility of the sources being *malicious*, i.e. competent but not necessarily honest. In this case, the interpretation of $P(rel_i) = x_i$ becomes the one of a *degree of honesty* of source S_i. Dishonest sources are different from incompetent sources in their attitude of deliberately stating the opposite of the truth. From a logical point of view, $\neg rel_i$ allows us thus to infer $\neg\varphi_i$, which we may express by additional material implications $\neg rel_i \rightarrow \neg\varphi_i$. This leads to an extended set of premises,

$$\Phi^\pm = \Phi^+ \cup \{\neg rel_1 \rightarrow \neg\varphi_1, \ldots, \neg rel_n \rightarrow \neg\varphi_n\} \equiv \{rel_1 \leftrightarrow \varphi_1, \ldots, rel_n \leftrightarrow \varphi_n\},$$

and a different probabilistic argumentation system $\mathcal{A}^\pm = (V \cup W, \mathcal{L}_{V \cup W}, \Phi^\pm, W, P)$. Note that the difference between the two interpretations may have a huge impact on the resulting degrees of support and possibility of ψ, and therefore produce quite different target sets Y.

Semantics 7: Unreliable Premises (Incompetent and Dishonest Sources)

For the more general case, where each $X_i = [\ell_i, u_i]$ is an interval, we will now consider a refined model of the above-mentioned idea of splitting up reliability into competence and honesty. Let X_i still refer to the reliability of the source, but consider now two auxiliary variables $comp_i$ (for competence) and hon_i (for honesty). This allows us to distinguish three exclusive and exhaustive cases,

namely $comp_i \wedge hon_i$ (the source is reliable), $comp_i \wedge \neg hon_i$ (the source is malicious), and $\neg comp_i$ (the source is incompetent). As before, we assume that φ_i holds if S_i is reliable, but also that $\neg\varphi_i$ holds if S_i is malicious. Statements from incompetent sources will again be neglected. Logically, the general behaviour of such a source can thus be modeled by two sentences $comp_i \wedge hon_i \rightarrow \varphi$ and $comp_i \wedge \neg hon_i \rightarrow \neg\varphi_i$, which can be merged into $comp_i \rightarrow (hon_i \leftrightarrow \varphi_i)$. This leads to the set of premises

$$\Phi^* = \{comp_1 \rightarrow (hon_1 \leftrightarrow \varphi_1), \ldots, comp_n \rightarrow (hon_n \leftrightarrow \varphi_n)\}.$$

To turn this model into a probabilistic argumentation system, we need to link the auxiliary variables $W = \{comp_1, \ldots, comp_n, hon_1, \ldots, hon_n\}$ to corresponding probabilities. For this, we assume independence between $comp_i$ and hon_i, which is often quite reasonable. If we assume the least restrictive interval $[0, 1]$ to represent a totally incompetent source, and similarly the most restrictive interval $[x_i, x_i]$ to represent a totally competent source, then $u_i - \ell_i$ surely represents the source's degree of incompetence, from which we obtain

$$P(comp_i) = 1 - (u_i - \ell_i) = 1 - u_i + \ell_i$$

for the marginal probability of $comp_i$. Following a similar line of reasoning, we first obtain $P(comp_i \wedge hon_i) = \ell_i$ for the combined event $comp_i \wedge hon_i$ of a reliable source, which then leads to

$$P(hon_i) = \frac{\ell_i}{P(comp_i)} = \frac{\ell_i}{1 - u_i + l_i}$$

for the marginal probability of hon_i. As before, we can use the independence assumption to multiply these values to obtain a fully specified probability measure P over all auxiliary variables. With $\mathcal{A}^* = (V \cup W, \mathcal{L}_{V \cup W}, \Phi^*, W, P)$ we denote the resulting probabilistic argumentation system, from which we obtain degrees of support and possibility for ψ, the bounds for the target interval Y. Note that $\mathcal{A}^+$ and $\mathcal{A}^\pm$ from the previous two semantics are special cases of $\mathcal{A}^*$, namely for $u_i = 1$ (hon_i becomes irrelevant, and rel_i undertakes the role of $comp_i$) and $\ell_i = u_i$ ($comp_i$ becomes irrelevant, and rel_i undertakes the role of hon_i), respectively.

4 Conclusion

Attaching probabilities to logical sentences is one of the most intuitive and popular starting points for the construction of a probabilistic logic. With the proposed progic framework, for which no particular semantics is imposed, the paper presents a unifying umbrella which covers many existing probabilistic logics. This is the first contribution of the paper.

The second contribution is the discussion of several possible semantics obtained by looking at it as different instances of a probabilistic argumentation system. This underlines the richness and diversity of the common framework.

The discussion also contributes to a better understanding of the connection between the theory of probabilistic argumentation and other probabilistic logics.

This paper is an important partial result in the context of a more comprehensive project, in which other possible semantics and a common computational machinery are currently under investigation.

Acknowledgement. This research is supported by the *Swiss National Science Foundation*, Project No. PP002-102652/1, and *The Leverhulme Trust.*

References

1. Adams, E.W.: A Primer of Probability Logic. CSLI Publications, Stanford (1998)
2. Anderson, K.A., Hooker, J.N.: A linear programming framework for logics of uncertainty. Decision Support Systems 16(1), 39–53 (1996)
3. Bergmann, G.: The logic of probability. American Journal of Physics 9, 263–272 (1941)
4. Carnap, R.: Logical Foundations of Probability. University of Chicago Press (1950)
5. Cozman, F.G.: Credal networks. Artificial Intelligence 120(2), 199–233 (2000)
6. de Finetti, B.: The logic of probability. Philosophical Studies 77(1), 181–190 (1995)
7. Fagin, R., Halpern, J.Y., Megiddo, N.: A logic for reasoning about probabilities. Information and Computation 87(1/2), 78–128 (1990)
8. Fox, J.: Probability, logic and the cognitive foundations of rational belief. Journal of Applied Logic 1(3–4), 197–224 (2003)
9. Gaifman, H.: Concerning measures in first order calculi. Israel Journal of Mathematics 2, 1–18 (1964)
10. Gerla, G.: Inferences in probability logic. Artificial Intelligence 70(1–2), 33–52 (1994)
11. Haenni, R.: Towards a unifying theory of logical and probabilistic reasoning. In: Cozman, F.B., Nau, R., Seidenfeld, T. (eds.) ISIPTA 2005, 4th International Symposium on Imprecise Probabilities and Their Applications, Pittsburgh, USA, pp. 193–202 (2005)
12. Haenni, R.: Using probabilistic argumentation for key validation in public-key cryptography. International Journal of Approximate Reasoning 38(3), 355–376 (2005)
13. Haenni, R.: Probabilistic argumentation. Journal of Applied Logic (forthcoming, 2008)
14. Haenni, R., Hartmann, S.: Modeling partially reliable information sources: A general approach based on Dempster-Shafer theory. International Journal of Information Fusion 7(4), 361–379 (2006)
15. Haenni, R., Kohlas, J., Lehmann, N.: Probabilistic argumentation systems. In: Gabbay, D.M., Smets, P. (eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems. Algorithms for Uncertainty and Defeasible Reasoning, vol. 5, pp. 221–288. Kluwer Academic Publishers, Dordrecht (2000)
16. Hailperin, T.: Probability logic. Notre Dame Journal of Formal Logic 25(3), 198–212 (1984)
17. Hailperin, T.: Sentential Probability Logic. Lehigh University Press (1996)
18. Halpern, J.Y.: An analysis of first-order logics of probability. Artificial Intelligence 46, 311–350 (1990)
19. Howson, C.: Probability and logic. Journal of Applied Logic 1(3–4), 151–165 (2003)

20. Jaynes, E.T.: Probability theory as logic. In: Fougère, P.F. (ed.) 9th Annual Workshop on Maximum Entropy and Bayesian Methods, pp. 1–16 (1990)
21. Kohlas, J.: Probabilistic argumentation systems: A new way to combine logic with probability. Journal of Applied Logic 1(3–4), 225–253 (2003)
22. Kyburg, H.E.: Probability and Inductive Logic. Macmillan, New York (1970)
23. Levi, I.: The Enterprise of Knowledge. The MIT Press, Cambridge (1980)
24. Ng, R., Subrahmanian, V.S.: Probabilistic logic programming. Information and Computation 101(2), 150–201 (1992)
25. Nilsson, N.J.: Probabilistic logic. Artificial Intelligence 28(1), 71–87 (1986)
26. Reichenbach, H.: Wahrscheinlichkeitslogik. Erkenntnis 5, 37–43 (1935)
27. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning 62(1–2), 107–136 (2006)
28. Ruspini, E.H.: The logical foundations of evidential reasoning. Tech. Rep. 408, SRI International, AI Center, Menlo Park, USA (1986)
29. Ruspini, E.H., Lowrance, J., Strat, T.: Understanding evidential reasoning. International Journal of Approximate Reasoning 6(3), 401–424 (1992)
30. Scott, D., Krauss, P.: Assigning probabilities to logical formulas. In: Hintikka, J., Suppes, P. (eds.) Aspects of Inductive Logic, pp. 219–264. North-Holland, Amsterdam (1966)
31. Williamson, J.: Probability logic. In: Gabbay, D., et al. (eds.) Handbook of the Logic of Argument and Inference: The Turn Toward the Practical, pp. 397–424. Elsevier, Amsterdam (2002)

A Unified Formulation of Deduction, Induction and Abduction Using Granularity Based on VPRS Models and Measure-Based Semantics for Modal Logics

Yasuo Kudo[1], Tetsuya Murai[2], and Seiki Akama[3]

[1] Dept. of Computer Science and Systems Eng., Muroran Institute of Technology
27-1, Mizumoto, Muroran 050-8585, Japan
kudo@csse.muroran-it.ac.jp
[2] Graduate School of Information Science and Technology, Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo 060-0814, Japan
murahiko@main.ist.hokudai.ac.jp
[3] 1-20-1, Higashi-Yurigaoka, Asao-ku, Kawasaki 215-0012, Japan
akama@jcom.home.ne.jp

Summary. In this paper, we propose a unified formulation of deduction, induction and abduction using granularity based on variable precision rough set models proposed by Ziarko [12] and measure-based semantics for modal logics proposed by Murai et al. [4].

1 Introduction

Rough set theory [8, 9] provides a theoretical basis of approximation and reasoning based on data. Variable precision rough set models (for short, VPRS) proposed by Ziarko [12] is one extension of rough set theory that enable us to treat probabilistic or inconsistent information in the framework of rough sets. From the viewpoint of modal logic, it is well-known that lower and upper approximations in rough sets and necessity and possibility in modal logic are closely related. From a viewpoint of reasoning based on rough set theory, Murai et al. have proposed a framework of granular reasoning [6, 7], which represents reasoning processes by control of granularity of equivalence classes. Moreover, we have discussed relationship between granularity and background knowledge in reasoning processes [3].

In this paper, we propose a unified formulation of deduction, induction and abduction using granularity based on VPRS and measure-based semantics for modal logics proposed by Murai et al. [4, 5]. Note that this paper is a revised version of the authors' previous paper [2].

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 280–290, 2008.
springerlink.com

2 Backgrounds

2.1 Rough Sets

In this subsection, we briefly review foundations of Pawlak's rough set theory and variable precision rough set models. The contents of this subsection is based on [10].

Let U be a non-empty and finite set of objects called the universe of discourse, and R be an equivalence relation on U called an *indiscernibility relation*. For any element $x \in U$, the equivalence class of x with respect to R is defined as follows:

$$[x]_R \stackrel{\text{def}}{=} \{y \in U \mid xRy\}. \tag{1}$$

The equivalence class $[x]_R$ is the set of objects which can not be discerned from x with respect to R. The quotient set $U/R \stackrel{\text{def}}{=} \{[x]_R \mid x \in U\}$ provides a partition of U. According to Pawlak [9], any set $X \subseteq U$ represents a *concept*, and a set of concepts is called *knowledge* about U is the set of concepts is a partition of U. Thus, we regard that R provides knowledge about U as the quotient set U/R.

The ordered pair (U, R) is called an *approximation space*, which provides the basis of approximation in rough set theory. For any set of objects $X \subseteq U$, the *lower approximation* $\underline{R}(X)$ of X and the *upper approximation* $\overline{R}(X)$ of X by R are defined as follows, respectively:

$$\underline{R}(X) \stackrel{\text{def}}{=} \{x \in U \mid [x]_R \subseteq X\} \tag{2}$$

$$\overline{R}(X) \stackrel{\text{def}}{=} \{x \in U \mid [x]_R \cap X \neq \emptyset\} \tag{3}$$

The lower approximation $\underline{R}(X)$ of X is the set of objects which are certainly included in X. On the other hand, the upper approximation $\overline{R}(X)$ of X is the set of objects which may be included in X.

If we have $\underline{R}(X) = X = \overline{R}(X)$, we say that X is R-definable, and otherwise, if we have $\underline{R}(X) \subset X \subset \overline{R}(X)$, we say that X is R-rough. The concept X is R-definable means that we can denote X correctly by using background knowledge by R. On the other hand, X is R-rough means that we can not denote the concept correctly based on the background knowledge.

As one generalization of approximation using rough sets, Yao et al. [11] have discussed generalized lower approximation and generalized upper approximation by replacing the equivalence class used in (2) and (3) by the set

$$U_R(X) \stackrel{\text{def}}{=} \{y \in U \mid xRy\} \tag{4}$$

with respect to any binary relation on U.

Variable precision rough set models (for short, VPRS) proposed by Ziarko [12] is one extension of Pawlak's rough set theory which provides a theoretical basis to treat probabilistic or inconsistent information in the framework of rough sets.

Let $X \subseteq U$ be an any set of objects, R be an indiscernibility relation on U, and a degree $\beta \in [0, 0.5)$ be a precision. The *β-lower approximation* $\underline{R}_\beta(X)$

of X and the *β-upper approximation* $\overline{R}_\beta(X)$ of X by R are defined as follows, respectively:

$$\underline{R}_\beta(X) \stackrel{\text{def}}{=} \left\{ x \in U \middle| \frac{|[x]_R \cap X|}{|[x]_R|} \geq 1 - \beta \right\} \tag{5}$$

$$\overline{R}_\beta(X) \stackrel{\text{def}}{=} \left\{ x \in U \middle| \frac{|[x]_R \cap X|}{|[x]_R|} > \beta \right\} \tag{6}$$

where $|X|$ is the cardinality of the set X. The precision β represents the threshold degree of misclassification of elements in the equivalence class $[x]_R$ to the set X. Thus, in VPRS, misclassification of elements are allowed if the ratio of misclassification is under the precision β. Note that the β-lower and upper approximations with $\beta = 0$ corresponds to Pawlak's lower and upper approximations. Table 1 represents some properties of β-lower and upper approximations. Two symbols "◯" and "×" appeared in Table 1 illustrate that, for each property in Table 1, whether the property is satisfied (denoted by "◯") or not satisfied (denoted by "×") in the case of $\beta = 0$ and $0 < \beta < 0.5$, respectively. For example, by the definition of β-lower approximation by (5), it is easy to check that the property **T**. $\underline{R}_\beta(X) \subseteq X$ are not guaranteed to satisfy in the case of $0 < \beta < 0.5$. Note that symbols assigned to properties like **T**. correspond to axiom schemas in modal logic (for detail, see [1] for example). In the next subsection, we briefly review modal logic, and relationship between rough set theory and modal logic.

Table 1. Some properties of β-lower and upper approximations

Properties		$\beta = 0$	$0 < \beta < 0.5$
Df $\Diamond$.	$\overline{R}_\beta(X) = \underline{R}_\beta(X^c)^c$	◯	◯
M.	$\underline{R}_\beta(X \cap Y) \subseteq \underline{R}_\beta(X) \cap \underline{R}_\beta(Y)$	◯	◯
C.	$\underline{R}_\beta(X) \cap \underline{R}_\beta(Y) \subseteq \underline{R}_\beta(X \cap Y)$	◯	×
N.	$\underline{R}_\beta(U) = U$	◯	◯
K.	$\underline{R}_\beta(X^c \cup Y) \subseteq \left(\underline{R}_\beta(X)^c \cup \underline{R}_\beta(Y)\right)$	◯	×
D.	$\underline{R}_\beta(X) \subseteq \overline{R}_\beta(X)$	◯	◯
P.	$\underline{R}_\beta(\emptyset) = \emptyset$	◯	◯
T.	$\underline{R}_\beta(X) \subseteq X$	◯	×
B.	$X \subseteq \underline{R}_\beta(\overline{R}_\beta(X))$	◯	×
4.	$\underline{R}_\beta(X) \subseteq \underline{R}_\beta(\underline{R}_\beta(X))$	◯	◯
5.	$\overline{R}_\beta(X) \subseteq \underline{R}_\beta(\overline{R}_\beta(X))$	◯	◯
	$X \subseteq Y \Rightarrow \underline{R}_\beta(X) \subseteq \underline{R}_\beta(Y)$	◯	◯
	$R \subseteq R' \Rightarrow \underline{R}_\beta(X) \supseteq \underline{R'}_\beta(X)$	◯	×

2.2 Kripke Models for Modal Logic

Propositional modal logic (hereafter, we simply denote modal logic) extends classical propositional logic by two unary operators $\Box$ and $\Diamond$ (called modal operators), and, for any proposition p, provides the following statements $\Box p$ (p is necessary) and $\Diamond p$ (p is possible).

Let $\mathcal{L}_{\mathrm{ML}}(\mathcal{P})$ be the set of sentences constructed from a given at most countably infinite set of atomic sentences $\mathcal{P} = \{\mathbf{p}_1, \cdots, \mathbf{p}_n(, \cdots)\}$ constant sentences $\top$ (truth) and $\bot$ (falsity), logical connectives $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditionality), $\leftrightarrow$ (biconditionality), $\neg$ (negation), and modal operators $\Box$ (necessity) and $\Diamond$ (possibility) by the following construction rules:

$$\begin{aligned}
&\mathbf{p} \in \mathcal{P} \Rightarrow \mathbf{p} \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P}),\ \top, \bot \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P}),\\
&p \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P}) \Rightarrow \neg p, \Box p, \Diamond p \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P}),\\
&p, q \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P}) \Rightarrow p \wedge q, p \vee q, p \rightarrow q, p \leftrightarrow q \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P}).
\end{aligned}$$

We say that a sentence is a modal sentence if the sentence contains at least one modal operator, and otherwise, we say the sentence is a non-modal sentence.

In this paper, we consider possible world semantics to interpret sentences used in modal logic. A *Kripke model*, one of the most popular frameworks of possible world semantics, is the following triple:

$$\mathcal{M} = (U, R, v), \tag{7}$$

where U ($\neq \emptyset$) is the set of possible worlds, R is a binary relation on U called an accessibility relation, and $v : \mathcal{P} \times U \rightarrow \{0, 1\}$ is a valuation function that assigns a truth value to each atomic sentence $\mathbf{p} \in \mathcal{P}$ at each world $w \in U$. We define that an atomic sentence $\mathbf{p}$ is true at a possible world x by the given Kripke model $\mathcal{M}$ if and only if we have $v(\mathbf{p}, x) = 1$. We say that a Kripke model is finite if its set of possible worlds is a finite set.

We denote $\mathcal{M}, x \models p$ to mean that the sentence p is true at the possible world $x \in U$ by the Kripke model $\mathcal{M}$. Interpretation of non-modal sentences is similar to the case of classical propositional logic. On the other hand, in possible world semantics using Kripke models, we use accessibility relations to interpret modal sentences. $\Box p$ is true at x if and only if p is true at every possible world y accessible from x. On the other hand, $\Diamond p$ is true at x if and only if there is at least one possible world y accessible from x and p is true at y. Formally, interpretation of modal sentences are defined as follows:

$$\begin{aligned}
&\mathcal{M}, x \models \Box p \overset{\mathrm{def}}{\Longleftrightarrow} \forall x \in U(xRy \Rightarrow \mathcal{M}, y \models p),\\
&\mathcal{M}, x \models \Diamond p \overset{\mathrm{def}}{\Longleftrightarrow} \exists x \in U(xRy \text{ and } \mathcal{M}, y \models p).
\end{aligned}$$

For any sentence $p \in \mathcal{L}_{\mathrm{ML}}(\mathcal{P})$, the truth set is the set of possible worlds at which p are true by the Kripke model $\mathcal{M}$, and the truth set is defined as follows:

$$\|p\|^{\mathcal{M}} \overset{\mathrm{def}}{=} \{x \in U \mid \mathcal{M}, x \models p\}. \tag{8}$$

We say that a sentence p is true in a Kripke model $\mathcal{M}$ if and only if p is true at every possible world in $\mathcal{M}$. We denote $\mathcal{M} \models p$ if p is true in $\mathcal{M}$.

In the case that the accessibility relation R in the given Kripke model $\mathcal{M}$ is an equivalence relation, the set of possible worlds $U_R(x)$ defined by (1) is identical to the equivalence class $[x]_R$. Thus, in this case, we can rewrite the definition of interpretation of modal sentences as follows:

$$\mathcal{M}, x \models \Box p \Longleftrightarrow U_R(x) \subseteq \|p\|^{\mathcal{M}},$$
$$\mathcal{M}, x \models \Diamond p \Longleftrightarrow U_R(x) \cap \|p\|^{\mathcal{M}} \neq \emptyset.$$

Therefore, in the case that R is an equivalence relation, we have the following correspondence relationship between Pawlak's lower approximation and necessity, and Pawlak's upper approximation and possibility, respectively:

$$\underline{R}(\|p\|^{\mathcal{M}}) = \|\Box p\|^{\mathcal{M}},$$
$$\overline{R}(\|p\|^{\mathcal{M}}) = \|\Diamond p\|^{\mathcal{M}}.$$

2.3 Scott-Montague Models for Modal Logic

Scott–Montague models are generalization of Kripke models which use the following functions

$$N : U \to 2^{2^U} \tag{9}$$

to interpret modal sentences instead of accessibility relations. Formally, a *Scott–Montague model* is the following triple:

$$\mathcal{M} \stackrel{\text{def}}{=} (U, N, v), \tag{10}$$

where U ($\neq \emptyset$) is the set of possible worlds, and $v : \mathcal{P} \times U \to \{0, 1\}$ is a valuation function. N is a function in (9).

Using the function N, interpretation of modal sentences in the Scott–Montague model are defined as follows:

$$\mathcal{M}, x \models \Box p \stackrel{\text{def}}{\Longleftrightarrow} \|p\|^{\mathcal{M}} \in N(x),$$
$$\mathcal{M}, x \models \Diamond p \stackrel{\text{def}}{\Longleftrightarrow} \left(\|p\|^{\mathcal{M}}\right)^c \notin N(x),$$

where the interpretation of $\Diamond p$ is based on duality of modal operators.

For any Scott–Montague model $\mathcal{M}$, if the function N in $\mathcal{M}$ satisfies the following four conditions:

(m) $X \cap Y \in N(x) \Rightarrow X \in N(x)$ and $Y \in N(x)$,
(c) $X, Y \in N(x) \Rightarrow X \cap Y \in N(x)$,
(n) $U \in N(x)$,
(a) $\bigcap_{X \in N_R(x)} X \in N_R(x)$,

we can construct an accessibility relation R_N as follows:

$$xR_N y \Longleftrightarrow y \in \bigcap_{X \in N(x)} X. \tag{11}$$

Note that the conditions (m), (c) and (n) correspond to axiom schemas **M**, **C** and **N**, respectively.

2.4 Measure-Based Semantics

A function $\mu : 2^U \to [0, 1]$ is called a *fuzzy measure* on U if the function μ satisfies the following conditions: (a) $\mu(U) = 1$, (b) $\mu(\emptyset) = 0$, and (c) $\forall X, Y \subseteq U,\ X \subseteq Y \Rightarrow \mu(X) \leq \mu(Y)$. In measure-based semantics [4, 5], we use fuzzy measures

assigned to each possible worlds to interpret modal sentences. Formally, a fuzzy measure model $\mathcal{M}_\mu$ is the following triple:

$$\mathcal{M}_\mu = (U, \{\mu_x\}_{x \in U}, v),$$

where U is a set of possible worlds, and v is a valuation. $\{\mu_x\}_{x \in U}$ is a class of fuzzy measures μ_x assigned to all possible worlds $x \in U$.

In measure-based semantics of modal logic, each degree $\alpha \in (0, 1]$ of fuzzy measures corresponds to a modal operator $\Box_\alpha$. In this paper, we fix a degree α, and consider *α-level fuzzy measure model*. Interpretation of modal sentences by the α-level fuzzy measure model M_μ is defined by

$$\mathcal{M}_\mu, x \models \Box p \overset{\text{def}}{\Longleftrightarrow} \mu_x\left(\|p\|^{\mathcal{M}_\mu}\right) \geq \alpha, \tag{12}$$

where μ_x is the fuzzy measure assigned to x.

It is easy to check that for any α-level fuzzy measure model $\mathcal{M}_\mu = (U, \{\mu_x\}_{x \in U}, v)$, there is a Scott-Montague model $\mathcal{M}_{\text{SM}} = (U, N_\alpha, v)$ with the following function N_α defined by

$$N_\alpha(x) \overset{\text{def}}{=} \{X \subseteq U \mid \mu_x(X) \geq \alpha\}, \tag{13}$$

and for all sentence $p \in \mathcal{L}_{\text{ML}}(\mathcal{P})$ and all possible world $x \in U$, $\mathcal{M}_{\text{SM}}$ satisfies the following property:

$$\mathcal{M}_\mu, x \models p \Longleftrightarrow \mathcal{M}_{\text{SM}}, x \models p. \tag{14}$$

3 A Unified Formulation of Deduction, Induction and Abduction Using Granularity

In this section, we introduce a unified framework of reasoning using granularity based on VPRS and measure-based semantics for modal logic.

3.1 Background Knowledge by Kripke Models Based on Approximation Spaces

As a basis of reasoning using granularity based on VPRS and measure-based semantics, suppose we have a Kripke model $\mathcal{M} = (U, R, v)$ that consists of the given approximation space (U, R) and a valuation v. In the Kripke model $\mathcal{M}$, any non-modal sentence p that represents a fact is characterized by its truth set $\|p\|^{\mathcal{M}}$. When we consider the fact represented by the non-modal sentence p, we may not consider *all* possible worlds in the truth set $\|p\|^{\mathcal{M}}$. In such cases, we often consider only *typical situations* about the fact p.

To capture the typical situations about the fact p, we consider the lower approximation of the truth set $\|p\|^{\mathcal{M}}$ by the indiscernibility relation R, and regard each possible world $x \in \|p\|^{\mathcal{M}}$ as a typical situation about p based on the "background knowledge" about U.

Moreover, we also need to consider misunderstandings about facts. In this paper, we capture misunderstandings about facts by regarding situations that are not typical about the facts as typical situations, and illustrate this characteristic by using β-lower approximations of the truth sets of sentences which represent the facts. Thus, using the background knowledge by the Kripke model $\mathcal{M}$, we can consider the following two sets of possible worlds about a fact p:

- $\|p\|^{\mathcal{M}}$: correct representation of the fact p.
- $\underline{R}_{\beta}(\|p\|^{\mathcal{M}})$: the set of "typical situations" about p (non-typical situations may be also included).

3.2 α-Level Fuzzy Measure Models Based on Background Knowledge

Using the given Kripke model as background knowledge, we define an α-level fuzzy measure model to treat "typical situations" about facts as β-lower approximations in the framework of modal logic.

Definition 1. *Let $\mathcal{M} = (U, R, v)$ be a Kripke model that consists of an approximation space (U, R) and a valuation v, and $\alpha \in (0.5, 1]$ is a fixed degree. An α-level fuzzy measure model $\mathcal{M}_{\alpha}^{R}$ based on the background knowledge is the following triple:*

$$\mathcal{M}_{\alpha}^{R} \stackrel{\text{def}}{=} (U, \{\mu_{x}^{R}\}_{x \in U}, v), \tag{15}$$

where U and v are the same ones in $\mathcal{M}$. The fuzzy measure μ_{x}^{R} assigned to each $x \in U$ is a probability measure based on the equivalence class $[x]_R$ with respect to R defined by

$$\mu_{x}^{R}(X) \stackrel{\text{def}}{=} \frac{|[x]_R \cap X|}{|[x]_R|}. \tag{16}$$

The constructed α-level fuzzy measure model $\mathcal{M}_{\alpha}^{R}$ from $\mathcal{M}$ has the following good properties.

Proposition 1. *Let $\mathcal{M}$ be a finite Kripke model such that its accessibility relation R is an equivalence relation, and $\mathcal{M}_{\alpha}^{R}$ be the α-level fuzzy measure model based on the background knowledge $\mathcal{M}$ defined by (15). For any non-modal sentence $p \in \mathcal{L}_{ML}(\mathcal{P})$, the following equations are satisfied:*

$$\|p\|^{\mathcal{M}_{\alpha}^{R}} = \|p\|^{\mathcal{M}}, \tag{17}$$

$$\|\Box p\|^{\mathcal{M}_{\alpha}^{R}} = \underline{R}_{1-\alpha}\left(\|p\|^{\mathcal{M}}\right). \tag{18}$$

By (18), we have the following soundness of systems of modal logic with respect to a class of all α-level fuzzy measure models based on background knowledge.

Proposition 2. *For any α-level fuzzy measure model $\mathcal{M}_{\alpha}^{R}$ defined by (15) based on any finite Kripke model $\mathcal{M}$ such that its accessibility relation R is an equivalence relation, the following soundness properties are satisfied in the case of $\alpha = 1$ and $\alpha \in (0.5, 1)$, respectively:*

- *If $\alpha = 1$, then all theorems of the system S5 are true in $\mathcal{M}_{\alpha}^{R}$.*
- *If $\alpha \in (0.5, 1)$, then all theorems of the system ENMD45 are true in $\mathcal{M}_{\alpha}^{R}$.*

3.3 Deduction

Deduction is the following reasoning form:

$p \to q$	If P, then Q.
p	P.
q.	Therefore, Q.

Deduction is logically valid inference, where logically valid means that if both the antecedent p and the rule $p \to q$ are true, then the consequent q is guaranteed to be true. Hereafter, we assume that all sentences p, q, etc. which represent some facts, and rules like $p \to q$ are non-modal sentences.

In the framework of possible world semantics, true rules are represented by inclusion relationship between truth sets as follows:

$$\mathcal{M}_\alpha^R \models p \to q \Longleftrightarrow \|p\|^{\mathcal{M}_\alpha^R} \subseteq \|q\|^{\mathcal{M}_\alpha^R}. \tag{19}$$

As we have shown in Table 1, the monotonicity of β-lower approximation is satisfied for all $\beta \in [0, 0.5)$, thus we have the following relationship:

$$\mathcal{M}_\alpha^R \models \Box p \to \Box q \Longleftrightarrow \|\Box p\|^{\mathcal{M}_\alpha^R} \subseteq \|\Box q\|^{\mathcal{M}_\alpha^R}. \tag{20}$$

If we regard the truth set of $\Box p$ as the set of "typical situations of p", then, by (20), we have that every elements $x \in \|\Box p\|^{\mathcal{M}_\alpha^R}$ are also elements in the truth set of $\Box q$, therefore we can conclude that all typical situations of p are also typical situations of q.

Consequently, using the α-level fuzzy measure model $\mathcal{M}_\alpha^R$, deduction is characterized by the following form of valid reasoning:

$\mathcal{M}_\alpha^R \models \Box p \to \Box q$	If (typically) P, then (typically) Q.
$\mathcal{M}_\alpha^R, x \models \Box p$	(Typically) P.
$\mathcal{M}_\alpha^R, x \models \Box q$	(Typically) Q.

As an example of deduction, Suppose sentences p and q have the following meaning:

- p: The sun rises from east.
- q: The sun sets to west.

Thus, deduction is illustrated as follows:

$\mathcal{M}_\alpha^R \models \Box p \to \Box q$	If the sun rises from east, then the sun sets to west.
$\mathcal{M}_\alpha^R, x \models \Box p$	Today, the sun raised from east.
$\mathcal{M}_\alpha^R, x \models \Box q$	The sun will set to west today.

3.4 Induction

Induction is the following reasoning form:

p	P.
q	Q.
$p \to q$.	Therefore, if P, then Q.

Induction is not logically valid, however, we use induction often to provide rules from facts.

Induction has the following characteristic: From the fact that all observed objects satisfying a property p also satisfy a property q, we conclude that if objects satisfy p, then the objects also satisfy q. Using α-level fuzzy measure model $\mathcal{M}_\alpha^R$ based on background knowledge, we characterize induction as follows: Suppose the lower approximation of the truth set $\|p\|^{\mathcal{M}_\alpha^R}$ of the sentence p illustrates the set of observed objects satisfying p. Then, induction has the following form:

$\mathcal{M}_\alpha^R \models \Box p \to q$	If observed objects satisfy P, then the objects also satisfy Q.
$\mathcal{M}_\alpha^R \models p \to q$	If P, then Q.

This form of reasoning is not valid, however, we can regard this reasoning as a valid reasoning by assuming the following property:

$$\|\Box_\alpha p\|^{\mathcal{M}_\alpha^R} = \|p\|^{\mathcal{M}_\alpha^R} \Longleftrightarrow \mathcal{M}_\alpha^R \models \Box_\alpha p \leftrightarrow p \tag{21}$$

This assumption mean that we regard the set of observed objects satisfying p as the set of all objects satisfying p. By repeating observations, we get more detailed background knowledge and this assumption may become more probable. As shown in Table 1, in variable precision rough set models, even though the partition becomes finer, that is, the current equivalence relation R changes to an other equivalence relation R' such that $R' \subseteq R$, the β-lower approximation may not become large. However, we may have the following situation by the more detailed equivalence relation R':

$$\text{For any } q,\ \mathcal{M}_\alpha^R \models \Box_\alpha p \to q \text{ but } \mathcal{M}_\alpha^{R'} \not\models \Box_\alpha p \to q. \tag{22}$$

This situation illustrates that, by getting more detailed background knowledge, we find some exception in the observed objects such that it does not satisfy q even though satisfying p. Therefore, in the framework of α-level fuzzy measure model based on background knowledge, induction has non-monotonicity.

As an example of induction and non-monotonic reasoning, Suppose sentences p and q have the following meaning:

- p: It is a bird.
- q: It can fly.

Thus, induction and non-monotonic reasoning are illustrated as follows:

$\mathcal{M}_\alpha^R \models \Box_\alpha p \to q$	All observed birds can fly.
$\mathcal{M}_\alpha^R \models \Box_\alpha p \leftrightarrow p$	(Generalization of observed results)
$\mathcal{M}_\alpha^R \models p \to q$	Therefore, all birds can fly.
The equivalence relation R changed to a more detailed equivalence relation R' by repeating observations,	
$\mathcal{M}_\alpha^{R'} \not\models \Box_\alpha p \to q$	Not all birds can fly.

3.5 Abduction

Abduction is the following reasoning form:

$p \to q$	If P, then Q.
q	Q.
p.	Therefore, P.

This corresponds to "affirming the consequent", and is not logically valid. However, we use this form of reasoning often to get some new ideas. Abduction is also called "hypothesis reasoning", and uses to explain the fact q by a hypothesis p based on the rule $p \to q$.

Generally, there may exist many rule which infer the sentence q representing the fact. Thus, we need to select one rule to use abduction from many rules $p_i \to q$ ($p_i \in \{p_1, \cdots, p_n(, \cdots)\}$) which infer q. Similar to the case of deduction, we consider the truth set $\|\Box_\alpha q\|^{\mathcal{M}^R_\alpha}$ of $\Box_\alpha q$ as the set of "typical situations" about q, and for each rule $p_i \to q$, we calculate the following degree of the truth set $\|p_i\|^{\mathcal{M}^R_\alpha}$ of the antecedent p_i of the rule:

$$\alpha(p_i) \stackrel{\text{def}}{=} \arg \min_{x \in \|\Box_\alpha q\|^{\mathcal{M}^R_\alpha}} \mu^R_x \left(\|p_i\|^{\mathcal{M}^R_\alpha} \right). \tag{23}$$

If the calculated degree $\alpha(p_i)$ satisfies $\alpha(p_i) \geq \alpha$, then the antecedent p_i satisfies the following property:

$$\mathcal{M}^R_\alpha \models \Box_\alpha q \to \Box_\alpha p. \tag{24}$$

Therefore, by using the rule $p_i \to q$ with the antecedent p, abduction that infer p_i from the fact q is characterized by the following form of valid reasoning:

$\mathcal{M}^R_\alpha, x \models \Box q$	(Actually) Q.
$\mathcal{M}^R_\alpha \models \Box q \to \Box p$	Selection of a rule "if P, then Q."
$\mathcal{M}^R_\alpha, x \models \Box p$	(Perhaps) P.

Note that the case there is no rule that satisfies (23) corresponds to the situation that we can not explain the fact q by the current background knowledge.

As an example of abduction (or hypothesis reasoning), we consider a reasoning based on fortune-telling we have used in [2]. Suppose sentences p and q have the following meaning:

- p: I wear some red items.
- q: I am lucky.

Then, reasoning based on fortune-telling is characterized by abduction as follows:

$\mathcal{M}^R_\alpha, x \models \Box q$	I am very lucky today!
$\mathcal{M}^R_\alpha \models \Box q \to \Box p$	In a magazine, I saw a fortune-telling "wearing red items makes you lucky".
$\mathcal{M}^R_\alpha, x \models \Box p$	Actually I wear red socks!

4 Conclusion

In this paper, we have introduced α-level fuzzy measure model based on background knowledge, and proposed a unified formulation of deduction, induction and abduction based on the α-level fuzzy measure model based on background knowledge. Using the proposed α-level fuzzy measure model, deduction and abduction are characterized as valid reasoning processes based on typical situations of facts. On the other hand, induction is characterized as a reasoning process of generalization of observations. Moreover, we have pointed out that non-monotonic reasoning has the same structure with induction in the α-level fuzzy measure model based on background knowledge. More refinement of the proposed framework is one of the most interesting future issues.

References

1. Chellas, B.F.: Modal Logic: An Introduction. Cambridge University Press, Cambridge (1980)
2. Kudo, Y., Murai, T.: Approximation of Concepts and Reasoning Based on Rough Sets. Journal of Japanese Society for Artificial Intelligence 22(5), 597–604 (in Japanese, 2007)
3. Murai, T., Kudo, Y., Akama, S.: A Role of Granularity and Background Knowledge in Reasoning Processes — Towards a Foundation of Kansei Representation in Human Reasoning. Kansei Engineering International 6(3), 41–46 (2006)
4. Murai, T., Miyakoshi, M., Shimbo, M.: Measure-Based Semantics for Modal Logic. In: Fuzzy Logic: State of the Art, pp. 395–405. Kluwer, Dordrecht (1993)
5. Murai, T., Miyakoshi, M., Shimbo, M.: A Logical Foundation of Graded Modal Operators Defined by Fuzzy Measures. In: Proc. of 4th FUZZ-IEEE, pp. 151–156 (1995)
6. Murai, T., et al.: Granular Reasoning Using Zooming In & Out: Part 2. Aristotlefs Categorical Syllogism. Electronic Notes in Theoretical Computer Science 82(4) (2003)
7. Murai, T., et al.: Granular Reasoning Using Zooming In & Out: Part 1. Propositional Reasoning. In: Wang, G., et al. (eds.) RSFDGrC 2003. LNCS (LNAI), vol. 2639, pp. 421–424. Springer, Heidelberg (2003)
8. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Science 11, 341–356 (1982)
9. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer, Dordrecht (1991)
10. Polkowski, L.: Rough sets: Mathematical Foundations. Advances in Soft Computing. Physica-Verlag (2002)
11. Yao, Y.Y., Wang, S.K.M., Lin, T.Y.: A Review of Rough Set Models. In: Rough Sets and Data Mining, pp. 47–75. Kluwer, Dordrecht (1997)
12. Ziarko, W.: Variable Precision Rough Set Model. Journal of Computer and System Science 46, 39–59 (1993)

Information from Inconsistent Knowledge: A Probability Logic Approach

Jeff B. Paris, David Picado-Muiño*, and Michael Rosefield**

School of Mathematics
University of Manchester
Manchester M13 9PL, UK
jeff.paris@manchester.ac.uk, davidpicado@yahoo.com,
rosyatrandom@gmail.com

Summary. We treat the sentences in a finite inconsistent knowledge base as assertions that are true with probability at least some primary threshold η and consider as consequences those assertions entailed to have probability at least some secondary threshold ζ.

1 Introduction

A frequent problem in practical applications of reasoning, even with just the basic propositional calculus, is that the knowledge base is actually inconsistent, so that if one was permitted to use the full force of logical consequence it would entail all sentences and so be useless for all practical purposes. On the other hand one's knowledge base, even if formally inconsistent, would normally contain potentially useful information, the problem being how to extract it whilst at the same time avoiding the classical explosion associated with inconsistency.

Numerous methods have been proposed to solve this problem, for example fragmenting the knowledge base into maximally consistent subsets and looking at the common consequences of these, limiting the proof theory or adopting non-classical semantics (see [1], [9] for surveys). Such a variety of approaches seems entirely appropriate here since one can envisage different ways in which one might have acquired an inconsistent knowledge base, for example by receiving some entirely erroneous information, by receiving information from different sources or by receiving information which has become corrupted in transmission.

In this paper we shall consider an approach suiting the situation (amongst others) where the knowledge base consists of the assertions made, or held, by a single rational agent, such as ourselves. In this context our pronounced knowledge frequently does exhibit inconsistency at a formal level, a feature commonly

* Supported by a UK Engineering and Physical Sciences Research Council (EPSRC) Research Studentship.
** Supported by a UK Engineering and Physical Sciences Research Council (EPSRC) Masters Training Programme Award.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 291–307, 2008.
springerlink.com

exploited by lawyers when cross examining in a court of law. Nevertheless, in our everyday lives this causes us little or no concern since under pressure we would maintain not that our assertions were necessarily undeniable facts (though that might happen) but rather that we assigned them a high, or at least reasonable, *degree of belief*, which for this paper we shall identify with subjective probability.

This paper extends results in [6]. To recap the basic approach from that paper we suppose that we have a finite inconsistent knowledge base in some finite propositional language L. Whilst overall the knowledge base is inconsistent in the simplest case we may have no reason to think that any one sentence from it is any more believable than any other. What we can do in this case, and this seems to fit in well with the way we treat inconsistent information in the real world, is to give each sentence some lower bound 'threshold probability' η that it is true. Given that we have accepted this threshold η it might then be argued that we should equally be willing to accept as consequences of our knowledge base any other sentences which as result must have (by probability logic, see for example [2], [7]) probability at least some suitable threshold ζ.

Of course the natural choice here might seem to be η, that is treating the believability requirement for the consequences only as rigorously as that for the initial pieces of knowledge. That was the approach taken in [6] (which we refer to for further background). However we might consider setting higher demands on the consequences, that is taking $\zeta > \eta$, or lower, or even measuring the relative believability of a consequence by the maximum ζ which can be sustained. This approach was mentioned very briefly in [6] and will be considered in detail in this paper.

The method of information extraction, or reasoning, from (possibly) inconsistent knowledge bases that we are proposing here has the advantage that it treats all the items in the knowledge base equally. At the same time by varying the choice of thresholds it allows one to provide graded beliefs. Of course the method does not entirely do away with inconsistency, if the initial threshold η is chosen too large then the resulting knowledge even within probability logic may be inconsistent. On the other hand the range of η for which we do have consistency within probability logic itself gives information about the initial knowledge base, as Knight describes in [3], [4].

2 Notation and Definitions

We shall work in a finite (but varying) propositional language L, denoting its propositional variables by $p_1, \ldots, p_l$ or $p, q, r \ldots$ in specific examples. We will denote its corresponding set of sentences by SL (boolean combinations of our primitive propositions in L) and its corresponding set of atoms by At^L. Recall that by atoms (see for example [7]) we mean sentences of the form

$$\pm p_1 \wedge ... \wedge \pm p_l$$

where $+p_i$ and $-p_i$ stand for p_i and $\neg p_i$ respectively.

Let $w : SL \longrightarrow [0,1]$. We say that w is a probability function on L if these two conditions hold for all θ, $\phi \in SL$:

(P1) If $\models \theta$ then $w(\theta) = 1$,

(P2) If $\models \neg(\theta \wedge \phi)$ then $w(\theta \vee \phi) = w(\theta) + w(\phi)$.

From (P1-2) all the standard properties of probability functions follow, for example that for general θ and ϕ,

$$w(\theta \vee \phi) = w(\theta) + w(\phi) - w(\theta \wedge \phi)$$

and if $\theta \models \phi$ then $w(\theta) \leq w(\phi)$ (for more details see [7]). In particular recall that a probability function w is determined uniquely by its values on the atoms,

$$\langle w(\alpha_1), \ldots, w(\alpha_{2^l})\rangle \in \{\langle x_1, \ldots, x_{2^l}\rangle | \, x_i \geq 0, \sum_i x_i = 1\}$$

via the identities

$$w(\theta) = \sum_{\alpha \models \theta} w(\alpha).$$

For $\eta \in [0,1]$ and Γ (here and throughout) a finite subset of SL we say (following Knight in [3],[4],[5]) that Γ is η-consistent if there is some probability function w on L such that $w(\phi) \geq \eta$ for all $\phi \in \Gamma$, in short $w(\Gamma) \geq \eta$. We say that Γ is maximally η-consistent, denoted $\mathrm{mc}(\Gamma) = \eta$, if η is maximal such that Γ is η-consistent. (As shown in [4] there is such a maximal η).

Following [6] and [12] we define for $\eta, \zeta \in [0,1]$,

$$\Gamma^{\eta} \triangleright_{\zeta} \theta \iff \text{for all probability functions } w \text{ on } L, \text{ if } w(\Gamma) \geq \eta \text{ then } w(\theta) \geq \zeta.$$

In the next two sections we will derive some basic properties of ${}^{\eta}\triangleright_{\zeta}$ and in the section following these we will set to work providing an equivalent formulation set entirely within the framework of the propositional calculus. This task will be completed in the final chapter when we shall consider the functional relationship between the thresholds imposed on the assumptions and those consequently applying to the conclusions.

As a simple example here suppose

$$\Gamma = \{\, p \wedge q, p \wedge \neg q \wedge r, \neg p \wedge q \wedge r \,\}. \tag{1}$$

If we are willing to assign belief (i.e. subjective probability) at least $1/3$ to each of these sentences in Γ being true then r will be true with at least probability $2/3$, that is $\Gamma^{1/3} \triangleright_{2/3} r$. Thus if we had set $2/3$ as our threshold for which we were willing to accept conclusions from this set we would accept r. This seems to us a more satisfactory conclusion (*in this context*) than, what is probably the currently most popular system of Rescher and Manor [11], where only conclusions of all maximal consistent subsets are accepted.

Of course one may argue that the Rescher and Manor approach at least yields a wholly consistent set of conclusions whereas even with a threshold $\zeta > 1/2$ the sentences inferred via ${}^{\eta}\triangleright_{\zeta}$ from Γ may be inconsistent when taken as a whole. In response we would aver that our aim is different from theirs, we are not aiming to produce a consistent set of conclusions but rather to produce a set of conclusions each of which is believable to some threshold degree given the assumed credibility of the original knowledge base. This seems to us closer to what is required in the case of a fallible agent (such as ourselves) assigning beliefs. The agent is not in general looking to discard knowledge in order to become internally consistent but rather to draw conclusions with a guaranteed acceptable probability.

In the next and final sections of this paper we investigate some aspects of the relation ${}^{\eta}\triangleright_{\zeta}$ pertinent to the above intention. For example its behavior at extreme values of η, ζ (Proposition 1), its invariance with respect to the underlying language (Theorem 1) and the way the optimal ζ varies as a function of η (Propositions 2, 4, Theorems 3, 4, Corollary 1).

3 Properties of ${}^{\eta}\triangleright_{\zeta}$

In this section we give a number of simple properties of the relation ${}^{\eta}\triangleright_{\zeta}$. Our first proposition just classifies ${}^{\eta}\triangleright_{\zeta}$ at some key extreme values of η, ζ. In the intervening section we will derive an equivalent of ${}^{\eta}\triangleright_{\zeta}$ formulated entirely within the propositional calculus.

Proposition 1. *For any Γ and θ,*

(i) *For all η, $\Gamma\, {}^{\eta}\triangleright_{0} \theta$.*

(ii) *For $\zeta > 0$, $\Gamma\, {}^{1}\triangleright_{\zeta} \theta \iff \Gamma \models \theta$.*

(iii) *For $\eta > mc(\Gamma)$, $\Gamma\, {}^{\eta}\triangleright_{1} \theta$.*

(iv) *For $\zeta > 0$, $\Gamma\, {}^{0}\triangleright_{\zeta} \theta \iff \models \theta$.*

Proof. Parts (i) and (iii) are immediate from the definition of ${}^{\eta}\triangleright_{\zeta}$. If Γ is inconsistent then (ii) follows trivially. Otherwise, notice that , since standard valuations V on L mapping into $\{0, 1\}$ are probability functions, if $\Gamma {}^{1}\triangleright_{\zeta} \theta$ then $V(\theta) \geq \zeta$ for all valuations V. But since $\zeta > 0$ the only possibility here is for $V(\theta) = 1$, so $\Gamma \models \theta$. Conversely suppose $\Gamma \models \theta$ and $w(\Gamma) = 1$. Then for any atom α, if $w(\alpha) > 0$ then $\alpha \models \phi$ for every $\phi \in \Gamma$, otherwise

$$w(\phi) = \sum_{\alpha \models \phi} w(\alpha) < 1,$$

so $\alpha \models \bigwedge \Gamma$. Hence, since $\bigwedge \Gamma \models \theta$,

$$\zeta \leq 1 = \sum_{\alpha \models \bigwedge \Gamma} w(\alpha) = w(\bigwedge \Gamma) \leq w(\theta),$$

as required.

For (iv), if it is not the case that $\models \theta$ then there is some valuation V such that $V(\theta) = 0$ and so since V is also a probability function $\Gamma^0 \mathrel{\triangleright_\zeta} \theta$ must also fail when $\zeta > 0$. Conversely if $\Gamma^0 \mathrel{\triangleright_\zeta} \theta$ fails then there must be some probability function w such that $w(\theta) < \zeta \leq 1$. Hence $\not\models \theta$ by (P1).

The next proposition gives us some general information on how the relation ${}^{\eta}\triangleright_{\zeta}$ is preserved when we alter η and ζ.

Proposition 2. *Assume that* $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \theta$.

(i) If $\tau \geq \eta$, $\nu \leq \zeta$ *then* $\Gamma^{\tau} \mathrel{\triangleright_\nu} \theta$.

(ii) If $\tau \geq 0$, $\zeta > 0$ *and* $\eta + \tau, \zeta + \tau \leq 1$ *then* $\Gamma^{(\eta+\tau)} \mathrel{\triangleright_{(\zeta+\tau)}} \theta$.

Proof. (i) follows trivially from the definition of ${}^{\eta}\triangleright_{\zeta}$. We omit the proof of (ii) since it is a straightforward adaptation of a similar result for ${}^{\eta}\triangleright_{\eta}$ in [6] (and furthermore Corollary 1 in Section 5 will improve on it).

Our next proposition gives us some right and left weakening properties of our consequence relation ${}^{\eta}\triangleright_{\zeta}$, both of which follow directly from the definition.

Proposition 3. *Assume* $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \theta$. *We then have what follows:*

(i) $\Gamma \cup \{\psi\}^{\eta}\triangleright_{\zeta}\theta$.

(ii) If $\theta \models \psi$ *then* $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \psi$.

Our next result demonstrates a closure property of the pairs (η, ζ) such that $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \theta$.

Proposition 4. *If* $\lim_{n\to\infty} \eta_n = \eta$, $\lim_{n\to\infty} \zeta_n = \zeta$ *with the* η_n *increasing and* $\Gamma^{\eta_n} \mathrel{\triangleright_{\zeta_n}} \theta$ *for all* n *then* $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \theta$.

Proof. Suppose on the contrary that under these assumptions $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \theta$ failed, say $w(\Gamma) \geq \eta$ but $w(\theta) < \zeta$ for some probability function w. Then for some n $w(\theta) < \zeta_n$ so $\Gamma^{\eta_n} \not\mathrel{\triangleright_{\zeta_n}} \theta$ since $\eta_n \leq \eta$, contradiction.

The next result shows that ${}^{\eta}\triangleright_{\zeta}$ is *Language Invariant* in the sense that it does not depend on the particular overlying language L that is chosen. More precisely if L_1, L_2 are finite propositional languages such that $\Gamma \subseteq SL_1 \cap SL_2$, $\theta \in SL_1 \cap SL_2$ then $w_1(\theta) \geq \zeta$ for every probability function w_1 on L_1 such that $w_1(\Gamma) \geq \eta$ if and only if $w_2(\theta) \geq \zeta$ for every probability function w_2 on L_2 such that $w_2(\Gamma) \geq \eta$

Theorem 1. *The relation* ${}^{\eta}\triangleright_{\zeta}$ *is Language Invariant.*

Proof. Assume that $\Gamma \subseteq SL$, $\theta \in SL$ and that $\Gamma^{\eta} \mathrel{\triangleright_\zeta} \theta$ in the context of the language L, in other words for all probability functions w on L if $w(\Gamma) \geq \eta$ then $w(\theta) \geq \zeta$. It is enough to show that if L' is the language obtained from L by adding a single new propositional variable p then for any probability function w' on L' if $w'(\Gamma) \geq \eta$ then $w'(\theta) \geq \zeta$, and conversely.

In the forward direction suppose that w' is a probability function on L' such that $w'(\Gamma) \geq \eta$ but $w'(\theta) < \zeta$. Then let w be the restriction (or marginalization) of w' to SL. The w is clearly again a probability function which agrees with w' on Γ and θ and so $\Gamma^{\eta} \mathrel{\triangleright_{\zeta}} \theta$ fails in the sense of L.

Conversely suppose w is a probability function on L such that $w(\Gamma) \geq \eta$ but $w(\theta) < \zeta$. Notice that the atoms of L' are of the form $\alpha \wedge p^{\epsilon}$ where $\epsilon \in \{0, 1\}$ and α is an atom of L. Define w' on the atoms L' by $w'(\alpha \wedge p) = w(\alpha)$, $w'(\alpha \wedge \neg p) = 0$. Then for ϕ a sentence of L

$$w(\phi) = \sum_{\alpha \models \phi} w(\alpha) = \sum_{\alpha \models \phi} w'(\alpha \wedge p) + w'(\alpha \wedge \neg p) = \sum_{\beta \models \phi} w'(\beta) = w'(\phi)$$

where the β range over the atoms of L', since for ϕ a sentence of L,

$$\alpha \models \phi \iff \alpha \wedge p \models \phi \iff \alpha \wedge \neg p \models \phi.$$

Hence $\Gamma^{\eta} \mathrel{\triangleright_{\zeta}} \theta$ also fails in the context of language L'.

Because of this proposition we can assume that the overlying language is in fact potentially infinite. It is just that at any one time we are restricting ourselves to some finite sublanguage. This is very convenient because probability functions are easier to deal with in this case, though it could certainly be dispensed with without changing any of the results.

Having derived the above basic properties of ${}^{\eta}\triangleright_{\zeta}$ we shall, starting in the next section, give an equivalent formulation of ${}^{\eta}\triangleright_{\zeta}$ set entirely within the context of the classical propositional calculus. In the section following that we shall complete the proof of this equivalence and give a rather more detailed picture of the relationship between pairs $\langle \eta, \zeta \rangle$ for which $\Gamma^{\eta} \mathrel{\triangleright_{\zeta}} \theta$ holds.

4 An Equivalent of ${}^{\eta}\triangleright_{\zeta}$ within Propositional Logic

As in [6] we derive the required equivalent via an extended discussion. Indeed, as anticipated in that paper, the derivation follows a similar pattern to that for ${}^{\eta}\triangleright_{\eta}$ (for rational η, ζ, see also [12]) and in consequence we shall be somewhat brief.

We start by considering the case of η, ζ rational, say $\eta = c/d, \zeta = e/f$ with $c, d, e, f \in \mathbb{N}$. We can assume $\eta, \zeta > 0$ since if either of these is zero we trivially have a suitable equivalent version (as will soon be apparent). So suppose that

$$\theta_1, \ldots, \theta_n \mathrel{{}^{c/d}\triangleright_{e/f}} \phi. \tag{2}$$

Let $\beta_1, \ldots, \beta_m$ enumerate the satisfiable sentences of the form

$$\pm\theta_1 \wedge \ldots \wedge \pm\theta_n$$

where $+\theta_i$ and $-\theta_i$ stand for θ_i and $\neg\theta_i$ respectively. Let $\boldsymbol{\theta_i}$ be that m-vector with jth coordinate 1 if $\beta_j \models \theta_i$ and 0 otherwise (i.e. in case $\beta_j \models \neg\theta_i$) and let $\boldsymbol{\phi}$ be the m-vector with jth coordinate 1 if $\beta_j \models \phi$ and 0 otherwise.

Then condition (2) is equivalent to

$$\text{For all } \mathbf{x} \in \mathbb{D}_m, \text{ if } \boldsymbol{\theta_i} \cdot \mathbf{x} \geq c/d \text{ for } 1 \leq i \leq n \text{ then } \boldsymbol{\phi} \cdot \mathbf{x} \geq e/f \tag{3}$$

where

$$\mathbb{D}_m = \{\langle x_1, \ldots, x_m \rangle \mid x_i \geq 0, \sum_i x_i = 1\}.$$

This follows since for any probability function w,

$$\langle w(\beta_1), \ldots, w(\beta_m) \rangle \in \mathbb{D}_m$$

and

$$w(\theta_i) = \sum_{\beta_j \models \theta_i} w(\beta_j) = \boldsymbol{\theta_i} \cdot \langle w(\beta_1), \ldots, w(\beta_m) \rangle.$$

Let $\mathbf{1}$ be the m-vector with 1's at each coordinate and let

$$\underline{\boldsymbol{\theta_i}} = \boldsymbol{\theta_i} - (c/d)\mathbf{1}, \qquad \underline{\boldsymbol{\phi}} = \boldsymbol{\phi} - (e/f)\mathbf{1}.$$

Then (3) can be restated as

$$\text{For all } \mathbf{x} \in \mathbb{D}_m, \text{ if } \underline{\boldsymbol{\theta_i}} \cdot \mathbf{x} \geq 0 \text{ for } 1 \leq i \leq n \text{ then } \underline{\boldsymbol{\phi}} \cdot \mathbf{x} \geq 0. \tag{4}$$

In turn this is equivalent to the assertion that $\underline{\boldsymbol{\phi}}$ is in the cone in $\mathbb{Q}^m$ (i.e. the m-vectors of rationals) given by

$$\left\{ \sum_{i=1}^{n} a_i \underline{\boldsymbol{\theta_i}} + \sum_{j=1}^{m} b_j \mathbf{u_j} \mid 0 \leq a_i, b_j \in \mathbb{Q} \right\}$$

where $\mathbf{u_j}$ is the m-vector with jth coordinate 1 and all other coordinates 0. In other words, it is equivalent to (4) that there are some $0 \leq a_i \in \mathbb{Q}$ such that

$$\underline{\boldsymbol{\phi}} \geq \sum_{i=1}^{m} a_i \underline{\boldsymbol{\theta_i}}. \tag{5}$$

Written in terms of a common denominator M let $a_i = N_i/M$ where the $M, N_i \in \mathbb{N}$. Then (5) becomes

$$M(df\boldsymbol{\phi} - de\mathbf{1}) \geq \sum_{i=1}^{n} N_i(df\boldsymbol{\theta_i} - cf\mathbf{1}), \tag{6}$$

equivalently

$$[Md(f-e) + cf\sum_{i=1}^{n} N_i]\mathbf{1} \geq Mdf\neg\boldsymbol{\phi} + \sum_{i=1}^{n} df N_i \boldsymbol{\theta_i} \tag{7}$$

where $\neg\boldsymbol{\phi} = \mathbf{1} - \boldsymbol{\phi}$. Conversely if (7) holds for some natural numbers $M > 0$, $N_1, \ldots, N_n \geq 0$ then we can reverse this chain to get back (2).

Now let $\chi_1, \ldots, \chi_N \in \{\theta_1, \ldots, \theta_n\}$ be such that amongst these $\chi_1, \ldots, \chi_N$ the sentence θ_i appears exactly $df N_i$ times for each $i = 1, \ldots, n$ (so $N = df \sum_i N_i$). Then for $\beta_r \models \neg\phi$ it follows from (7) that the rth coordinate of $\boldsymbol{\chi}_\mathbf{j}$ is non-zero for at most $-deM + cf \sum_i N_i = (cN - d^2eM)/d$ many j. Hence

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| > (cN - d^2eM)/d}} \bigwedge_{j \in S} \chi_j \models \phi. \tag{8}$$

Similarly if $\beta_r \models \phi$ then it follows from (7) that the rth coordinate of $\boldsymbol{\chi}_\mathbf{j}$ is non-zero for at most $Md(f-e) + cf \sum_{i=1}^n N_i = (cN + d^2M(f-e))/d$ many j. Hence

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| > (cN + d^2M(f-e))/d}} \bigwedge_{j \in S} \chi_j \models \bot. \tag{9}$$

Now let

$$Z = 1 + (cN + d^2M(f-e))/d$$
$$T = 1 + (cN - d^2eM)/d$$

so $T < Z$ and

$$Td(f-e) = fcN - edZ + df.$$

From (8), (9) we have that

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| = Z}} \bigwedge_{j \in S} \chi_j \models \bot, \tag{10}$$

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| = T}} \bigwedge_{j \in S} \chi_j \models \phi, \tag{11}$$

$$Td(f-e) = fcN - edZ + df \text{ and } T < Z. \tag{12}$$

Conversely suppose that for *some* $Z \in \mathbb{N}$ and $\chi_1, \ldots, \chi_N$ (not necessarily those above) (10), (11) and

$$Td(f-e) \leq fcN - edZ + df \text{ and } T < Z \tag{13}$$

hold. Then for any atom α of L, if $\alpha \models \neg\phi$ then for at most $T-1$ many j can we have that $\alpha \models \chi_j$. Similarly if $\alpha \models \phi$ then there can be at most $Z-1$ such j. Hence, using the earlier vector notation but now with the genuine atoms $\alpha_1, \ldots, \alpha_{2^l}$ replacing the β's

$$\sum_{j=1}^{N} \boldsymbol{\chi}_\mathbf{j} \leq (T-1)\mathbf{1} + (Z-T)\boldsymbol{\phi}. \tag{14}$$

Now suppose $\mathbf{x} \in \mathbb{D}_{2^l}$ and $\boldsymbol{\chi}_\mathbf{j} \cdot \mathbf{x} \geq c/d$ for $j = 1, \ldots, N$. Then dotting each side of (14) with $\mathbf{x}$ we obtain

$$(Z - T)\boldsymbol{\phi} \cdot \mathbf{x} \geq (c/d)N - T + 1.$$

But from (13) we have that

$$\frac{(c/d)N - T + 1}{Z - T} \geq e/f$$

so $\boldsymbol{\phi} \cdot \mathbf{x} \geq e/f$.

To sum up, if (10), (11), (13) hold then

$$\chi_1, \ldots, \chi_N \, {}^{c/d}\!\triangleright_{e/f} \, \phi$$

and by Proposition 3(i) (if necessary) we have

$$\theta_1, \ldots, \theta_n \, {}^{c/d}\!\triangleright_{e/f} \, \phi.$$

Conversely if

$$\theta_1, \ldots, \theta_n \, {}^{c/d}\!\triangleright_{e/f} \, \phi$$

then there are sentences $\chi_1, \ldots, \chi_N \in \Gamma$ (possibly with repeats) such that for some Z, T (10), (11), (13) hold. [Indeed we can even have equality in the first inequality in (13) though for practical purposes it is very convenient to adopt the weaker version.]

Taking $\eta = c/d$, $\zeta = e/f$ we now obtain the following propositional equivalent of ${}^{\eta}\!\triangleright_{\zeta}$ in the case when η, ζ are rational and non-zero[1]. We will complete the proof of this theorem for possibly irrational η, ζ at the end of the next section (which will assume this theorem but only in the proven rational case).

Theorem 2. *Let $\eta, \zeta \in (0, 1]$. Then for $\theta_1, \ldots, \theta_n, \phi \in SL$,*

$\theta_1, \ldots, \theta_n \, {}^{\eta}\!\triangleright_{\zeta} \, \phi \iff$

$\exists \chi_1, \ldots, \chi_N \in \{\theta_1, \ldots, \theta_n\}$ *(possibly with repeats) and T, Z such that*

$$T(1 - \zeta) \leq \eta N - \zeta Z + 1, \quad T < Z \text{ and}$$

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| = Z}} \bigwedge_{j \in S} \chi_j \models \bot,$$

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| = T}} \bigwedge_{j \in S} \chi_j \models \phi.$$

[1] In [6] a somewhat less amenable equivalent is given for the special case $\eta = \zeta$ and the necessary side condition $T < Z$ is not explicitly stated.

Theorem 2 allows one to formulate and work with ${}^{\eta}\!\triangleright_{\zeta}$ entirely within the familiar framework of the propositional calculus, though in practice the N can be at least almost exponentially larger than n, see [8]. Nevertheless in small examples this does seem to be of some practical use, certainly so once one is willing to engage with classical propositional reasoning.

To give an idea of how this works in practice let us take our earlier example (1) where $\Gamma = \{ p \wedge q, p \wedge \neg q \wedge r, \neg p \wedge q \wedge r \}$ and $\phi = r$. To see that indeed $\Gamma^{1/3} \triangleright_{2/3} \phi$ it is enough here to take $\chi_1 = p \wedge \neg q \wedge r$ and $\chi_2 = \neg p \wedge q \wedge r$ (so $N = 2$). Then, for $Z = 2$ and $T = 1$, the conditions (10), (11) and (13) in the above discussion (and previous theorem) hold.

In the next section we turn our attention to considering the best ζ, as a function of η, such that $\Gamma^{\eta} \triangleright_{\zeta} \theta$ for given Γ, θ.

5 The Function $F_{\Gamma,\theta}$

For $\Gamma \subseteq SL$, $\theta \in SL$ and $\eta \in [0,1]$ define

$$F_{\Gamma,\theta}(\eta) = \sup \{ \zeta \in [0,1] \mid \Gamma^{\eta} \triangleright_{\zeta} \theta \}.$$

Notice that by Proposition 1(i) this supremum is certainly well defined. Furthermore by Proposition 4 this supremum is actually attained, that is if $F_{\Gamma,\theta}(\eta) = \gamma$ then $\Gamma^{\eta} \triangleright_{\gamma} \theta$, and moreover unless $\eta > \mathrm{mc}(\Gamma)$ and $\models \neg\theta$ there must be some probability function w such that $w(\Gamma) \geq \eta$ and $w(\theta) = \gamma$, otherwise γ would not be the claimed maximum.

Our next theorem gives what turn out to be the key properties of these functions $F_{\Gamma,\theta}$.

Theorem 3. *The function $F_{\Gamma,\theta}$ is increasing and $F_{\Gamma,\theta}(0), F_{\Gamma,\theta}(1) \in \{0,1\}$.*

On the interval $[0, mc(\Gamma)]$ the function $F_{\Gamma,\theta}$ is convex and continuous and is made up of a finite number of straight line segments $y = q_1 x + q_2$ with $q_1, q_2 \in \mathbb{Q}$ such that $q_1 = q_2 = 0$ or $q_1 = 0$, $q_2 = 1$ or $q_1 \geq 1 - q_2 \geq 1$.

On the interval $(mc(\Gamma), 1]$ $F_{\Gamma,\theta}$ has constant value 1.

Proof. That $F_{\Gamma,\theta}$ is increasing follows by Proposition 2(i). The fact that its value on 0 is either 0 or 1 follows from Proposition 1(iv),(ii), and the fact that its value on 1 is either 0 or 1 follows from Proposition 1(ii),(i).

We now consider $F_{\Gamma,\theta}$ on the interval $[0, \mathrm{mc}(\Gamma)]$. To show that $F_{\Gamma,\theta}$ is convex on here suppose that $0 \leq \eta_1 < \eta_2 \leq \mathrm{mc}(\Gamma)$ and $0 < \mu < 1$. Pick probability functions w_1, w_2 such that $w_2(\Gamma) \geq \eta_2$ and $w_2(\theta) = F_{\Gamma,\theta}(\eta_2)$ and $w_1(\Gamma) \geq \eta_1$ and $w_1(\theta) = F_{\Gamma,\theta}(\eta_1)$. Let w be the probability function $\mu w_1 + (1-\mu) w_2$. Then

$$w(\Gamma) \geq \mu w_1(\Gamma) + (1-\mu) w_2(\Gamma) \geq \mu\eta_1 + (1-\mu)\eta_2$$

$$w(\theta) = \mu F_{\Gamma,\theta}(\eta_1) + (1-\mu) F_{\Gamma,\theta}(\eta_2).$$

Then, as required,

$$F_{\Gamma,\theta}(\mu\eta_1 + (1-\mu)\eta_2) \leq \mu F_{\Gamma,\theta}(\eta_1) + (1-\mu) F_{\Gamma,\theta}(\eta_2).$$

Given now that $F_{\Gamma,\theta}$ is increasing and convex on $[0, \mathrm{mc}(\Gamma)]$ to show it is also continuous it suffices by standard results on convex functions, see for example [13], to show that for $r = \mathrm{mc}(\Gamma)$, $\lim_{x \nearrow r} F_{\Gamma,\theta}(x) = F_{\Gamma,\theta}(r)$. By Proposition 4, $\lim_{x \nearrow r} F_{\Gamma,\theta}(x) \leq F_{\Gamma,\theta}(r)$. If strict inequality held here, say

$$\lim_{x \nearrow r} F_{\Gamma,\theta}(x) < \lambda < F_{\Gamma,\theta}(r),$$

then we can find increasing γ_n converging to r and probability functions w_n such that $w_n(\Gamma) \geq \gamma_n$, $w_n(\theta) < \lambda$. Since the w_n are determined by their values on the fixed finitely many atoms of L these w_n must have a subsequence w_{k_n} with a limit, w say. Then

$$w(\Gamma) \geq \lim_{n \to \infty} w_{k_n}(\Gamma) = r$$

whilst

$$w(\theta) = \lim_{n \to \infty} w_{k_n}(\theta) \leq \lambda < F_{\Gamma,\theta}(r),$$

which is the required contradiction.

To show that on $[0, \mathrm{mc}(\Gamma)]$ $F_{\Gamma,\theta}$ is made up of a finite number of straight line segments $y = q_1 x + q_2$ notice that there is a formula $\Psi(x, y)$ of the language of the structure

$$\mathcal{R} = \langle \mathbb{R}, +, \leq, 0, 1 \rangle$$

such that for $\eta, \zeta \in [0, 1]$,

$$\mathcal{R} \models \Psi(\eta, \zeta) \iff F_{\Gamma,\theta}(\eta) = \zeta.$$

Since $\mathcal{R}$ satisfies quantifier elimination and is an elementary extension of the structure

$$\mathcal{Q} = \langle \mathbb{Q}, +, \leq, 0, 1 \rangle$$

we can suppose that $\Psi(x, y)$ is of the form

$$\bigvee_{i=1}^{h} \bigwedge_{j=1}^{g_i} (m_{ij} y \leq n_{ij} x + k_{ij})$$

for some $m_{ij}, n_{ij}, k_{ij} \in \mathbb{Z}$. The set of pairs $\langle \eta, \zeta \rangle$ such that

$$\mathcal{R} \models \bigwedge_{j=1}^{g_i} (m_{ij} y \leq n_{ij} x + k_{ij})$$

is a convex set so since $F_{\Gamma,\theta}$ is a function it must actually be a line segment, and with coefficients in $\mathbb{Q}$.

To see that these coefficients q_1, q_2 have the required properties assume $q_1 > 0$ (otherwise $q_2 \in \{0, 1\}$ by what has already been proved) and pick an interior rational point $\langle \eta, \zeta \rangle$ on this line segment (so $\eta, \zeta < 1$). By the already proven result for rationals there exist such N, Z, T etc. for this pair. Notice that $T \geq 1$, otherwise $\models \theta$ and $q_1 = 0$, $q_2 = 1$, and $T \leq N$, otherwise $q_1 = q_2 = 0$. So,

$$(1-\zeta)T \leq \eta N - \zeta Z + 1$$

and

$$\zeta \leq \frac{\eta N - T + 1}{Z - T}. \tag{15}$$

Clearly we must have equality in (15) since otherwise we could increase ζ to some ζ' and the rule would give $\Gamma \, {}^{\eta}\!\rhd_{\zeta'} \theta$, contradicting $F_{\Gamma,\theta}(\eta) = \zeta$. But then on this line

$$y = \frac{xN - T + 1}{Z - T}$$

we must have $F_{\Gamma,\theta}(x) = y$, so it must be the case that $q_1 = N/(Z-T)$, $q_2 = (1-T)/(Z-T)$. Notice that $Z \leq N+1$ since otherwise we could replace Z by $N+1$ without changing the required conditions and that would contradict the fact that $F_{\Gamma,\theta}(\eta) = \zeta$. The required inequalities $q_1 \geq 1 - q_2 \geq 1$ follow.

This shows that $F_{\Gamma,\theta}$ has the required properties on $[0, \mathrm{mc}(\Gamma)]$. The last part of the theorem follows directly from Proposition 1(iii).

The next result improves on Proposition 2(ii).

Corollary 1. *Suppose* $\eta \in [0, mc(\Gamma)]$, $F_{\Gamma,\theta}(\eta) = \zeta$ *and* $0 \leq \tau \leq mc(\Gamma) - \eta$. *Then:*

(i) If $\zeta > 0$ *then* $F_{\Gamma,\theta}(\eta + \tau) \geq \zeta + \tau(1-\zeta)(1-\eta)^{-1}$.

(ii) If $\eta > 0$ *then* $F_{\Gamma,\theta}(\eta + \tau) \geq min\{1, \tau\zeta\eta^{-1}\}$.

Proof. Part (i) follows by noticing that for arguments between η and $\mathrm{mc}(\Gamma)$ the graph of $F_{\Gamma,\theta}$ cannot dip below the straight line joining $\langle \eta, \zeta \rangle$ and $\langle 1, 1 \rangle$ (notice that by Proposition 1(i) and Proposition 2(i) $F_{\Gamma,\theta}(1) = 1$).

If $\not\models \theta$ part (ii) similarly follows by noting that by Proposition 1(iv) $F_{\Gamma,\theta}(0) = 0$ so for arguments between η and $\mathrm{mc}(\Gamma)$ the graph of $F_{\Gamma,\theta}$ cannot dip below the straight line passing through $\langle 0, 0 \rangle$ and $\langle \eta, \zeta \rangle$. On the other hand if $\models \theta$ then $F_{\Gamma,\theta}$ just takes constant value 1 so the conclusion is trivially true.

We now show the converse to Theorem 3, namely that any function satisfying the properties of $F_{\Gamma,\theta}$ proved in Theorem 3 is in fact of the form $F_{\Gamma,\theta}$ for some Γ, θ (in some finite language). The next lemma is key to showing this.

Lemma 1. *Given* $\Gamma_1, \Gamma_2, \theta_1, \theta_2$ *there are* Γ, θ *(possibly on a finite extension of the language of* Γ_1, Γ_2 *etc.) such that*

$$F_{\Gamma,\theta}(x) = max\{F_{\Gamma_1,\theta_1}(x), F_{\Gamma_2,\theta_2}(x)\}.$$

Proof. We may assume that $\Gamma_i \subseteq SL_i$, $\theta_i \in SL_i$ for $i = 1, 2$ were $L_1 = \{p_1, ...p_n\}$ and $L_2 = \{q_1, ...q_m\}$ are disjoint languages with atoms $\{\alpha_1, ..., \alpha_{2^n}\}$ and $\{\beta_1, ..., \beta_{2^m}\}$ respectively. Let $L = L_1 \cup L_2$ and set $\Gamma = \Gamma_1 \cup \Gamma_2 \subseteq SL$ and $\theta = \theta_1 \vee \theta_2 \in SL$.

First note that by the language invariance of ${}^{\eta}\!\rhd_{\zeta}$ if w is a probability function on L such that $w(\Gamma) \geq \eta$ then $w(\theta_1) \geq F_{\Gamma_1,\theta_1}(\eta)$ and $w(\theta_2) \geq F_{\Gamma_2,\theta_2}(\eta)$, so certainly

$$w(\theta) \geq \max\{F_{\Gamma_1,\theta_1}(\eta), F_{\Gamma_2,\theta_2}(\eta)\}.$$

Thus it only remains to show that there is some probability function w which takes exactly this value. Without loss of generality assume that

$$F_{\Gamma_1,\theta_1}(\eta) \geq F_{\Gamma_2,\theta_2}(\eta)$$

and let w_i, $i = 1, 2$, be a probability function on L_i such that $w_i(\Gamma_i) \geq \eta$, $w_i(\theta_i) = F_{\Gamma_i,\theta_i}(\eta)$. We define a finite sequence of probability functions w^r on L such that for each r

$$\begin{aligned} w^r(\alpha_i) &= w_1(\alpha_i) \text{ for } i = 1, \ldots, 2^n, \\ w^r(\beta_i) &= w_1(\beta_j) \text{ for } j = 1, \ldots, 2^m, \end{aligned} \tag{16}$$

so in consequence

$$w^r(\theta_i) = F_{\Gamma_i,\theta_i}(\eta)$$

for $i = 1, 2$, and such that for the final w^r in this sequence

$$w^r(\theta) = w^r(\theta_1),$$

equivalently

$$w^r(\alpha_i \wedge \beta_j) = 0 \text{ whenever } \alpha_i \not\models \theta_1,\ \beta_j \models \theta_2. \tag{17}$$

To start with set

$$w^0(\alpha_i \wedge \beta_j) = w_1(\alpha_i) \cdot w_2(\beta_j).$$

Now suppose we have successfully constructed w^r. If (17) holds for this w^r then we are done. Otherwise take the atoms $\alpha_i \wedge \beta_j$ with $w^r(\alpha_i \wedge \beta_j) > 0$, $\beta_j \models \theta_2$, $\alpha_i \not\models \theta_1$. In this case we can find an atom $\alpha_p \wedge \beta_q$ with $w^r(\alpha_p \wedge \beta_q) > 0$, $\alpha_p \models \theta_1$, $\beta_q \not\models \theta_2$. Such an atom of L must exist since if not then

$$w^r(\theta_2) = \sum_t w^r(\alpha_{i_t} \wedge \beta_{j_t}) + \sum_s w^r(\alpha_{i_s} \wedge \beta_{j_s})$$

and

$$w^r(\theta_1) = \sum_t w^r(\alpha_{i_t} \wedge \beta_{j_t})$$

for t, s such that $\beta_{j_t} \models \theta_2$, $\alpha_{i_t} \models \theta_1$, $\beta_{j_s} \models \theta_2$, $\alpha_{i_s} \not\models \theta_1$. But then

$$F_{\Gamma_2,\theta_2}(\eta) = w^r(\theta_2) > w^r(\theta_1) = F_{\Gamma_1,\theta_1}(\eta),$$

contradiction.

Now define w^{r+1} as follows, for i, j, p and q as above:

$$\begin{aligned} w^{r+1}(\alpha_i \wedge \beta_j) &= w^r(\alpha_i \wedge \beta_j) - \min\{w^r(\alpha_i \wedge \beta_j), w^r(\alpha_p \wedge \beta_q)\}, \\ w^{r+1}(\alpha_i \wedge \beta_q) &= w^r(\alpha_i \wedge \beta_q) + \min\{w^r(\alpha_i \wedge \beta_j), w^r(\alpha_p \wedge \beta_q)\}, \\ w^{r+1}(\alpha_p \wedge \beta_j) &= w^r(\alpha_p \wedge \beta_j) + \min\{w^r(\alpha_i \wedge \beta_j), w^r(\alpha_p \wedge \beta_q)\}, \\ w^{r+1}(\alpha_p \wedge \beta_q) &= w^r(\alpha_p \wedge \beta_q) - \min\{w^r(\alpha_i \wedge \beta_j), w^r(\alpha_p \wedge \beta_q)\}, \end{aligned}$$

and w^{r+1} agreeing with w^r on all other atoms of L. Then again we have (16) holding for w^{r+1} in place of w^r and compared with w^r the probability function w^{r+1} gives non-zero probability to strictly fewer atoms $\alpha_i \wedge \beta_j$ with $\beta_j \models \theta_2$, $\alpha_i \not\models \theta_1$. Clearly then this process eventually terminates at the required probability function.

We now prove the converse to Theorem 3, that any such function satisfying the conditions proved of $F_{\Gamma,\theta}$ in that theorem is in fact of the form $F_{\Gamma,\theta}$ for *some* Γ, θ. This is clear for the functions which are identically 0 or 1 so we now drop them from consideration in the next result.

Theorem 4. *Let $r \in [0,1] \cap \mathbb{Q}$ and let F be any function such that*

(i) $F(0) = 0$, $F(1) = 1$, F is increasing.

(ii) On [0,r] F is continuous and convex and made up of a finite set of straight line segments $q_1 x + q_2$ with $q_1, q_2 \in \mathbb{Q}$ and $q_1 \geq 1 - q_2 \geq 1$.

(iii) On $(r, 1]$ $F(x) \equiv 1$.

Then there are Γ, θ such that $F = F_{\Gamma,\theta}$ on $[0,1]$.

Proof. In view of Lemma 1 it is enough to show

(A) If $0 \leq r < 1$, $r \in \mathbb{Q}$, then there are $\Gamma \subseteq SL$ and $\theta \in SL$, for some finite language L, such that

$$F_{\Gamma,\theta}(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq r, \\ 1 & \text{for } r < x \leq 1. \end{cases}$$

(B) If $q_1, q_2 \in \mathbb{Q}$, $q_1 \geq 1 - q_2 \geq 1$ then there are $\Gamma \subseteq SL$ and $\theta \in SL$ for some finite language L such that

$$F_{\Gamma,\theta}(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq -q_2/q_1, \\ q_1 x + q_2 & \text{for } -q_2/q_1 \leq x \leq (1-q_2)/q_1, \\ 1 & \text{for } (1-q_2)/q_1 \leq x \leq 1. \end{cases}$$

To show (A), if $r = 0$ just take θ to be a contradiction and $\Gamma = \{\theta\}$. Otherwise for $r = s/t > 0$, $s, t \in \mathbb{N}$, take L large enough and set[2] Γ

$$\Gamma = \{ \bigvee_{i \in S} \alpha_i \mid S \subseteq \{1, \ldots, t\}, |S| = s \}.$$

If the probability function w gives Γ its maximum consistency then so does any permutation of w (when identifying w with a vector in $\mathbb{D}_{2^l}$) and in turn the average over these permutations. Hence we see that Γ attains its maximum consistence of s/t for the probability function which gives each α_i for $i = 1, 2, \ldots, t$ probability $1/t$.

[2] We have chosen this large Γ to make the construction easier to understand. There are much smaller choices, see for example [8].

To show (B) let $-q_2/q_1 = r/t$, $(1-q_2)/q_1 = s/t$ where $r, s, t \in \mathbb{N}$. By the conditions on q_1, q_2, $0 \leq r/t \leq s/t \leq 1$. Again let L be large and set

$$\theta = \bigvee_{j=t+1}^{2t} \alpha_j,$$

and

$$\Gamma = \{ \bigvee_{i \in S} \alpha_i \vee \bigvee_{j \in T} \alpha_j \mid S \subseteq \{1, \ldots, t\}, |S| = r, T \subseteq \{t+1, \ldots, 2t\}, |T| = s \}.$$

Then if $w_1(\alpha_i) = 1/t$ for $i = 1, 2, \ldots, t$, $w_2(\Gamma) = r/t$, $w_1(\theta) = 0$, so $F_{\Gamma,\theta}(r/t) = 0$. Now suppose that $r/t \leq g < s/t$ $F_{\Gamma,\theta}(g) = h < 1$ and let w witness this. As above we can assume that $w(\alpha_i)$ has constant value, a say, for $i = 1, 2, \ldots, t$ and constant value, b say, for $i = t+1, \ldots, 2t$ and that all the probability is assigned to the α_i for $i \leq 2t$ (otherwise assign what is left evenly to the α_i for $i \leq t$). Then $h = tb$, $g = sb + ra$, $ta + tb = 1$ so $h = r/t + (1 - s/t)g = q_1 g + q_2$. From this and the properties of $F_{\Gamma,\theta}$ shown in Theorem 3 part (B) follows.

At this point we finally return to complete the proof of Theorem 2 in the case when one or both of η, ζ are irrational.

Proof. **of Theorem 2 continued.** We first consider the case where η is irrational and ζ rational. In this case if $\Gamma^{\eta} \triangleright_{\zeta} \phi$ then by Theorem 3 $F_{\Gamma,\phi}(x) = q_1 x + q_2$ for some $q_1, q_2 \in \mathbb{Q}$ for all x in some open non-empty neighborhood $(\eta - \epsilon, \eta + \epsilon)$. Since $q_1\eta + q_2$ is irrational ($q_1 \neq 0$ otherwise $\zeta = 0$) it must be that $q_1\eta + q_2 > \zeta$ so there are $r_1, r_2 \in \mathbb{Q}$ such that $r_1 < \eta$, $r_2 > \zeta$, $q_1 r_1 + q_2 > r_2$. Taking r_1 within ϵ of η then $F_{\Gamma,\phi}(r_1) > r_2$ so there is some $\chi_1, \ldots, \chi_N \in \Gamma$ and Z, T such that $T(1 - r_2) \leq r_1 N - r_2 Z + 1$, $T < Z$ and

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| = Z}} \bigwedge_{j \in S} \chi_j \models \bot, \tag{18}$$

$$\bigvee_{\substack{S \subseteq \{1,\ldots,N\} \\ |S| = T}} \bigwedge_{j \in S} \chi_j \models \phi. \tag{19}$$

But then

$$T(1 - \zeta) \leq \eta N - \zeta Z + 1 \tag{20}$$

as required. Conversely if we have $\chi_1, \ldots, \chi_N \in \Gamma$ and Z, T satisfying (18), (19), (20) then again there must be $r_1 < \eta$, $r_2 > \zeta$ such that

$$T(1 - r_2) \leq r_1 N - r_2 Z + 1.$$

Thus by the two rational case already proved $\Gamma^{r_1} \triangleright_{r_2} \phi$.

The case where $\eta \in \mathbb{Q}, \zeta \notin \mathbb{Q}$ is proved similarly. Finally suppose that η, ζ are both irrational. If $\Gamma^\eta \mathrel{|\!\sim}_\zeta \phi$ and $F_{\Gamma,\phi}(\eta) = q_1\eta + q_2 > \zeta$ then just as in the previous case we can show there are χ_i etc. to give the required propositional equivalent. So suppose $F_{\Gamma,\phi}(\eta) = q_1\eta + q_2 = \zeta$. In this case by Theorem 3 $F_{\Gamma,\phi}(x) = q_1 x + q_2$ for x in some non-empty open neighborhood $(\eta - \epsilon, \eta + \epsilon)$. Pick r_1 in this interval and set $r_2 = q_1 r_1 + q_2$. Then by the two rational case there are some $\chi_1, \ldots, \chi_N \in \Gamma$ and Z, T such that $T < Z$ and $T(1-r_2) \leq r_1 N - r_2 Z + 1$ etc.. Indeed we must have equality here, otherwise we could increase r_2 with r_1 fixed and so (using the two rational case) show that $F_{\Gamma,\phi}(r_1) > r_2$. But since this holds for r_1 arbitrarily close to η it must also hold for η. Hence this provides the required equivalent to $\Gamma^\eta \mathrel{|\!\sim}_\zeta \phi$.

Finally in the other direction in the case $\eta, \zeta \notin \mathbb{Q}$, suppose that we have the required $T, Z, \chi_1, \ldots, \chi_N$ satisfying (19),(18),(20). Then for rational r_1 close to η and $r_2 \leq (r_1 N - T + 1)/(Z - T)$, r_2 close to ζ these same $\chi_1, \ldots, \chi_N, \phi$ and Z, T give $\Gamma^{r_1} \mathrel{|\!\sim}_{r_2} \phi$. Since r_1, r_2 can be made arbitrarily close to η, ζ respectively we can conclude by Proposition 4 that $\Gamma^\eta \mathrel{|\!\sim}_\zeta \phi$, as required.

6 Conclusion

We have introduced and investigated a parameterized propositional consequence relation, ${}^\eta\!\mathrel{|\!\sim}_\zeta$, which we have argued is appropriate for making acceptable inferences in the case of an agent's subjective, and possibly inconsistent, knowledge base. Similar results may also be obtained for conditional probabilities, [14], the predicate calculus, [10], and multiple thresholds, which in the guise of a proof theory for probability logic will be the subject of a future paper.

References

1. Batens, D., Mortensen, C., Priest, G.: Frontiers of paraconsistent logic. In: Van Bendegem, J.P. (ed.) Studies in Logic and Computation, vol. 8, Research Studies Press, Baldock (2000)
2. Fagin, R., Halpern, J.Y., Meggido, N.: A Logic for Reasoning about Probabilities. Information and Computation 87, 78–128 (1990)
3. Knight, K.M.: A theory of inconsistency. PhD Thesis, University of Manchester, UK (2002), http://www.maths.man.ac.uk/DeptWeb/Homepages/jbp/
4. Knight, K.M.: Measuring inconsistency. Journal of Philosophical Logic 31(1), 77–98 (2002)
5. Knight, K.M.: Probabilistic entailment and a non-probabilistic logic. Logic Journal of the IGPL 11(3), 353–365 (2003)
6. Paris, J.B.: Deriving information from inconsistent knowledge bases: A completeness theorem for ${}^\eta\!\mathrel{|\!\sim}_\eta$. Logic Journal of the IGPL 12(5), 345–353 (2004)
7. Paris, J.B.: The Uncertain Reasoner's Companion. Cambridge University Press, Cambridge (1994)
8. Picado M.D.: PhD Thesis, University of Manchester, UK (to appear)
9. Priest, G., Routley, R.: Paraconsistent logic: Essays on the inconsistent. In: Norman, J. (ed.) Philosophia Verlag, Müchen (1989)

10. Rafiee R.S.: PhD Thesis, University of Manchester, UK (to appear)
11. Rescher, N., Manor, R.: On inference from inconsistent premises. In: Theory and decision, vol. 1, pp. 179–217 (1970)
12. Rosefield, M.: A Probabilistic Approach to Paraconsistent Reasoning. MA Thesis, University of Manchester, UK (2005)
13. Royden, H.L.: Real Analysis. The Macmillan Company, London (1971)
14. Simmonds, R.: PhD Thesis, University of Manchester, UK (to appear)

Part VI

Fuzziness and Uncertainty Analysis in Applications

Personalized Recommendation for Traditional Crafts Using Fuzzy Correspondence Analysis with Kansei Data and OWA Operator

Mina Ryoke[1], Yoshiteru Nakamori[2], and Van Nam Huynh[2]

[1] University of Tsukuba
Bunkyo, Tokyo 112-0012, Japan
ryoke@mbaib.gsbs.tsukuba.ac.jp
[2] Japan Advanced Institute of Science and Technology
Nomi, Ishikawa 923-1292, Japan
{nakamori,huynh}@jaist.ac.jp

Summary. In this paper, we first apply a technique of fuzzy correspondence analysis to deal with fuzziness in subjective evaluation data of traditional crafts in which evaluated objects are assessed by multiple evaluators against multiple kansei features, where each kansei feature corresponds to two (adjectival) kansei words having opposite meanings. Basically, the variation in subjective evaluation data is transformed into a parameter space trying to preserve information as much as possible, and thereby we define fuzzy sets which express relative fuzziness between principal component scores in the parameter space. Then we define a so-called *fitness measure* which expresses how well an object fits with a given kansei word, taking both the distance and the relative fuzziness into account. This fitness measure is then utilized together with ordered weighted averaging (OWA, for short) operator for aggregation to define a ranking function that quantifies how well an object fits with the feeling preference specified as a set of kansei words by a potential consumer. A prototyping system for visualizing the fuzzy correspondence analysis of subjective evaluation data as well as for personalized recommendation is also developed and illustrated.

1 Introduction

Nowadays, consumers and customers have put higher demands not only on the quality but also on their satisfaction in terms of psychological feelings of products and services to be purchased. Therefore, in an increasingly competitive world market, it is important for manufacturers to have a customer-focused approach in order to improve attractiveness in development of new products, which should not only satisfy requirements of physical quality, defined objectively, but should also satisfy consumers' psychological needs, by essence subjective [15]. Kansei engineering [11, 12] defined as a "translating technology of a consumer's feeling and image for a product into design elements" has been a great interest and concern since its inception in the 1970s. Kansei is a Japanese term which

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 311–325, 2008.
springerlink.com

is, according to Nagamachi [16], 'the impression somebody gets from a certain artefact, environment or situation using all her senses of sight, hearing, feeling, smell, taste [and sense of balance] as well as their recognition'. Typically, Kansei engineering focuses on the issue of how to transform human beings' feelings in terms of subjective evaluation into the design of products and the customer-oriented manufacturing process. It has been rapidly developed and successfully applied to a variety of industries, e.g., [1, 2, 3, 5, 7, 9, 12].

While evaluation and ranking for decision support are often involved in many studies of Kansei engineering or other consumer-oriented design techniques [15], the evaluation process using kansei data of subjective feelings for commercial products have generally received less attention [10], especially those for traditional crafts have not been addressed yet. Such evaluations would be powerful strength in marketing strategy as, for traditional crafts, decisions on which product items to buy or use might be strongly influenced by personal feelings/characteristics. Also, they are essentially helpful for recommendation purposes particularly in the era of e-commerce where recommendation systems have been getting popular and important.

In Japan there are a large number of traditional craft products which are so closely connected to Japanese traditional culture. As explained in the Web site of The Association for the Promotion of Traditional Craft Industries [17], each of which is 'unique fostered through regional differences and loving dedication and provides a continual wealth of pleasure'. However, due to the rapid changing of lifestyles of younger generations plus the prevalence of modern industrial products with their advantage in cost and usage, the market of traditional crafts in Japan has been shrinking over the last recent decades. Under such circumstances, in 1974 Japanese government (METI) enacted the so-called Densan Law [18] for the "Promotion of Traditional Craft Industries" in order that traditional crafts bring richness and elegance to people's living and contribute to the development of local economy.

On the other hand, with the fast growing of e-commerce in today's business, the Internet can be a great help in revitalizing traditional craft industries. Manufacturers and retailers via their Web sites can make their marketing better as providing more attractive introduction and, hopefully, personalized recommendation, or even helping bring people back to traditional and cultural values concerning their products. Our main concern is with the evaluation of traditional crafts for the personalized recommendation problem. A particular focus is put on the evaluation of traditional craft products of Ishikawa prefecture, where Japan Advanced Institute of Science and Technology is located.

The objective of this paper is two-fold. Firstly, due to a large variance in subjective evaluations of traditional crafts against kansei features using human sensibility, we apply a technique of fuzzy correspondence analysis developed in [13] to model individual differences in evaluation and in the spread of vagueness, which hopefully helps us to understand about relative vagueness as well as about relationship between objects and kansei words behind these evaluations. Secondly, based on that understanding, we propose the so-called *fitness*

measure which determines how well an object fits with a given kansei word, and this fitness measure is then combined with OWA operator [19] for aggregation to define a ranking function for quantifying how well an object fits with the feeling preference of a person specified by a set of kansei words. In addition, we have also implemented a prototyping software system for visualizing the fuzzy correspondence analysis of subjective evaluation data as well as for personalized recommendation on traditional products.

The rest of this paper is organized as follows. Section 2 briefly describes the research problem and the formulation of kansei data structure. Section 3 introduces a technique of fuzzy correspondence analysis using kansei data. Section 4 first proposes a fitness measure between evaluation objects and kansei words and then defines a ranking function for the problem of personalized recommendation. Section 5 illustrates the developed prototyping system with a case study of evaluating Yamanaka lacquer product of Ishikawa prefecture. Finally, Section 6 presents some conclusions.

2 Preliminaries

Let us denote $O = \{o_1, \ldots, o_M\}$ the set of objects such as traditional ceramic products to be evaluated. The first task in the evaluation process is to identify what kansei features human beings often use to express our feelings regarding evaluated objects. Each kansei feature is defined by an opposite pair of (adjectival) kansei words, for example the *fun* feature determines the pair of kansei words *solemn* and *funny*. Let us denote $F = \{f_1, \ldots, f_N\}$ the set of kansei features selected, and $\mathbf{w}_n^+$ and $\mathbf{w}_n^-$ be the opposite pair of kansei words corresponding to F_n, for $n = 1, \ldots, N$. Denote $\mathbf{W}$ the set of kansei words, i.e. $\mathbf{W} = \{\mathbf{w}_n^+, \mathbf{w}_n^- | n = 1, \ldots, N\}$.

We then design a questionnaire for gathering kansei evaluation data using the semantic differential (SD, for short) method [14] as a measurement instrument. Note that SD method has been widely used in application, e.g., [4, 5, 8, 10, 13, 15]. The questionnaire using SD method for gathering information consists of the list of kansei features, each of which corresponds to an opposite pair of kansei words located at either end of a 7-point scale $\{1, \ldots, 7\}$. Let us now denote $E = \{e_1, \ldots, e_K\}$ the set of evaluators who are invited to evaluate objects from O against kansei features from F using the designed questionnaire. Several situations may happen with the data set obtained by this way as described below.

Assume that E_m is the set of evaluators who have evaluated object o_m, and O_k is the set of objects which are evaluated by evaluator e_k. The following conditions should be satisfied:

$$E = \bigcup_{m=1}^{M} E_m, \text{ where } E_m \neq \emptyset, \forall m \tag{1}$$

$$O = \bigcup_{k=1}^{K} O_k, \text{ where } O_k \neq \emptyset, \forall k \tag{2}$$

Then the obtained data set highly depends on each particular application as well as the designed questionnaire survey. Particularly, we have at least three different cases of the obtained data structure as follows.

♢ Case 1: (complete 3-way data)

$$E_m = E, \forall m; \text{ and } O_k = O, \forall k$$

♢ Case 2: (one person evaluates only one object)

$$|O_k| = 1, \forall k; E_m \cap E_{m'} = \emptyset, \text{ for } m \neq m'; \text{ and } \sum_{m=1}^{M} |E_m| = K$$

♢ Case 3: (some evaluate several objects)

$$|O_k| \geq 1, \forall k; \text{ and } |E| = K$$

Let us denote z_{mnk} the subjectively evaluation value of evaluator e_k regarding object o_m against kansei feature f_n, where $z_{mnk} \in \{1, 2, 3, 4, 5, 6, 7\}$. In addition, we assume that there is no missing data about the evaluated objects. Denote

$$z_{mk} = (z_{m1k}, z_{m2k}, \cdots, z_{mNk})^t, \text{ where } k \in E_m.$$

For Case 1, the averaged data is determined by

$$z_{mn} = \frac{1}{K} \sum_{k=1}^{K} z_{mnk}, \text{ for } m = 1, 2, \ldots, M, \ n = 1, 2, \ldots, N.$$

For Cases 2 and 3, the averaged data is defined as follows.

$$z_{mn} = \frac{1}{|E_m|} \sum_{k=1}^{K} z_{mnk}, \text{ for } m = 1, 2, \ldots, M, \ n = 1, 2, \ldots, N.$$

The data treated in this paper corresponds to Case 1 above, and the technique of fuzzy correspondence analysis developed in [13] will be adapted for this case in the following section.

3 Fuzzy Correspondence Analysis Using Kansei Data

3.1 Analysis Based on the Average Data

Let $P = [p_{mn}]_{M \times N}$ be the average data matrix for evaluators, where its elements are determined by

$$p_{mn} = \frac{z_{mn}}{\sum_{m=1}^{M} \sum_{n=1}^{N} z_{mn}} \tag{3}$$

Denote

$$p_{m\bullet} = \sum_{n=1}^{N} p_{mn}, \quad p_{\bullet n} = \sum_{m=1}^{M} p_{mn}$$

then we have

$$\sum_{m=1}^{M} p_{m\bullet} = \sum_{n=1}^{N} p_{\bullet n} = 1 \tag{4}$$

In correspondence analysis, a quantity x_m is associated with evaluated object o_m, while a quantity y_n is associated with a kansei word $\mathbf{w}_n$. It should be worth noting here that for each kansei feature f_n, we have two opposite kansei words $\mathbf{w}_n^+$ and $\mathbf{w}_n^-$ $(n = 1, \ldots, N)$, and in our model we use originally obtained data $[z_{mnk}]$ to do analysis for kansei words $\mathbf{w}_n^+$ $(n = 1, \ldots, N)$, which locate on the right hand side in the questionnaire; while the analysis for kansei words $\mathbf{w}_n^-$ $(n = 1, \ldots, N)$ locating on the left hand side in the questionnaire will be done similarly using the reverse data, i.e. $[8-z_{mnk}]$. Abusing the notation, but without danger of confusion, from now on we use only notation $\mathbf{w}_n$ to indicate a kansei word being considered.

Now let us denote

$$x = (x_1, x_2, \cdots, x_M)^t$$
$$y = (y_1, y_2, \cdots, y_N)^t$$

These vectors are determined by maximizing the correlation coefficient ρ_{xy} in terms of the following equation:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \to \max \tag{5}$$

where σ_{xy}, σ_x, σ_y are defined as follows:

$$\sigma_{xy} = \sum_{m=1}^{M} \sum_{n=1}^{N} p_{mn} x_m y_n - \sum_{m=1}^{M} p_{m\bullet} x_m \sum_{n=1}^{N} p_{\bullet n} y_n \tag{6}$$

$$\sigma_x = \sqrt{\sum_{m=1}^{M} p_{m\bullet} x_m^2 - \left(\sum_{m=1}^{M} p_{m\bullet} x_m \right)^2} \tag{7}$$

$$\sigma_y = \sqrt{\sum_{n=1}^{N} p_{\bullet n} y_n^2 - \left(\sum_{n=1}^{N} p_{\bullet n} y_n \right)^2} \tag{8}$$

Then the solution of the optimization problem (5) can be derived as an eigenvalue problem. The scores regarding to object o_m and kansei word $\mathbf{w}_n$ on a two-dimensional plane using the second and third eigenvectors are given as follows:

$$\begin{aligned} &(x_{2m}, x_{3m}), \ m = 1, 2, \ldots, M \\ &(y_{2n}, y_{3n}), \ n = 1, 2, \ldots, N. \end{aligned} \tag{9}$$

3.2 Modeling Fluctuation of Subjective Evaluations

As mentioned above, the location defined by equation (9) on the eigenvector space is determined by maximizing the correlation coefficient. In order to express the fluctuation on subjective evaluations, each evaluator's eigenvector is

expressed using the distance from the location defined by average data and individual evaluation of the evaluator. This can be done as follows.

Let us define

$$b_{mk} = \frac{1}{N}\sum_{n=1}^{N}(z_{mnk} - z_{mn})$$
$$b_m = \frac{1}{K}\sum_{k=1}^{K} b_{mk}, m = 1, 2, \ldots, M$$

Using these vectors, the individual eigenvector is defined by the following equation

$$\tilde{\boldsymbol{x}}_{ik} = \tilde{\boldsymbol{x}}_i + \lambda_i(\boldsymbol{b}_k - \boldsymbol{b}), \text{ for } i = 2, 3 \tag{10}$$

where λ_i are designed parameters, and

$$\boldsymbol{b}_k = (b_{1k}, b_{2k}, \cdots, b_{Mk})^t$$
$$\boldsymbol{b} = (b_1, b_2, \cdots, b_M)^t$$

The individual eigenvectors are then featured by the following equation

$$\frac{1}{K}\sum_{k=1}^{K}\tilde{\boldsymbol{x}}_{ik} = \tilde{\boldsymbol{x}}_i$$

where

$$\tilde{\boldsymbol{x}}_{\boldsymbol{ik}} = (\tilde{x}_{i1k}, \tilde{x}_{i2k}, \cdots, \tilde{x}_{iMk})^t$$
$$\tilde{\boldsymbol{x}}_{\boldsymbol{i}} = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{iM})^t$$

Note that when the individual evaluation vectors are averaged by evaluators, it results in the same with the original eigenvector.

Fuzzy sets of evaluated objects and kansei words

We now define for each object a fuzzy set which expresses relative fuzziness of the object in the parameter space. Using Zadeh's extension principle [21], second and third elements of the eigenvector are mapped into a two-dimensional fuzzy set. In particular, let us denote

$$\tilde{X}_i = (\tilde{X}_{i1}, \cdots, \tilde{X}_{iM})^t \tag{11}$$

a fuzzy vector whose membership function is defined by

$$\mu_{\tilde{X}_i}(\boldsymbol{x}) = \exp\{-(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_i)^t D_X^{-1}(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_i)\} \tag{12}$$

where D_X is independent of i and determined as follows

$$D_X = \frac{1}{K}\sum_{k=1}^{K}(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_i)(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_i)^t = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{b}_k\boldsymbol{b}_k^t \tag{13}$$

In order to make a mapping to the two-dimensional plane consisting of second and third elements of the eigenvector, we define $\boldsymbol{a}_m$ by

$$\boldsymbol{a}_m = (a_{m1}, a_{m2}, \cdots, a_{mM})^t \tag{14}$$

where

$$a_{mm'} = \begin{cases} 1, & m = m' \\ 0, & m \neq m' \end{cases}$$

Then we consider the following mapping

$$X_{im} = \boldsymbol{a}_m^t \tilde{X}_i \tag{15}$$

According to Zadeh's extension principle, the membership function of X_{im} is identified as follows

$$\mu_{X_{im}}(x) = \max_{\{\tilde{\boldsymbol{x}} | \boldsymbol{x} = \boldsymbol{a}_m^t \tilde{\boldsymbol{x}}\}} \mu_{\tilde{x}_i}(\boldsymbol{x}) = \exp\Big\{ - \frac{(x - \boldsymbol{a}_m^t \tilde{\boldsymbol{x}})^2}{(\boldsymbol{a}_m^t D_X \boldsymbol{a}_m)} \Big\} \tag{16}$$

Now we can define a fuzzy set $X_{2m} \times X_{3m}$ on the two-dimensional plane for object o_m having the membership function defined by

$$\begin{aligned} \mu_{X_{2m} \times X_{3m}}(x_2, x_3) &= \mu_{X_{2m}}(x_2) \times \mu_{X_{3m}}(x_3) \\ &= \exp\Big\{ - \frac{(x_2 - \boldsymbol{a}_m^t \tilde{\boldsymbol{x}}_2)^2 + (x_3 - \boldsymbol{a}_m^t \tilde{\boldsymbol{x}}_3)^2}{(\boldsymbol{a}_m^t D_X \boldsymbol{a}_m)} \Big\} \end{aligned}$$

Then, for each $\alpha \in (0, 1]$, the α-level set of the fuzzy set of o_m is determined by the following circle

$$(x_2 - \boldsymbol{a}_m^t \tilde{\boldsymbol{x}}_2)^2 + (x_3 - \boldsymbol{a}_m^t \tilde{\boldsymbol{x}}_3)^2 = (\boldsymbol{a}_m^t D_X \boldsymbol{a}_m)(-log(\alpha)) \triangleq s_m^2 \tag{17}$$

Similarly, we can also define for each kansei word $\mathbf{w}_n$ a fuzzy set $Y_{2n} \times Y_{3n}$ on the two-dimensional plane, whose membership function is defined by

$$\begin{aligned} \mu_{Y_{2n} \times Y_{3n}}(y_2, y_3) &= \mu_{Y_{2n}}(y_2) \times \mu_{Y_{3n}}(y_3) \\ &= \exp\Big\{ - \frac{(y_2 - \boldsymbol{a}_n^t \tilde{\boldsymbol{y}}_2)^2 + (y_3 - \boldsymbol{a}_n^t \tilde{\boldsymbol{y}}_3)^2}{(\boldsymbol{a}_n^t D_Y \boldsymbol{a}_n)} \Big\} \end{aligned}$$

and the α-level set of the fuzzy set of $\mathbf{w}_n$ is described as follows

$$(y_2 - \boldsymbol{a}_m^t \tilde{\boldsymbol{y}}_2)^2 + (y_3 - \boldsymbol{a}_m^t \tilde{\boldsymbol{y}}_3)^2 = (\boldsymbol{a}_m^t D_Y \boldsymbol{a}_m)(-log(\alpha)) \triangleq t_n^2 \tag{18}$$

Putting these together in the two-dimensional plane, we can visualize relative fuzziness as well as the relationship between the evaluated objects and the kansei words in evaluations.

4 A Ranking Procedure for Personalized Recommendation

In this section, we propose a ranking procedure for the following problem of personalized recommendation. Assume that an agent A wants to look at objects according to her preference given by a set $\mathbf{W}_A$ of kansei words, which is a subset of $\mathbf{W}$. Let

$$\mathbf{W}_A = \{\mathbf{w}_{j_1}, \ldots, \mathbf{w}_{j_p}\}$$

Practically, agent A may state her preference in forms of the following statement

"I would like objects which meet LQ (of) attributes from $\mathbf{W}_A$."

where LQ is a linguistic quantifier such as *all*, *most*, *at least half*, etc.

Based on the fuzzy correspondence analysis discussed above, we first define a fitness measure which quantifies how well an object is compatible with a given kansei word, given subjective evaluations from a population of evaluators. Then we propose to use an OWA operator [19] for aggregating fitness values of an object with respect to kansei words from $\mathbf{W}_A$ to define an overall fitness value, where weights of OWA operator are determined using the fuzzy set based semantics of linguistic quantifier involved.

4.1 A Fitness Measure

Consider an object o_m and a kansei word $\mathbf{w}_n$, which are fuzzily expressed in the two-dimensional space by fuzzy sets $X_{2m} \times X_{3m}$ and $Y_{2n} \times Y_{3n}$, respectively. Note that points (x_{2m}, x_{3m}) and (y_{2n}, y_{3n}) (refer to (9)) are modal points of these two-dimensional fuzzy sets respectively. In order to define a fitness measure between object o_m and kansei word $\mathbf{w}_n$, we fix an $\alpha \in (0, 1]$, and then the fitness measure is determined by the distance between these modal points and the radiuses of corresponding α-level sets. Particularly, the fitness measure between o_m and $\mathbf{w}_n$ is defined by

$$fitness(o_m, \mathbf{w}_n) \triangleq S_{mn} = \frac{\exp\{s_m + t_n\}}{\exp\{2\max\{s_m, t_n\}\}\exp\{d_{mn}\}} \tag{19}$$

where s_m and t_n (refer to (17) and (18)) are radiuses of α-cuts of $X_{2m} \times X_{3m}$ and $Y_{2n} \times Y_{3n}$, respectively, and d_{mn} is the Euclidean distance between (x_{2m}, x_{3m}) and (y_{2n}, y_{3n}).

It is easily seen that $S_{mn} = 1$ when two circles (i.e., α-cuts) are completely identical, and furthermore, the following condition is also satisfied

$$0 \leq S_{mn} \leq 1 \tag{20}$$

Before introducing a procedure for ranking objects according to an agent A's preference as defined above, it is necessary to recall the notion of OWA operators.

4.2 OWA Operators

The notion of OWA operators was introduced in [19] regarding the problem of information aggregation. A mapping

$$F : [0, 1]^p \to [0, 1]$$

is called an OWA operator of dimension p if it is associated with a weighting vector $W = [w_1, \ldots, w_p]$, such that 1) $w_i \in [0, 1]$ and 2) $\sum_i w_i = 1$, and

$$F(a_1, \ldots, a_p) = \sum_{i=1}^{p} w_i b_i$$

where b_i is the i-th largest element in the collection $a_1, \ldots, a_p$.

Table 1. Linguistic quantifiers

Linguistic quantifier	Membership function Q
there exists	$Q(r) = \begin{cases} 0 & \text{if } r < 1/n \\ 1 & \text{if } r \geq 1/n \end{cases}$
for all	$Q(r) = \begin{cases} 1 & \text{if } r = 1 \\ 0 & \text{if } r \neq 1 \end{cases}$
almost all	$Q(r) = r$
at least half	$Q(r) = \begin{cases} 2r & \text{if } 0 \leq r \leq 0.5 \\ 1 & \text{if } 0.5 \leq r \leq 1 \end{cases}$
as many as possible	$Q(r) = \begin{cases} 0 & \text{if } 0 \leq r \leq 0.5 \\ 2r-1 & \text{if } 0.5 \leq r \leq 1 \end{cases}$
most	$Q(r) = \begin{cases} 0 & \text{if } 0 \leq r \leq 0.3 \\ 2r-0.6 & \text{if } 0.3 \leq r \leq 0.8 \\ 1 & \text{if } 0.8 \leq r \leq 1 \end{cases}$

One of approaches for obtaining weights w_i's for an OWA operator is to associate some semantics or meaning to them and then, based on these semantics, we can directly determine their values. In the following we introduce the semantics based on fuzzy linguistic quantifiers for weights.

The fuzzy linguistic quantifiers were introduced by Zadeh in [20]. According to Zadeh, there are basically two types of quantifiers: absolute and relative. Here we focus on the relative quantifiers typified by terms such as *most*, *at least half*, *as many as possible*. A relative quantifier Q is defined as a mapping $Q : [0,1] \to [0,1]$ verifying $Q(0) = 0$, there exists $r \in [0,1]$ such that $Q(r) = 1$, and Q is a non-decreasing function.

Table 1 provides a list of linguistic quantifiers associated with their membership functions, all of which are taken from the literature.

Then, Yager [19] proposed to compute the weights w_i's based on the linguistic quantifier represented by Q as follows:

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right), \text{ for } i = 1, \ldots, n. \tag{21}$$

4.3 A Ranking Procedure

With those discussed previously, we now introduce the following procedure which provides a way of evaluating objects in O according to agent A's preference.

1. Determine the membership function Q of linguistic quantifier LQ and the corresponding weighting vector (refer to (21))

$$W_{LQ} = [w_1, \ldots, w_p]$$

2. For each $o_m \in O$, the evaluation of o_m is calculated by

$$V(o_m) = F_{W_{LQ}}(S_{ij_1}, \ldots, S_{ij_p}) = \sum_{k=1}^{K} w_k S_k \tag{22}$$

where S_k **is the** k**-th largest element in the collection** $S_{ij_1}, \ldots, S_{ij_p}$.

3. Rank objects in O according to $V(o_m)$ by a sorting algorithm for crisp numbers.

5 A Case Study for Yamanaka Lacquer

Based on the above considerations, we have developed a prototyping software system for fuzzy correspondence analysis as well as for personalized recommendation. Figure 1 graphically shows a screenshot of the system.

Fig. 1. A screenshot of the system

In the following we will briefly describe a case study for evaluating Yamanaka Lacquer, a traditional craft industry of Ishikawa prefecture. Within the framework of a research project supported by the local government, a total of 45 patterns of Yamanaka Lacquer have been collected for evaluation using kansei data. Due to the limitation of space, we has not presented the detailed analysis of this case study as well as experimental results here.

5.1 Identification of Kansei Features

Before gathering kansei data of these patterns for evaluation, a preliminary research was carried out to select kansei features, consulting with local manufacturers and trading shops. As for the purpose of this research, we did not consider aspects like type or function of products here, but we mainly focus on psychological feelings such as the esteem and aesthetic aspects of human beings on traditional crafts. Criteria used to select kansei features can be summarized as follows:

(i) frequent adjectives used in traditional craft shops, and
(ii) impressive adjectives given by designers from evaluated samples.

At the end of the preliminary research, 26 pairs of opposite kansei words were selected which are shown in Table 2.

Table 2. Pairs of kansei words

n	left kansei word ($\mathbf{w}_n^-$)	$\longleftrightarrow$	right kansei word ($\mathbf{w}_n^+$)
1	standard	□□□□□□□	unconventional
2	simple	□□□□□□□	rich
3	ceremonious	□□□□□□□	funny
4	formal	□□□□□□□	causal
5	serene	□□□□□□□	powerful
6	stillness	□□□□□□□	carousing
7	pretty	□□□□□□□	austere
8	friendly	□□□□□□□	inaccessible
9	soft	□□□□□□□	hard
10	blase	□□□□□□□	attractive
11	flowery	□□□□□□□	quiet
12	happy occasions	□□□□□□□	usual
13	elegant	□□□□□□□	loose
14	delicate	□□□□□□□	largehearted
15	luxurious	□□□□□□□	plain
16	gentle	□□□□□□□	pithy
17	bright	□□□□□□□	dark
18	reserved	□□□□□□□	imperious
19	free	□□□□□□□	regular
20	level	□□□□□□□	indented
21	lustered	□□□□□□□	matte
22	translucent	□□□□□□□	dim
23	warm	□□□□□□□	cool
24	moist	□□□□□□□	arid
25	colorfull	□□□□□□□	sober
26	plain	□□□□□□□	gaudy, loud

5.2 Evaluated Objects and Results

The evaluated objects are collected from local manufactures. A total of 45 patterns of Yamanaka Lacquer (evaluated objects) is prepared together with the

Fig. 2. Yamanaka lacquer #14

Fig. 3. Yamanaka lacquer #15

Fig. 4. Yamanaka lacquer #19

Fig. 5. Yamanaka lacquer #20

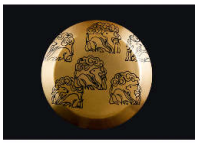


Fig. 6. Yamanaka lacquer #27

Fig. 7. Yamanaka lacquer #36

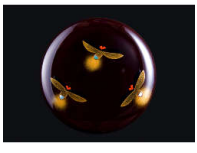


Fig. 8. Yamanaka lacquer #37

Fig. 9. Yamanaka lacquer #42

questionnaire survey. Figures 2–9 show some patterns to be evaluated in this case study. All evaluators are invited to come to an assessment session to provide their evaluations. The obtained data is 3-way data of which each pattern is assessed by all participated evaluators on all 26 kansei features.

Figure 10 shows a screenshot of the result of fuzzy correspondence analysis. In this figure, evaluated objects are displayed as circles in red color, (adjectival) kansei words located at the left hand side (respectively, the right hand side) are displayed as circles in yellow (respectively, blue) color. The circles with semi-transparent color are utilized to see whatever is beneath the shapes when they are overlapped. When we move the mouse across any circle, the name of the corresponding kansei word or object will be appeared such as "yamanaka32" in the graph. In addition, when an evaluated object is selected to see the circumstance, the corresponding circle is stressed by elastic movement.

As we can see at the screenshot shown in Figure 1, the second column shows the list of pairs of opposite kansei words so that from which an user can select a set of kansei words as her preference, where only one kansei word from each pair can be selected. This is due to the observation that, for example, if the user is interested in craft items being *funny*, then she is not interested in those being *ceremonious*, the opposite kansei word of *funny*. In addition, the system also allows users to select different linguistic quantifiers for aggregation according to their own attitudes.

For example, let us consider a recommendation request consisting of three kansei words '*unconventional*', '*funny*' and '*casual*'. Figure 11 shows the result

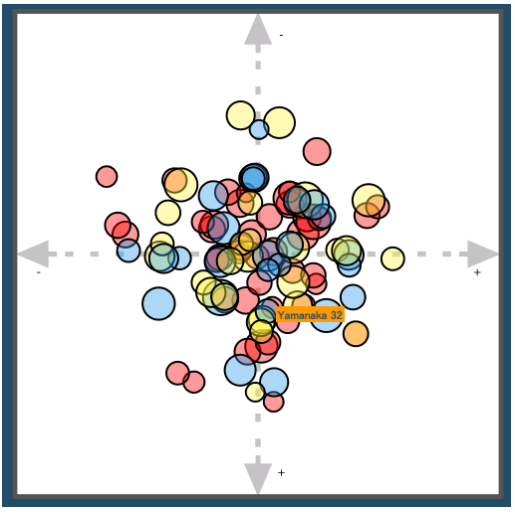


Fig. 10. A result of the fuzzy correspondence analysis

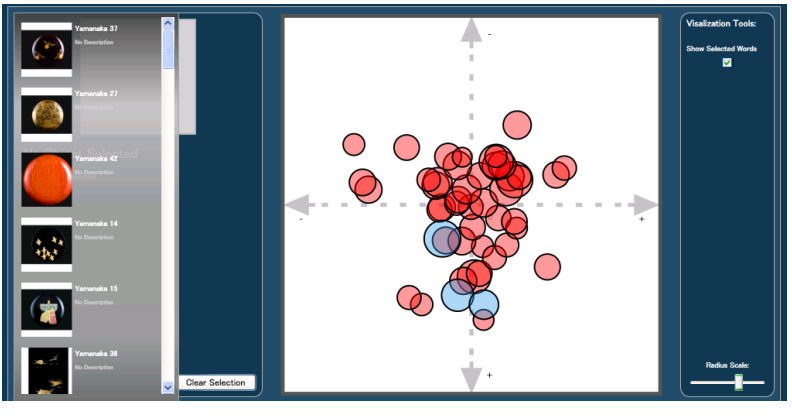


Fig. 11. A ranking of objects (LQ = *all*)

for the case of linguistic quantifier '*all*' being selected for aggregation. While the frame in the left hand side of Figure 11 shows the obtained ranking of patterns according to the recommendation request, the central area displays the result of fuzzy correspondence analysis illustrating the relationship between these kansei words and objects, where three selected kansei words are expressed by three blue circles. Table 3 shows the top 5 of recommended patterns (graphically, see Figures 2–9) according to these three kansei words with different linguistic quantifiers used.

Table 3. Top 5 of recommended patterns

Fuzzy Quantifier	ID Numbers
LQ = *all*	#37 #27 #42 #14 #15
LQ = *almost all*	#37 #27 #42 #15 #14
LQ = *as many as possible*	#27 #37 #20 #42 #19
LQ = *most*	#37 #27 #42 #19 #36
LQ = *at least half*	#20 #27 #19 #37 #36
LQ = *there exists*	#20 #27 #19 #36 #37

6 Concluding Remarks

In this paper, we have presented a technique of for analyzing relative fuzziness in subjective evaluations of traditional crafts using kansei data. Based on this analysis, we have introduced a fitness measure for quantifying how well an evaluated object is compatible with a given kansei word and then proposed a ranking procedure for the problem of personalized recommendation. A case study of evaluating patterns of Yamanaka Lacquer, a traditional craft industry of Ishikawa prefecture, has been also illustrated with the developed prototyping software system.

References

1. Chang, H.-C., Lai, H.-H., Chang, Y.-M.: Expression modes used by consumers in conveying desire for product form: A case study of a car. International Journal of Industrial Ergonomics 36, 3–10 (2006)
2. Choi, K., Jun, C.: A systematic approach to the kansei factors of tactile sense regarding the surface roughness. Applied Ergonomics 38, 53–63 (2007)
3. Horiguchi, A., Suetomi, T.: A kansei engineering approach to a drivervehicle system. International Journal of Industrial Ergonomics 15, 25–37 (1995)
4. Hsu, S.-H., Chuang, M.-C., Chang, C.-C.: A semantic differential study of designers' and users' product form perception. International Journal of Industrial Ergonomics 25, 375–391 (2000)
5. Jindo, T., Hirasago, K.: Application studies to car interior of kansei engineering. International Journal of Industrial Ergonomics 19, 105–114 (1997)
6. Jindo, T., Hirasago, K., Nagamachi, M.: Development of a design support system for office chairs using 3-D graphics. International Journal of Industrial Ergonomics 15, 49–62 (1995)
7. Jiao, J.R., Zhang, Y., Helander, M.: A kansei mining system for affective design. Expert System with Applications 30, 658–673 (2006)
8. Kawano, S., et al.: A context-dependent knowledge model for evaluation of regional environment. Environmental Modelling and Software 20(3), 343–352 (2005)
9. Lai, H.-H., et al.: User-oriented design for the optimal combination on product design. Production Economics 100, 253–267 (2006)
10. Mondragón, S., Company, P., Vergara, M.: Semantic differential applied to the evaluation of machine tool design. International Journal of Industrial Ergonomics 35, 1021–1029 (2005)

11. Nagamachi, M.: Kansei Engineering: a new ergonomic consumer-oriented technology for product development. International Journal of Industrial Ergonomics 15, 3–11 (1995)
12. Nagamachi, M.: Kansei engineering as a powerful consumer-oriented technology for product development. Applied Ergonomics 33, 289–294 (2002)
13. Nakamori, Y., Ryoke, M.: Treating fuzziness in subjective evaluation data. Information Sciences 176, 3610–3644 (2006)
14. Osgood, Ch.E., Suci, G.J., Tannenbaum, P.H.: The Measurement of Meaning. University of Illinois Press, Urbana (1957)
15. Petiot, J.-F., Yannou, B.: Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics. International Journal of Industrial Ergonomics 33, 507–525 (2004)
16. Schütte, S.: Engineering Emotional Values in Product Design – Kansei Engineering in Development. Institute of Technology, Linköpings Universitet, Dissertation 951 (2005)
17. The Association for the Promotion of Traditional Craft Industries, http://www.kougei.or.jp/english/promotion.html
18. The Japanese Ministry of Economy, Trade and Industry (METI), http://www.kansai.meti.go.jp/english/dentousangyou/top_page.html
19. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man, and Cybernetics 18, 183–190 (1988)
20. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Computers and Mathematics with Applications 9, 149–184 (1983)
21. Zadeh, L.A.: The concept of a linguistic variable and its applications to approximate reasoning. Part I. Information Sciences 8, 199–249 (1975)

A Probability-Based Approach to Consumer Oriented Evaluation of Traditional Craft Items Using Kansai Data

Hongbin Yan, Van Nam Huynh, and Yoshiteru Nakamori

Japan Advanced Institute of Science and Technology (JAIST)
Nomi, Ishikawa 923-1292, Japan
{hongbinyan,huynh,nakamori}@jaist.ac.jp

Summary. This paper deals with the evaluation problem of Japanese traditional crafts in which product items are evaluated according so-called kansei features by means of the semantic differential method. As for traditional crafts, decisions on which items to buy or use are usually influenced by personal feelings/characteristics, we shall propose a consumer-oriented evaluation model targeting on those recommendation requests specified by consumers. Particularly, given a consumer's recommendation request, the proposed model aims at defining an evaluation function that quantifies how well a product item meets the consumer's feeling preference. An application for evaluating patterns of Kutani porcelain is conducted to illustrate how the proposed evaluation model works practically.

1 Introduction

Kansei Engineering, invented by Mitsuo Nagamachi at Hiroshima University in the 1970s, is defined by its founder as a translating technology of a consumer's feeling and image for a product into design elements [11]. According to Nagamachi [18], Kansei is "the impression somebody gets from a certain artefact, environment or situation using all her senses of sight, hearing, feeling, smell, taste [and sense of balance] as well as their recognition". For building a kansei database on psychological feelings of products, the most commonly-used method is to choose (adjectival) kansei words first and then ask people to express their feelings using those words in which the semantic differential (SD) method [14] or its modifications are often used, e.g., [4, 8, 12, 18].

Generally, evaluations for ranking and for selection are two closely related and common facets of human decision-making activities in practice. Many studies of Kansei Engineering or other consumer-oriented design techniques have involved an evaluation process in which for example a design could be selected for production [15]. However, kansei data-based evaluation for commercial products have generally received less attention [10], in particular those for traditional products have not been addressed yet [17]. These evaluations would be helpful for marketing or recommendation purposes and particularly more important in the

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 326–340, 2008.
springerlink.com

era of e-commerce, where recommender systems have become an important research area [1].

The main purpose of this paper is to propose and develop an evaluation model for Japanese traditional crafts using kansei data and consumer-specified preferences. A particular focus is put on the evaluation of traditional craft products of Ishikawa prefecture, located in Japan. It should be emphasized here that the esteem and aesthetic aspects play a crucial role in human beings perception on traditional crafts, therefore kansei data are essential and necessary for the evaluation.

In order to do so, a preliminary research is conducted to select kansei words and then a population of subjects is called for gathering evaluation data of traditional crafts, here the SD method is used for this purpose. Using the voting statistics, these kansei data are then used to generate kansei profiles of evaluated patterns which will be served as the knowledge for a consumer-oriented evaluation later. Because the preference on traditional crafts varies from person to person depending on personal character, feeling and aesthetic, assuming we are given a consumer c's preference also expressed by kansei words like *bright*, *elegant*, etc., we then define an evaluation function $V : \mathcal{O} \to [0, 1]$, where $\mathcal{O}$ is the set of evaluated patterns or objects and $V(o)$ is interpreted as the degree to which the pattern o meets c's preference. Basically, our approach is based on the appealing idea of target-based decision analysis [2] that a consumer would be only interested in patterns meeting her feeling needs. It particular, following our model, the preference order on the qualitative scale according to a kansei feature will be determined adaptively depending on a particular consumer's preference. Furthermore, viewing multi-person assessments as uncertain judgments regarding kansei features of traditional craft items, making use of a similar idea as in uncertain decision making with fuzzy targets [7], we shall be able to work out the probability that judgment on a kansei feature of each item meets the feeling target set on this feature by the consumer. Then, guided by the linguistic quantifier also specified by the consumer, an appropriate ordered weighted averaging (OWA) operator [19] can be used to define the evaluation function V as mentioned above.

The paper is organized as follows. Section 2 begins with a brief description of the research context and follows by the formulation of research problem. Section 3 introduces a consumer-oriented evaluation model using kansei data and Section 4 applies the proposed model to a case study of evaluating Kutani porcelain, one of traditional craft products of Ishikawa. Finally, some concluding remarks are presented in Section 5.

2 Preliminaries

2.1 OWA Operators and Linguistic Quantifiers

The notion of OWA operators was first introduced in [19] regarding the problem of aggregating multi-criteria to form an overall decision function. By definition, an OWA operator of dimension n is a mapping

$$\mathcal{F} : [0,1]^n \to [0,1]$$

associated with a weighting vector $w = [w_1, \ldots, w_n]$, such that 1) $w_i \in [0,1]$ and 2) $\sum_i w_i = 1$, and

$$\mathcal{F}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i b_i$$

where b_i is the i-th largest element in the collection $\{a_1, \ldots, a_n\}$.

OWA operators provide a type of aggregation operators which lie between the "AND" and the "OR" aggregation. As suggested by Yager [19], there exist at least two methods for obtaining weights w_i's. The first approach is to use some kind of learning mechanism. That is, we use some sample data, arguments and associated aggregated values and try to fit the weights to this collection of sample data. The second approach is to give some semantics or meaning to the weights. Then, based on these semantics we can directly provide the values for the weights. As for the purpose of this paper, let us introduce the semantics based on linguistic quantifiers for the weights.

The fuzzy linguistic quantifiers were introduced by Zadeh in [22]. According to Zadeh, there are basically two types of quantifiers: absolute, and relative. Here we focus on the relative quantifiers typified by terms such as *most*, *at least half*, *as many as possible*. A relative quantifier Q is defined as a mapping $Q : [0,1] \to [0,1]$ verifying $Q(0) = 0$, there exists $r \in [0,1]$ such that $Q(r) = 1$, and Q is a non-decreasing function. For example, the membership function of relative quantifiers can be simply defined as

$$Q(r) = \begin{cases} 0 & \text{if } r < a \\ \frac{r-a}{b-a} & \text{if } a \le r \le b \\ 1 & \text{if } r > b \end{cases} \tag{1}$$

with parameters $a, b \in [0,1]$.

Then, Yager [19] proposed to compute the weights w_i's based on the linguistic quantifier represented by Q as follows:

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right), \text{ for } i = 1, \ldots, n. \tag{2}$$

In addition, the so-called measure '*orness*' of OWA operator $\mathcal{F}$ associated with weighting vector $w = [w_1, \ldots, w_n]$ is defined as

$$orness(\mathcal{F}) = \frac{1}{n-1} \sum_{i=1}^{n} ((n-i) \times w_i) \tag{3}$$

This measure of '*orness*' indicates how much degree the operator $\mathcal{F}$ behaves like an 'OR' aggregation. Also, the measure of '*andness*' associated with $\mathcal{F}$ is defined as the complement of its '*orness*', then

$$andness(\mathcal{F}) = 1 - orness(\mathcal{F}) \tag{4}$$

which indicates how much degree the operator $\mathcal{F}$ behaves like an 'AND' aggregation.

2.2 Formulation of the Problem

The consumer-oriented evaluation process for traditional crafts using kansei data is graphically described as in Fig. 1. Let us denote $\mathcal{O}$ the collection of craft patterns to be evaluated and N is the cardinality of $\mathcal{O}$, i.e. $N = |\mathcal{O}|$.

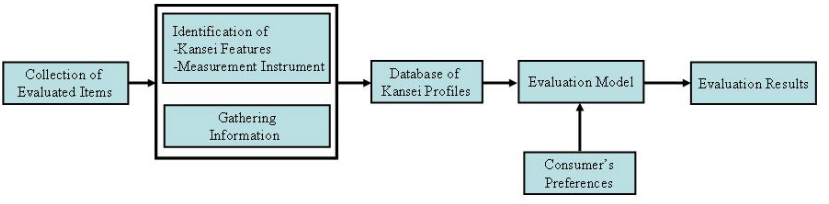


Fig. 1. The framework of consumer-oriented evaluation using kansei data

Identification of kansei features and measurement instrument

As mentioned previously, the first task in the kansei data-based evaluation process is to identify what kansei features we as human beings often use to express our feelings regarding traditional crafts. In the present research project, kansei features are selected through a brainstorming process by those attending including relevant researchers, senior residents and certificated masters of traditional crafts. Each kansei feature is defined by an opposite pair of (adjectival) kansei words, for example the *fun* feature determines the pair of kansei words *solemn* and *funny*. Let

1. $\{F_1, \ldots, F_K\}$ be the set of kansei features selected,
2. $\mathbf{w}_k^+$ and $\mathbf{w}_k^-$ be the opposite pair of kansei words corresponding to F_k, for $k = 1, \ldots, K$. Denote $\mathbf{W}$ the set of kansei words, i.e. $\mathbf{W} = \{\mathbf{w}_k^+, \mathbf{w}_k^- | k = 1, \ldots, K\}$.

Then, SD method [14] is used as a measurement instrument to design the questionnaire for gathering kansei evaluation data.

Gathering information

The questionnaire using SD method for gathering information consists in listing the kansei features, each of which corresponds to an opposite pair of kansei words that lie at either end of a qualitative M-point scale, where M is an odd positive integer as used, for example, in 5-point scale [13], 7-point scale [10] or 9-point scale [9]. In our model, the qualitative scale are treated as categorical scale, then we symbolically denote the M-point scale by

$$\mathbb{V} = \{v_1, \ldots, v_M\}$$

where $\mathbf{w}_k^+$ and $\mathbf{w}_k^-$ are respectively assumed to be at the ends v_1 and v_M.

The questionnaire is then distributed to a population $\mathcal{P}$ of subjects who are invited to express their emotional assessments according each kansei feature of

craft patterns in $\mathcal{O}$ by using the M-point scale. Formally, we can model the kansei data of each craft pattern $o_i \in \mathcal{O}$ according to kansei features obtained from the assessment of subjects s_j in $\mathcal{P}$ as shown in Table 1, where $x_{jk}(o_i) \in \mathbb{V}$, for $j = 1, \ldots, P = |\mathcal{P}|$ and $k = 1, \ldots, K$.

Table 1. The kansei assessment data of pattern o_i

	Kansei Features			
Subjects	F_1	D_2	$\cdots$	F_K
s_1	$x_{11}(o_i)$	$x_{12}(o_i)$	$\cdots$	$x_{1K}(o_i)$
s_2	$x_{21}(o_i)$	$x_{22}(o_i)$	$\cdots$	$x_{2K}(o_i)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
s_m	$x_{P1}(o_i)$	$x_{p2}(o_i)$	$\cdots$	$x_{pK}(o_i)$

Problem

The kansei assessment database built, as described above, will be utilized to generate the knowledge serving for the following evaluation problem. Assume that an agent as a potential consumer is interested in looking for a craft pattern which would meet her preference given by a proper subset W of the set $\mathbf{W}$ of kansei words as defined below. She may then want to rate craft patterns available in $\mathcal{O}$ according to her preference. In particular, we are concerned with consumer-specified recommendation requests which can be stated generally in forms of the following statement:

> "I should like craft patterns which would best meet LQ (of) my preference specified in $W \subset \mathbf{W}$" ($\bigstar$)

where LQ is a linguistic quantifier such as *all*, *most*, *at least half*, *as many as possible*, etc. Formally, the problem can be formulated as follows.

Given $W = \{\mathbf{w}^*_{k_1}, \ldots, \mathbf{w}^*_{k_n}\}$ and LQ corresponding to the recommendation request specified by an agent as linguistically stated in ($\bigstar$), where $*$ stands for either $+$ or $-$, and $\{k_1, \ldots, k_n\} \subseteq \{1, \ldots, K\}$, the problem now is how to evaluate craft patterns in $\mathcal{O}$ using kansei data and the recommendation request specified as the pair $[W, LQ]$? Here, by $*$ standing for either $+$ or $-$ as above, it indicates that only one of the two $\mathbf{w}^+_{k_l}$ and $\mathbf{w}^-_{k_l}$ $(l = 1, \ldots, n)$ presents in W, which may be psychologically reasonable to assume. For example, if the agent is interested in craft items being *funny* according to kansei feature of *fun*, then she is not interested in those being *solemn*, the opposite kansei word of *funny*. This evaluation problem will be solved by a so-called consumer-oriented evaluation model presented in the next section.

3 A Consumer-Oriented Evaluation Model

In this section we shall propose a consumer-oriented evaluation model based on the idea that a consumer should be only interested in craft patterns that would

best meet her psychological needs from the esteem/aesthetic point of view. Let us denote $\mathbf{D}$ the kansei assessment database about a finite set $\mathcal{O}$ of craft patterns using SD method as mentioned previously, and $\mathbf{D}[o_i]$ the data of pattern o_i $(i = 1, \ldots, N)$ as shown in Table 1. The proposed model basically consists of the following main steps. The first step is to generate a kansei profile for each pattern o_i using its data $\mathbf{D}[o_i]$ based on voting statistics. Then in the second step, given the recommendation request specified by a consumer c as a pair $[W, LQ]$, an evaluation function $V : \mathcal{O} \to [0, 1]$ is defined taking c's recommendation request into consideration. Lastly, a ranking order for all patterns in $\mathcal{O}$ is determined according to this function V as an answer to the recommendation request. In the following, we will describe these three steps in detail.

3.1 Generating Kansei Profiles

For each pattern o_i with its assessment data $\mathbf{D}[o_i]$ shown in Table 1, we define for each kansei feature F_k, $k = 1, \ldots, K$, a probability distribution $f_{ik} : \mathbb{V} \to [0, 1]$ as follows:

$$f_{ik}(v_h) = \frac{|\{s_j \in \mathcal{P} : x_{jk}(o_i) = v_h\}|}{|\mathcal{P}|} \tag{5}$$

This distribution f_{ik} is considered as an uncertain judgment of craft pattern o_i according to kansei feature F_k. By the same way, we can obtain a K-tuple of distributions $[f_{i1}, \ldots, f_{iK}]$ regarding the kansei assessment of o_i and call the tuple the kansei profile of o_i. Similarly, kansei profiles of all patterns in $\mathcal{O}$ can be generated from $\mathbf{D}$.

It should be also emphasized here that in many papers of Kansei Engineering or other methodologies of consumer-oriented design support using SD method for gathering data, populations of subjects involved in experimental studies have a size ranging typically from 10 to 35 people, (cf. [3, 9, 10, 15]). However, for the purpose of consumer-oriented evaluation which the present paper is aiming at, such a small size of the population $\mathcal{P}$ may cause a statistical bias as well as may provide not enough information from various points of view to possibly reduce the subjectivity of the evaluation. Therefore, a population of subjects with a much larger size has been used for gathering kansei data. In addition, in order to increase the reliability of subjective judgements, all subjects were invited to participate at a centralized face-to-pattern evaluation session on a designed date. For example, in the case study of Kutani porcelain described in the following section, a population of 211 subjects was used at a centralized evaluation session.

3.2 Evaluation Function

Having generated kansei profiles for all patterns $o_i \in \mathcal{O}$ as above, we now define the evaluation function V corresponding to the recommendation request $(\bigstar)$ symbolically specified as $[W, LQ]$, where $W = \{\mathbf{w}^*_{k_1}, \ldots, \mathbf{w}^*_{k_n}\}$ and LQ is a linguistic quantifier.

Intuitively, if a consumer expresses her preference on a kansei feature such as *color contrast* with kansei word *bright*, she may implicitly assume a preference order on the semantic differential scale corresponding to *color contrast* towards the end v_1 where *bright* is placed. Conversely, if the consumer's preference on *color contrast* was *dark*, i.e. the opposite kansei word of *bright*, she would assume an inverse preference order on the scale towards the end v_M where *dark* is placed. In other words, in consumer-oriented evaluation using kansei data, the preference order on the semantic differential scale corresponding to a kansei feature should be determined adaptively depending on a particular consumer's preference. This can be formally formulated as below.

For each $\mathbf{w}^*_{k_l} \in W$, we define a linear preference order $\succeq_l$ on $\mathbb{V}$ according to the kansei feature F_{k_l} as follows

$$v_h \succeq_l v_{h'} \Leftrightarrow \begin{cases} h' \geq h, & \text{if } \mathbf{w}^*_{k_l} = \mathbf{w}^+_{k_l} \\ h \geq h', & \text{if } \mathbf{w}^*_{k_l} = \mathbf{w}^-_{k_l} \end{cases} \tag{6}$$

In addition, due to vagueness inherent in consumer's expression of preference in terms of kansei words, each $\mathbf{w}^*_{k_l}$ is considered as the feeling target, denoted by T_{k_l}, of the consumer according to kansei feature F_{k_l}, which can be represented as a possibility variable [21] on $\mathbb{V}$ whose possibility distribution is defined as

$$\pi_{k_l}(v_h) = \begin{cases} \frac{M-h}{M-1}, & \text{if } \mathbf{w}^*_{k_l} = \mathbf{w}^+_{k_l} \\ \frac{h-1}{M-1}, & \text{if } \mathbf{w}^*_{k_l} = \mathbf{w}^-_{k_l} \end{cases} \tag{7}$$

Fig. 2 graphically illustrates these concepts.

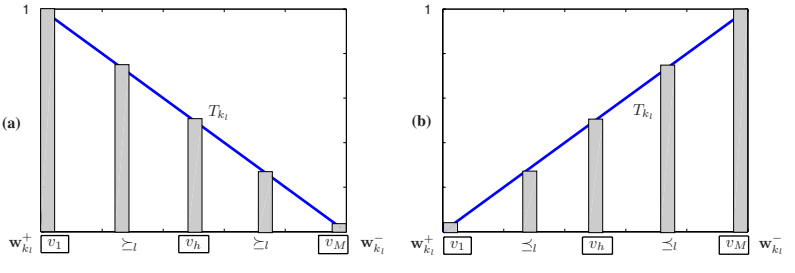


Fig. 2. The preference order $\succeq_l$ and the possibility distribution of feeling target T_{k_l}: (a) $\mathbf{w}^*_{k_l} = \mathbf{w}^+_{k_l}$; (b) $\mathbf{w}^*_{k_l} = \mathbf{w}^-_{k_l}$

As such, with the consumer's preference specified by W, we obtain n feeling targets T_{k_l} $(l = 1, \ldots, n)$ accompanying with n preference orders $\succeq_l$ $(l = 1, \ldots, n)$ on the semantic differential scale of kansei features F_{k_l} $(l = 1, \ldots, n)$, respectively. Recall that, for each $l = 1, \ldots, n$, the uncertain judgment of each craft pattern o_i regarding the kansei feature F_{k_l} is represented by the probability distribution f_{ik_l} over $\mathbb{V}$, as defined previously. Bearing these considerations

in mind, we are now able to evaluate, for each $l = 1, \ldots, n$, how the feeling performance of a pattern o_i on F_{k_l}, denoted by $F_{k_l}(o_i)$ and represented by f_{ik_l}, meets the feeling target T_{k_l} representing consumer's preference on F_{k_l}. This can be done as follows.

Firstly, making use of the possibility-probability conversion method [20] we can transform the possibility distribution of feeling target T_{k_l} into an associated probability distribution, denoted by $\hat{p}_{k_l}$, via the simple normalization as follows

$$\hat{p}_{k_l}(v_h) = \frac{\pi_{k_l}(v_h)}{\sum\limits_{v \in \mathbb{V}} \pi_{k_l}(v)} \tag{8}$$

Then, by accepting the assumption that the feeling target T_{k_l} is stochastically independent of feeling performance on F_{k_l} of any pattern o_i, we can work out the probability that the feeling performance $F_{k_l}(o_i)$ meets the feeling target T_{k_l}, denoted by $\mathbf{P}(F_{k_l}(o_i) \succeq T_{k_l})$, in terms of the preference order $\succeq_l$ as

$$\begin{aligned} \mathbf{P}(F_{k_l}(o_i) \succeq T_{k_l}) &\triangleq P(f_{ik_l} \succeq_l \hat{p}_{k_l}) \\ &= \textstyle\sum_{h=1}^{M} f_{ik_l}(v_h) P(v_h \succeq_l \hat{p}_{k_l}) \end{aligned} \tag{9}$$

where $P(v_h \succeq_l \hat{p}_{k_l})$ is the cumulative probability function defined by

$$P(v_h \succeq_l \hat{p}_{k_l}) = \sum_{v_h \succeq_l v_{h'}} \hat{p}_{k_l}(v_{h'}) \tag{10}$$

It is of interest noting here that a similar idea has also been recently used in [6] for developing the so-called satisfactory-oriented linguistic decision model. Intuitively, the quantity $\mathbf{P}(F_{k_l}(o_i) \succeq T_{k_l})$ defined above could be interpreted as the probability of "the feeling performance on F_{k_l} of o_i meeting the feeling target T_{k_l} specified by a consumer on F_{k_l}". Then, after having these probabilities $\mathbf{P}(F_{k_l}(o_i) \succeq T_{k_l}) = \mathbf{P}_{k_l i}$, for $l = 1, \ldots, n$, we are able to aggregate all of them to obtain an aggregated value with taking the linguistic quantifier LQ into account, making use of the so-called ordered weighted averaging (OWA) aggregation operator [19].

Under such a semantics of OWA operators, now we are ready to define the evaluation function, for any $o_i \in \mathcal{O}$, as follows

$$\begin{aligned} V(o_i) &= \mathcal{F}(\mathbf{P}_{k_1 i}, \ldots, \mathbf{P}_{k_n i}) \\ &= \textstyle\sum_{l=1}^{n} w_l \mathbf{P}_{li} \end{aligned} \tag{11}$$

where $\mathbf{P}_{li}$ is the l-th largest element in the collection $\mathbf{P}_{k_1 i}, \ldots, \mathbf{P}_{k_n i}$ and weighting vector $[w_1, \ldots, w_n]$ is determined directly by using a fuzzy set-based semantics of the linguistic quantifier LQ. As interpreted previously on quantities $\mathbf{P}_{k_l i}$ $(l = 1, \ldots, n)$, the aggregated value $V(o_i)$ therefore indicates the degree to which craft pattern o_i meets the feeling preference derived from the recommendation request specified by a consumer as $[W, LQ]$.

3.3 Rating Craft Patterns

Based on the evaluation function (11) defined above, a rating of all the craft patterns o_i in $\mathcal{O}$ can be straightforwardly determined according to their values $V(o_i)$ by a sorting algorithm for real numbers. The obtained rating is then considered as the solution to the recommendation request $[W, LQ]$. For the sake of convenience, the evaluation procedure for recommendation described above is summarized and algorithmically presented in Fig. 3.

Input: A recommendation request $[W, LQ]$

- $W = \{\mathbf{w}^*_{k_1}, \ldots, \mathbf{w}^*_{k_n}\}$–feeling preference.
- LQ – linguistic quantifier.

Output: A rating of all items in $\mathcal{O}$

for each $\mathbf{w}^*_{k_l} \in W$ **do**
  determine the preference order $\succeq_l$ on $\mathbb{V}$ for kansei feature F_{k_l} via (6)
  determine the feeling target T_{k_l} on F_{k_l} and its possibility distribution via (7)
end for
determine weighting vector $[w_1, \ldots, w_n]$ using the fuzzy set based semantics of linguistic quantifier LQ
for each $o_i \in \mathcal{O}$ **do**
  for each $\mathbf{w}^*_{k_l} \in W$ **do**
    compute $\mathbf{P}(F_{k_l}(o_i) \succeq T_{k_l})$ via (9)
  end for
  compute $V(o_i)$ via (11)
end for
rank items o_i according to their values $V(o_i)$

Fig. 3. The recommendation procedure

4 Application to Kutani Porcelain

In this section we shall apply the proposed model to evaluating Kutani porcelain, a traditional craft industry in Japan, historically back to the seventeenth century, of Kutani Pottery Village in Ishikawa prefecture[1]. Within the framework of a research project supported by the local government, a total of 30 patterns of Kutani porcelain have been collected for the kansei data-based evaluation [16].

Before gathering kansei assessment data of these patterns for evaluation, a preliminary research was carried out to select kansei features, consulting with local manufacturers and trading shops. Finally, 26 opposite pairs of kansei words were selected at the end of a brainstorming process. Kansei words are approximately translated into English as shown in Table 2 (adapted from [16]).

[1] http://shofu.pref.ishikawa.jp/shofu/intro_e/HTML/H_S50402.html

Table 2. Opposite pairs of kansei words used for the evaluation

F_k	*Left kansei word*	v_1	v_2	v_3	v_4	v_5	v_6	v_7	*Right kansei word*
1	conventional($\mathbf{w}_1^+$)	□	□	□	□	□	□	□	unconventional($\mathbf{w}_1^-$)
2	simple($\mathbf{w}_2^+$)	□	□	□	□	□	□	□	compound($\mathbf{w}_2^-$)
3	solemn($\mathbf{w}_3^+$)	□	□	□	□	□	□	□	funny($\mathbf{w}_3^-$)
4	formal($\mathbf{w}_4^+$)	□	□	□	□	□	□	□	causal($\mathbf{w}_4^-$)
5	serene($\mathbf{w}_5^+$)	□	□	□	□	□	□	□	forceful($\mathbf{w}_5^-$)
6	still($\mathbf{w}_6^+$)	□	□	□	□	□	□	□	moving($\mathbf{w}_6^-$)
7	pretty($\mathbf{w}_7^+$)	□	□	□	□	□	□	□	austere($\mathbf{w}_7^-$)
8	friendly($\mathbf{w}_8^+$)	□	□	□	□	□	□	□	unfriendly($\mathbf{w}_8^-$)
9	soft($\mathbf{w}_9^+$)	□	□	□	□	□	□	□	hard($\mathbf{w}_9^-$)
10	blase($\mathbf{w}_{10}^+$)	□	□	□	□	□	□	□	attractive($\mathbf{w}_{10}^-$)
11	flowery($\mathbf{w}_{11}^+$)	□	□	□	□	□	□	□	quiet($\mathbf{w}_{11}^-$)
12	happy($\mathbf{w}_{12}^+$)	□	□	□	□	□	□	□	normal($\mathbf{w}_{12}^-$)
13	elegant($\mathbf{w}_{13}^+$)	□	□	□	□	□	□	□	loose($\mathbf{w}_{13}^-$)
14	delicate($\mathbf{w}_{14}^+$)	□	□	□	□	□	□	□	large-hearted($\mathbf{w}_{14}^-$)
15	luxurious($\mathbf{w}_{15}^+$)	□	□	□	□	□	□	□	frugal($\mathbf{w}_{15}^-$)
16	gentle($\mathbf{w}_{16}^+$)	□	□	□	□	□	□	□	pithy($\mathbf{w}_{16}^-$)
17	bright($\mathbf{w}_{17}^+$)	□	□	□	□	□	□	□	dark($\mathbf{w}_{17}^-$)
18	reserved($\mathbf{w}_{18}^+$)	□	□	□	□	□	□	□	imperious($\mathbf{w}_{18}^-$)
19	free($\mathbf{w}_{19}^+$)	□	□	□	□	□	□	□	regular($\mathbf{w}_{19}^-$)
20	level($\mathbf{w}_{20}^+$)	□	□	□	□	□	□	□	indented($\mathbf{w}_{20}^-$)
21	lustered($\mathbf{w}_{21}^+$)	□	□	□	□	□	□	□	matte($\mathbf{w}_{21}^-$)
22	transpicuous($\mathbf{w}_{22}^+$)	□	□	□	□	□	□	□	dim($\mathbf{w}_{22}^-$)
23	warm($\mathbf{w}_{23}^+$)	□	□	□	□	□	□	□	cool($\mathbf{w}_{23}^-$)
24	moist($\mathbf{w}_{24}^+$)	□	□	□	□	□	□	□	arid($\mathbf{w}_{24}^-$)
25	colorful($\mathbf{w}_{25}^+$)	□	□	□	□	□	□	□	sober($\mathbf{w}_{25}^-$)
26	plain($\mathbf{w}_{26}^+$)	□	□	□	□	□	□	□	gaudy, loud($\mathbf{w}_{26}^-$)

4.1 Gathering Data and Kansei Profiles

Several assessment sessions, with a total of 211 subjects invited to participate in, were held to gather kansei data. The data obtained is 3-way data of which each pattern Kutani#i ($i = 1, \ldots, 30$) is assessed by all participated subjects on all kansei features F_k, $k = 1, \ldots, 26$.

The 3-way data is then used to generate kansei profiles for patterns via (5) as mentioned previously. These kansei profiles are considered as (uncertain) feeling assessments of patterns serving as the knowledge for consumer-oriented evaluation.

4.2 Consumer-Oriented Evaluation

To illustrate how the model proposed in the preceding section works, let us consider the following example.

Assuming a consumer's recommendation request is specified as

$$[\{\mathbf{w}_3^-, \mathbf{w}_7^+, \mathbf{w}_{10}^-, \mathbf{w}_{11}^+\}, \textit{as many as possible}]$$

That is, verbally, she would ask for craft patterns meeting *as many as possible* her feeling preference of *funny*, *pretty*, *attractive* and *flowery*.

According to the evaluation procedure shown in Fig. 3, we first determine preference orders on $\mathbb{V} = \{v_1, \ldots, v_7\}$ for features F_3, F_7, F_{10}, F_{11}. Using (6), we have $\succeq_3 = \succeq_{10}$, $\succeq_7 = \succeq_{11}$ and $v_7 \succeq_3 \ldots \succeq_3 v_1$ and $v_1 \succeq_7 \ldots \succeq_7 v_7$. Then, using (7) we define feeling targets T_3, T_7, T_{10}, T_{11} on features F_3, F_7, F_{10}, F_{11}, respectively. In particular, we again have $T_3 \equiv T_{10}$, $T_7 \equiv T_{11}$ with possibility distributions shown in Fig. 4.

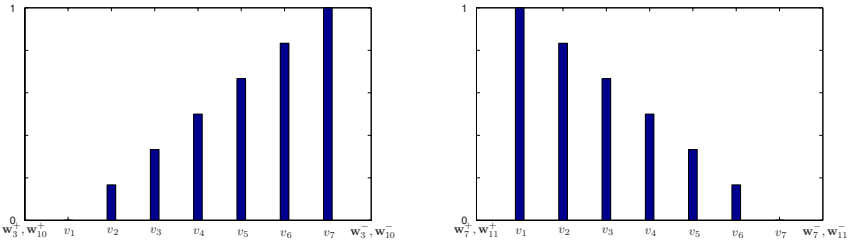


Fig. 4. Possibility distribution of feeling targets

We now determine the weighting vector $[w_1, w_2, w_3, w_4]$ according to the fuzzy set based semantics of linguistic quantifier '*as many as possible*'. Assume that the membership function of the quantifier '*as many as possible*' is defined as a mapping $Q : [0, 1] \to [0, 1]$ such that [5]

$$Q(r) = \begin{cases} 0 & \text{if } 0 \le r \le 0.5 \\ 2r - 1 & \text{if } 0.5 \le r \le 1 \end{cases}$$

We then obtain the weighting vector as $[0, 0, 0.5, 0.5]$ using Yager's method (refer to (2)).

With these preparations done, we are now ready to use (8) and (9) for computing probabilities $\mathbf{P}_{3i}$, $\mathbf{P}_{7i}$, $\mathbf{P}_{10i}$, $\mathbf{P}_{11i}$ of meeting corresponding feeling targets T_3, T_7, T_{10}, T_{11} for each pattern Kutani#i ($i = 1, \ldots, 30$). Then, using (11) we have

$$V(\text{Kutani}\#i) = \mathcal{F}(\mathbf{P}_{3i}, \mathbf{P}_{7i}, \mathbf{P}_{10i}, \mathbf{P}_{11i})$$

where $\mathcal{F}$ is the OWA operator of dimension 4 associated with the weighting vector $[0, 0, 0.5, 0.5]$.

Table 3. Quantifiers used and corresponding top 3 patterns

Linguistic quantifier	Weighting vector	The top 3 patterns
As many as possible (AMAP)	$[0, 0, 0.5, 0.5]$	$\#14 \succeq \#10 \succeq \#18$
All	$[0, 0, 0, 1]$	$\#14 \succeq \#18 \succeq \#30$
There exists ($\exists$)	$[1, 0, 0, 0]$	$\#10 \succeq \#29 \succeq \#3$
At least haft (ALH)	$[0.5, 0.5, 0, 0]$	$\#10 \succeq \#29 \succeq \#3$

Finally, a ranking of patterns Kutani#i, $i = 1, \ldots, 30$, according to their values V(Kutani#i) can be easily obtained. Table 3 shows the top three patterns that would best meet the feeling preference of *funny*, *pretty*, *attractive* and *flowery* with different typical linguistic quantifiers used.

4.3 Discussion

To facilitate the discussion of the obtained results, all the target achievements of recommended patterns on selected features as well as their aggregated values corresponding to different quantifiers used are shown in Table 4. Accordingly, the target achievements of recommended patterns on selected features as well as their aggregated values corresponding to different quantifiers used are shown in Table 4.

Table 4. Target Achievements on Selected Features of Recommended Patterns

	Target Achievements				Aggregated Values			
Patterns	F_3	F_7	F_{10}	F_{11}	$\exists$	ALH	AMAP	*All*
#03	0.4159	0.61776	0.5099	0.18767	**0.61776**	**0.56383**	0.30148	0.18767
#10	0.59957	0.538	0.32924	0.799	**0.799**	**0.69929**	**0.43362**	0.32924
#14	0.48605	0.59976	0.39843	0.5259	0.59976	0.56283	**0.44224**	**0.39843**
#18	0.40033	0.53248	0.39043	0.43967	0.53248	0.48607	**0.39538**	**0.39043**
#29	0.6269	0.55824	0.35995	0.35771	**0.6269**	**0.59257**	0.35883	0.35771
#30	0.49081	0.48695	0.36848	0.363	0.49081	0.48888	0.36574	**0.363**

First, let us consider the result according to the use of quantifier '*there exists*'. As shown in Table 3, in this case we have the top three patterns are, in order of preference, #10, #29, and #03. According to Table 4 and as graphically illustrated by Fig. 5, one can intuitively observe that pattern #10 meets very well the feeling target *flowery* ($\mathbf{w}_{11}^{+}$), follow by *funny* ($\mathbf{w}_{3}^{-}$) of pattern #29 and *pretty* ($\mathbf{w}_{7}^{+}$) of pattern #3. In the case where quantifier '*at least half*' is used instead of '*there exists*', we still obtain the same result. This is due to, beside the feeling target *flowery*, pattern #10 has well met the target *funny* and, moreover, beside *funny* and *pretty*, pattern #29 and #03 are quite good at meeting *pretty* and *attractive*, respectively. It would be worth noting here that aggregation operator $\mathcal{F}$ with weighting vector corresponding to quantifier '*there exists*' is a pure 'OR' operator, and the one corresponding to quantifier '*at least half*' still behaves toward an 'OR' aggregation as well; namely, the degree of '*orness*' (refer to (3)) associated with the operator $\mathcal{F}$ of quantifier *at least half* is

$$orness(\mathcal{F}) = \frac{1}{3}(3 \times 0.5 + 2 \times 0.5) = 0.833$$

Now let us look at the case of using quantifier '*as many as possible*'. Then we obtain, in the order of preference, patterns #14, #10 and #18 as the top three. In this case, due to the requirement of meeting as many as possible of the four feeling targets {*funny*, *pretty*, *attractive*, *flowery*}, the aggregation operator $\mathcal{F}$

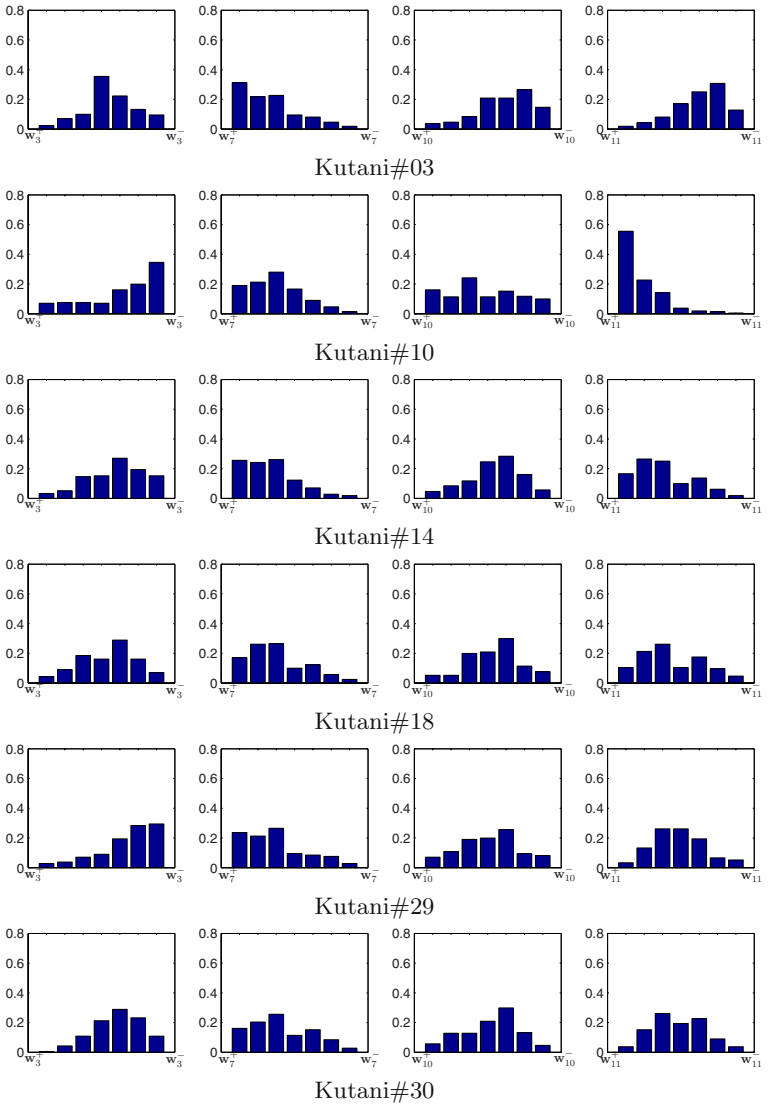


Fig. 5. Recommended patterns' uncertain judgments on selected features

behaves toward an 'AND' aggregation with the corresponding degree of '*andness*' (refer to (4)) being

$$\begin{aligned} andness(\mathcal{F}) &= 1 - orness(\mathcal{F}) \\ &= 1 - \tfrac{1}{3}(1 \times 0.5 + 0 \times 0.5) \\ &= 0.833 \end{aligned}$$

Then, pattern #14 in this case becomes the most recommended one as having the highest aggregated value which is the average of its two lowest degrees of target

achievement for *funny* and *attractive*, while having good scores in achieving *pretty* and *flowery* targets. Pattern #10 appears as the second recommended item because of having a good score in meeting *pretty* target and not so bad score in meeting *attractive* one, beside very well scores of target achievement on *flowery* and *funny*. Looking at Fig. 5, one may have an impression that the uncertain judgments of patterns #14 and #18 on correspondingly selected features are somewhat similar. More concretely, pattern #18 has the third highest aggregated value which is the average of its two lowest degrees of target achievement also for *funny* and *attractive* like the case of pattern #14 as shown in Table 4.

Finally, if quantifier '*all*' is used, the aggregation operator $\mathcal{F}$ is a pure 'AND' operator, i.e. $andness(\mathcal{F}) = 1$. In this case, we see that pattern #10 disappears from the top three recommended items due to the target achievement for *attractive*, while pattern #30 becomes the third recommended item, after patterns #14 and #18, as having the third highest aggregated value.

5 Conclusion

In this paper, firstly we have formulated the evaluation problem of Japanese traditional crafts in which product items are essentially evaluated according kansei features reflecting esteem and aesthetic aspects of human beings perception on them. Practically, decisions on which traditional items to purchase or use are heavily influenced by personal feelings/characteristics, we have proposed a consumer-oriented evaluation model targeting on recommendation requests specified by consumers' feeling preferences. Particularly, the proposed evaluation model aims at providing a recommendation to a particular consumer that which product items would best meet her feeling preference predefined. An application for Kutani porcelain has been provided to illustrate how the proposed evaluation model works.

Acknowledgements

The work on this paper was supported by JSPS Grant-in-Aid for Scientific Research [KAKENHI (C) #19500174].

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)
2. Bordley, R.: Foundations of target-based decision theory. In: Brachinger, H.W., Monney, P.A. (eds.) Decision Analysis. From: Encyclopedia of Life Support Systems (EOLSS), Eolss Publishers, Oxford (2002), http://www.eolss.net
3. Childs, T., et al.: Affective Design (Kansei Engineering) in Japan. Faraday Packaging Partnership, University of Leeds, UK (2001)

4. Grimsæth, K.: Kansei Engineering: Linking Emotions and Product Features. Norwegian University of Science and Technology, Report (2005)
5. Herrera, F., Herrera-Viedma, E., Verdegay, J.L.: A linguistic decision process in group decision making. Group Decision Negotiation 5, 165–176 (1996)
6. Huynh, V.N., Nakamori, Y.: A satisfactory-oriented approach to multi-expert decision-making under linguistic assessments. IEEE Trans. Systems, Man, and Cybernetics–Part B SMC–35(2), 184–196 (2005)
7. Huynh, V.N., et al.: A fuzzy target based model for decision making under uncertainty. In: Proceedings of IEEE International Conference on Fuzzy Systems, Vancouver, Canada, pp. 1611–1618. IEEE, Los Alamitos (2006)
8. Jindo, T., Hirasago, K.: Application studies to car interior of kansei engineering. International Journal of Industrial Ergonomic 19, 105–114 (1997)
9. Lai, H.-H., Chang, Y.-M., Chang, H.-C.: A robust design approach for enhancing the feeling quality of a product: A car profile case study. International Journal of Industrial Ergonomic 35, 445–460 (2005)
10. Mondragón, S., Company, P., Vergara, M.: Semantic differential applied to the evaluation of machine tool design. International Journal of Industrial Ergonomicsc 35, 1021–1029 (2005)
11. Nagamachi, M.: Kansei engineering: a new ergonomic consumer-oriented technology for product development. International Journal of Industrial Ergonomics 15, 3–11 (1995)
12. Nakada, K.: Kansei engineering research on the design of construction machinery. International Journal of Industrial Ergonomics 19, 129–146 (1997)
13. Nakamori, Y., Ryoke, M.: Treating fuzziness in subjective evaluation data. Information Sciences 176, 3610–3644 (2006)
14. Osgood, C.E., Suci, G.J., Tannenbaum, P.H.: The Measurement of Meaning. University of Illinois Press, Urbana, USA (1957)
15. Petiot, J.-F., Yannou, B.: Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics. International Journal of Industrial Ergonomics 33, 507–525 (2004)
16. Ryoke, M., Nagase, K., Nakamori, Y.: Quantification of relationships between evaluation targets and adjectives through fuzzy correspondence analysis and OWA operator. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montréal, Canada, pp. 656–660 (2007)
17. Ryoke, M., Nakamori, Y., Huynh, V.N.: Personalized recommendation for traditional crafts using fuzzy correspondence analysis with kansei data and OWA operator (in these proceedings)
18. Schütte, S.: Engineering emotional values in product design – kansei engineering in development. Institute of Technology, Linköpings Universitet, Dissertation, 951, (2005)
19. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans. Systems, Man, and Cybernetics SMC–18, 183–190 (1988)
20. Yager, R.R.: On the instantiation of possibility distributions. Fuzzy Sets and Systems 128(2), 261–266 (2002)
21. Zadeh, L.A.: Fuzzy sets as a basic for a theory of possibility. Fuzzy Sets and Systems 1, 3–28 (1978)
22. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Computers and Mathematics with Applications 9, 149–184 (1983)

Using Interval Function Approximation to Estimate Uncertainty

Chenyi Hu

Computer Science Department
University of Central Arkansas
Conway, AR 72035, USA
CHu@uca.edu

Summary. Uncertainties in the real world often appear as variabilities of observed data under similar conditions. In this paper, we use interval functions to model uncertainty and function volatility. To estimate such kinds of functions, we propose a practical interval function approximation algorithm. Applying this algorithm, we have studied stock market forecasting with real economic data from 1930-2004. The computational results indicate that interval function approximation can produce better quality forecasts than that obtained with other methods[1].

1 Introduction

1.1 Interval Function

Functions have been among the most studied topics in mathematics and applications. Provided in analytical form, a function can be easily examined for its properties. However, in real world applications, the analytical form of a function is often unknown. To discover a function that properly models an application is a major challenge. Hence, computational methods on interpolation and approximation are often applied in estimating a function. Also, the main objective of studying differential and integral equations is to search for the unknown function that satisfies given conditions either theoretically or computationally.

Real world observations often differ from from the exact mathematical definition of a function. Even for a fixed x, the observed values of y may be different from time to time. These kinds of uncertainties are traditionally considered as effects of random noise and are modeled with probability theory. These variations of the value of a function f, for a given x, are often within a finite interval rather than completely random. Also, due to imprecise measurement and control, the value of x can be within an interval rather than an exact point. This means that an observed data pair can be represented as an interval valued pair $(\mathbf{x}, \mathbf{y})$ rather than precise point (x, y).

[1] The computational results have been published recently [12]. This paper is a generalized abstraction.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 341–352, 2008.
springerlink.com

Definition 1. Let f be a mapping from $\Re^n \to \Re$ and $\mathbf{x}$ be an interval vector in $\Re^n$ (i. e. each component of $\mathbf{x}$ is an interval in $\Re$). If for any interval vector $\mathbf{x}$ there is an interval $\mathbf{y}$ such that $f(\mathbf{x}) = \mathbf{y}$, then f is an interval function.

We use Figure 1 to illustrate a volatile function that can be better modeled with a interval function. In observing the function values in the figure repeatedly, due to imprecise measurement and control, one may obtain different values of y even for a "fixed" x. More importantly, even one can control x precisely and get the exact y, the point data pairs (x, y) can be misleading when use them in classical function interpolation and approximation. Therefore, an observation recorded as $(\mathbf{x}, \mathbf{y})$ should be more appropriate.

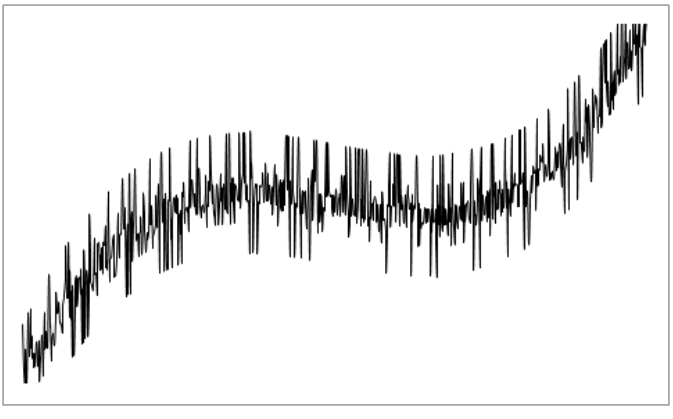


Fig. 1. A volatile function

We say that a real valued function f is 'volatile' in a domain D, if within any small subset of D the sign of the derivatives of f alternates frequently. Figure 1 presents a 'volatile' function. Real world examples of volatile functions include stock prices during any volatile trading day and recorded seismic wave. As shown in Figure 1, for a volatile function, it would be more appropriate to use an interval valued pair to record an observation.

1.2 The Objective of This Paper

Using observed discrete data pairs (x, y) to computationally approximate an unknown function has been intensively studied. Numerical polynomial interpolation and the least squares approximation are the classical methods in scientific computing. In this paper, we view uncertainty as function volatility modeled with interval valued function. Our objective is to establish a general algorithm that can approximate an unknown interval function. In other words, we try to estimate an interval function from a collection of interval valued pairs $(\mathbf{x}, \mathbf{y})$. Algorithms on interpolating interval functions have been discussed in [13]. In

this paper, we focus on approximation, specifically, on the least squares approximation, since it is probably the most broadly used computational method in function approximation.

The rest of this paper is organized as the follow. Section 2 reviews briefly the classical least-squares approximation. Section 3 presents an algorithm for interval function least squares approximation. Section 4 discusses assessment indicators. Section 5 presents a case study. Section 6 concludes the paper.

2 Least Squares Approximation

In this section, we briefly review the principle and computational methods of the ordinary least squares approximation.

2.1 Basis of a Function Space

Let us start with basic concepts related to a function space first.

Definition 2. Let F be a function space and $\Phi = \{\phi_0, \phi_1, \cdots, \phi_n, \cdots\}$ be a set of functions in F. We say that Φ is a basis of F if for any function $f \in F$ and any given $\epsilon > 0$ there is a linear combination of ϕ, $f = \sum_j \alpha_j \phi_j$, such that $|f(x) - \sum_j \alpha_j \phi_j| < \epsilon$ for all x in the domain.

For example, the set $\{1, x, x^2, \cdots\}$ is a basis of polynomial function space as well as a basis of a function space that consists of all continuous functions. Of course, there are other bases for a function space. For example, Chebychev polynomials, Legendre polynomials, sine/cosine functions, and others are commonly used as bases in approximating continuous functions.

2.2 The Least Squares Principle

For a continuous function f (even with countable discontinuities), we may approximate it as $f(x) \approx \sum_{0 \le j \le m} \alpha_j \phi_j(x)$, where $\phi_j(x)$ is a preselected set of m basis functions. To determine the coefficient vector $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_m)^T$, the least squares principle requires that the integral of the squares of the deferences between $f(x)$ and $\sum_{0 \le j \le m} \alpha_j \phi_j(x)$ is minimized. In other words, applying the least-squares principle in approximating a function f, one selects the vector $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_m)^T$ that minimizes $\int \left(f(x) - \sum_{0 \le j \le m} \alpha_j \phi_j(x) \right)^2 dx$.

2.3 Discrete Algorithm

In real world applications, one usually only knows a collection of N pairs of (x_i, y_i) rather than the function $y = f(x)$. Therefore, one minimizes the total

sum $\sum_{i=1}^{N}\left(y_i - \sum_{0\le j\le m}\alpha_j\phi_j(x_i)\right)^2 dx$ instead. The classical algorithm that computationally determines the coefficient vector α is as the follow:

Algorithm 1

(i) Evaluate the basis functions $\phi_j(x)$ at x_i for all $1 \le i \le N$ and $1 \le j \le m$;
(ii) Form the matrix

$$A = \begin{pmatrix} N & \Sigma_i\phi_1 & \Sigma_i\phi_2 & \cdots & \Sigma_i\phi_m \\ \Sigma_i\phi_1 & \Sigma_i\phi_1^2 & \Sigma_i\phi_1\phi_2 & \cdots & \Sigma_i\phi_1\phi_m \\ \Sigma_i\phi_2 & \Sigma_i\phi_2\phi_1 & \Sigma_i\phi_2^2 & \cdots & \Sigma_i\phi_2\phi_m \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma_i\phi_m & \Sigma_i\phi_m\phi_1 & \Sigma_i\phi_m\phi_2 & \cdots & \Sigma_i\phi_m^2 \end{pmatrix} \tag{1}$$

and the vector

$$b = (\Sigma_i y_i \;\; \Sigma_i y_i\phi_1(x_i) \;\; \Sigma_i y_i\phi_2(x_i) \;\; \cdots \Sigma_i y_i\phi_m(x_i))^T; \tag{2}$$

(iii) Solve the linear system of equations $A\alpha = b$ for α.

The linear system of equations $A\alpha = b$ above is called the normal equation. In stead of normal equations, a more current approach applies a design matrix with a sequence of Householder transformations to estimate the vector α. For details about Householder transformations, it is out of the main scope of this paper, readers may refer [16] or most books that cover computational linear algebra. Although the basic idea of this paper can be applied to both approaches, we use the normal equation approach in the rest of is paper for its simplicity.

2.4 Time Series and Slicing-Window

We now switch our attention to the dataset. In real world applications, an observed data pair (x_i, y_i) is often associated with a specific time. The collection of data pairs, if ordered chronically, is called a time series. Time series have been extensively studied for prediction and forecasting [5] and [7]. Rules and functions often rely on a specific time period. We call it time-varying, that is, the relationship is valid for a limited time period. Therefore, in applying function approximation on a time series, one should use only data inside an appropriate time-window to estimate the relationship. By slicing the time-window (also called rolling), one obtains a sequence of function approximations such that each of them valid only for a specific time-window.

3 Interval Function Approximation

Previous studies on least squares approximation mostly assume, if not all, point valued data. There are several computational issues that need to be considered in order to apply Algorithm 1 on interval valued pairs $(\mathbf{x}, \mathbf{y})$ to approximate an interval function.

3.1 Computational Challenges

With interval arithmetic [17], it is straightforward to perform both steps 1 and 2 in Algorithms 1. However, it presents a challenge in the step 3. This is because the normal equations are now interval systems of linear equations $\mathbf{A}\alpha = \mathbf{b}$. The solution set of an interval linear system of equations is mostly irregular shaped and non-convex [18]. A naive application of interval arithmetic to bound the solution vector α may cause serious overestimation due to the wrap effects, and then negatively affect the approximation quality. Using the design matrix approach would not solve the problem since finding a Householder transformation for an interval matrix remains a challenge.

3.2 An Inner Approximation Approach

While an interval $\mathbf{x}$ is usually presented by its lower and upper bounds as $\mathbf{x} = [\underline{x}, \overline{x}]$, it can also be represented by its midpoint $\text{mid}(\mathbf{x}) = \frac{\underline{x} + \overline{x}}{2}$ and its width $w(\mathbf{x}) = \overline{x} - \underline{x}$. This creates a two-step approach where we consider the midpoint and width separately in each of the two steps.

Instead of finding the lower and upper bounds of the interval vector α in the step 3 of Algorithm 1, let us first try to find its midpoint vector, which is a scalar vector. This suggests us to match the center of two interval vectors $\mathbf{A}\alpha$ and $\mathbf{b}$ in the interval linear system of equations $\mathbf{A}\alpha = \mathbf{b}$. Let A_{mid} be the midpoint matrix of $\mathbf{A}$, and b_{mid} be the midpoint vector of $\mathbf{b}$. We solve the non-interval linear system of equations $A_{mid}\alpha = b_{mid}$ for α.

We would like to emphasize that the result of $\mathbf{y} = f(\mathbf{x}) \approx \alpha_0 + \sum_{1 \le j \le m} \alpha_j \phi_j(\mathbf{x})$ is an interval even when we use the midpoint of the interval vector α in the calculation. This is because of that the independent variable $\mathbf{x}$ is interval valued. However, by collapsing an interval vector α to its midpoint, we could reasonably expect that the approximation is an *inner interval approximation*.

3.3 Width Adjustment

Now, let us consider the width. One may try to look for the width vector of α vector or for the width of $\mathbf{y}$. There can be different computational heuristics too. For example, one may select a width vector that makes $\mathbf{A}\alpha$ as close as possible to $\mathbf{b}$ in the step 3 of Algorithm 1. One may also use widths to perform least-square approximation to estimate the width of $\mathbf{y}$. Another computational heuristic is to adjust the width by multiplying a scale factor to the inner approximation. In our case study, we adopted the later approach. We believe that there are still many open questions for further study on width adjustment.

3.4 Interval Least-Squares Approximation

By summarizing the above discussions, we revise Algorithm 1 to Algorithm 2 below for interval function least squares approximation.

Algorithm 2

(i) Input available interval data pairs $(\mathbf{x}_i, \mathbf{y}_i)$ for $1 \leq i \leq N$;
(ii) Evaluate matrix A and vector b with interval arithmetic;
(iii) Find A_{mid} and b_{mid}, the midpoint matrix of $\mathbf{A}$ and the midpoint vector of $\mathbf{b}$, respectively;
(iv) Solve the linear systems of equations: $A_{mid}\alpha = b_{mid}$;
(v) Apply the vector α to calculate an inner approximation with interval arithmetic;
(vi) Modify the initial approximation with a width adjustment.

3.5 Other Approaches to Obtain an Interval Approximation

One may obtain an interval approximation without using interval arithmetic at all. The lower and upper bounds of interval data pairs $(\mathbf{x}_i, \mathbf{y}_i)$ form two collections of point data $(\underline{x}_i, \underline{y}_i)$ and $(\overline{x}_i, \overline{y}_i)$. By applying point least square approximation to them separately, one can obtain two point estimations. These two estimations can form an interval estimation. We call this approach the *min-max interval approximation*. This has been reported and applied in [10] and [11].

Another way to obtain an interval approximation is to apply classical statistic/probabilistic approach. By adding to and subtracting from a point approximation a certain percentage of standard deviations, one can obtain forecasting intervals. In the literature, this is called a *confidence interval*. However, the case study in Section 5 of this paper implies that, at least in certain cases, interval function lease squares approximation may produce better computational results than that obtained with the min-max interval and confidence interval.

4 Assessing Interval Function Approximation

There are different ways to produce an interval approximation. A immediate question is how to assess the quality of different interval estimations. We define two measurements for quality assessment of an interval approximation.

Definition 3. Let $\mathbf{y}_{est}$ be an approximation for the interval $\mathbf{y}$. The absolute error of the approximation is the absolute sum of the lower and upper bounds errors, i.e. $|\underline{y}_{est} - \underline{y}| + |\overline{y}_{est} - \overline{y}|$.

Example 1. If one obtained $[-1.02, 1.95]$ as an approximation for the interval $[-1.0, 2.0]$, then the absolute error of the estimation is $|(-1.02)-(-1.0)|+|1.95-2.0| = 0.02 + 0.05 = 0.07$.

Since both $\mathbf{y}_{est}$ and $\mathbf{y}$ are intervals, an additional meaningful quality measurement can be defined. The larger the overlap between the two intervals the better the approximation should be. By the same token, the less the non-overlap between the two intervals the more accurate the forecast is. In addition, the accuracy of an interval estimation should be between 0% and 100%. By using the notion of interval width, which is the difference of the upper and lower bounds

of an interval, we can measure the intersection and the union (or the convex hull) of the two intervals. Let $w(\ )$ be the function that returns the width of an interval. Then, we define the concept, named the *accuracy ratio of an interval approximation*, as the follow.

Definition 4. Let $\mathbf{y}_{est}$ be an approximation for the interval $\mathbf{y}$. The accuracy ratio of the approximation is $\frac{w(\mathbf{y} \cap \mathbf{y}_{est})}{w(\mathbf{y} \cup \mathbf{y}_{est})}$ if $(\mathbf{y} \cap \mathbf{y}_{est}) \neq \emptyset$. Otherwise, the accuracy ratio is zero.

Example 2. Using $[-1.02, 1.95]$ to approximate the interval $[-1.0, 2.0]$, the accuracy ratio is $\frac{w([-1.02, 1.95] \cap [-1.0, 2.0])}{w([-1.02, 1.95] \cup [-1.0, 2.0])} = \frac{w([-1.0, 1.95])}{w([-1.02, 2.0])} = \frac{2.95}{3.02} = 97.68\%$.

As in classical statistics, for a collection of interval estimations, one can calculate the mean and standard deviation of the absolute error and the accuracy ratio. Furthermore, one may also apply probability theory to perform comparisons of different approximations.

5 Case Study: Forecasting the S & P 500 Index

The S & P 500 index is a broadly used indicator for the overall stock market. Using interval least squares approximation, we have performed S & P 500 annual forecast with astonishing computational results[9] and [12]. We report it here again as a case study with comparisons against the result obtained with traditional ordinary least squares forecasting.

5.1 The Model

Driven by macroeconomic and social factors, the stock market usually varies with time. The main challenge in studying the stock market is its volatility and uncertainty. The arbitrage pricing theory (APT) [20] provides a framework that identifies macroeconomic variables that significantly and systematically influence stock prices. By modeling the relationship between the stock market and relevant macroeconomic variables, one may try to forecast the overall level of the stock market.

The model we use in this case study is a broadly accepted one by Chen, Roll and Ross (1986). According to their model, the changes in the overall stock market value (SP_t) are linearly determined by the following five macroeconomic factors: the growth rate variations of seasonally-adjusted Industrial Production Index (IP_t); changes in expected inflation (DI_t) and unexpected inflation (UI_t); default risk premiums (DF_t); and unexpected changes in interest rates (TM_t). This relationship can be expressed as

$$SP_t = a_t + I_t(IP_t) + U_t(UI_t) + D_t(DI_t) + F_t(DF_t) + T_t(TM_t)$$

By using historic data, one may estimate the coefficients of the above equation to forecast changes of the overall stock market. There is a general consensus in

the financial literature, that relationships between financial market and macroeconomic variables are time-varying. Hence the coefficients are associated with a time-window.

In the literature, it is called an *in-sample forecast* if using the obtained coefficients in a time-window and the equation above to calculate the SP for the last time period in the time-window. It is called an *out-of-sample forecast* if using the obtained coefficients in a time-window to calculate the SP for the first time period that immediately follows the time-window [4]. By slicing the time-window (also called rolling), one obtains a sequence of coefficients and forecasted SP values. The overall quality of forecasting can be measured by comparing the forecasts against actual SP values. In practice, the out-of-sample-forecast is more useful than in-sample-forecast because it can make predictions.

5.2 The Data

So far the primary measurements used in economics and finance are quantified points. For instance, a monthly closing value of an index is used to represent the index for that month even though the index actually varies during that month. The available data in this case study are monthly data from January 1930 to December 2004. We list a portion of the data here.

Date	UI	DI	SP	IP	DF	TM
30-Jan	-0.00897673	0	0.014382062	-0.003860512	0.0116	-0.0094
30-Feb	-0.00671673	-0.0023	0.060760088	-0.015592832	-0.0057	0.0115
30-Mar	-0.00834673	0.0016	0.037017628	-0.00788855	0.0055	0.0053
30-Apr	0.00295327	0.0005	0.061557893	-0.015966279	0.01	-0.0051
30-May	-0.00744673	-0.0014	-0.061557893	-0.028707502	-0.0082	0.0118
30-Jun	-0.00797673	0.0005	-0.106567965	-0.046763234	0.0059	0.0025
.....	.....	...	...	...	...	...
04-Jun	0.00312327	-0.0002	0.026818986	0.005903385	-0.0028	0.0115
04-Jul	-0.00182673	0.0002	-0.024043354	0.00306212	0.0029	0.0147
04-Aug	0.00008127	0.0002	-0.015411102	-0.002424198	0	0.0385
04-Sep	0.00156327	0.0001	0.026033651	0.007217235	0.0005	0.0085
04-Oct	0.00470327	0	0.000368476	0.002001341	0.001	0.0143
04-Nov	-0.00002273	0	0.044493038	0.006654848	0.0034	-0.0245
04-Dec	-0.00461673	0.0004	0.025567309	0.001918659	0.0007	0.0235

In this case study, we use a time-window of ten years to obtain the out-of-sample annual forecasts for 1940-2004.

5.3 Interval Rolling Least Squares Forecasts

To perform interval rolling least squares forecasts, we need interval input data. From the provided monthly data, for each of the attributes, we choose its annual minimum and maximum to form the interval input data. By applying Algorithm 2, we obtain initial forecasts first. For each of them, we then adjust

the width of the predicted S & P 500 interval to the average of that of those within the time-window. The program was written in C++. The software package IntBLAS [19] was applied for interval and related linear algebra operations. Figure 2 illustrates the out-of-sample annual interval forecasts.

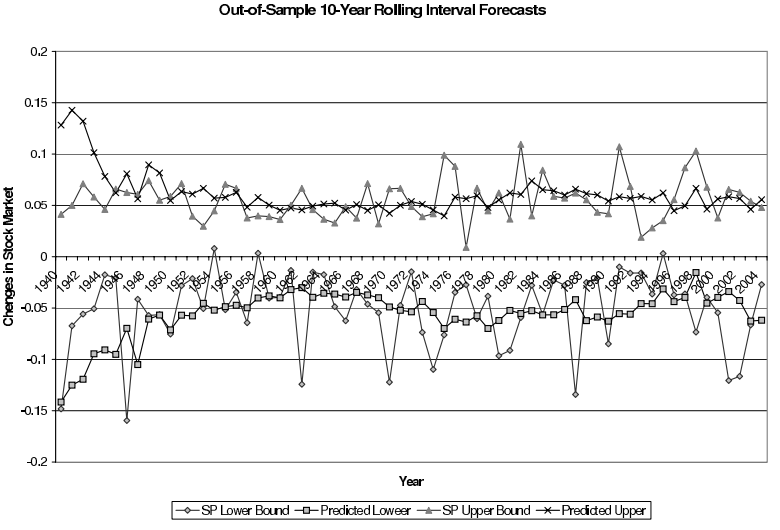


Fig. 2. Out-of-Sample Annual Interval Forecasts(1940-2004)

For the purpose of quality comparison, we calculated the annual point forecasts that are commonly used in financial study. We obtained the out-of-sample annual forecasts (in percent) for a period of 1940-2004. The out-of-sample annual point forecasts have an average absolute forecasting error of 20.6% with a standard deviation of 0.19. By adding to and subtracting from the point-forecasts with a proportion of the standard deviation, we may form confidence interval forecasts with 95% statistical confidences.

It is worth pointing out that the ranges of Figure 2 are significantly less than that of Figure 3 at the ratio only about 14%.

5.4 Quality Comparisons

To assess the quality of the above forecasts, we use the following indicators: (1) the average absolute forecast error, (2) the standard deviation of forecast errors, (3) the average accuracy ratio, and (4) the number of forecasts with 0% accuracy. We summarize the statistics of the quality indicators in the table below.

All measured indicators for forecasting quality in the table suggest that interval OLS significantly outperform point-based forecasts with a much less mean forecast error. The much smaller standard deviations produced by the interval approaches indicate that the interval forecasting is more stable than other

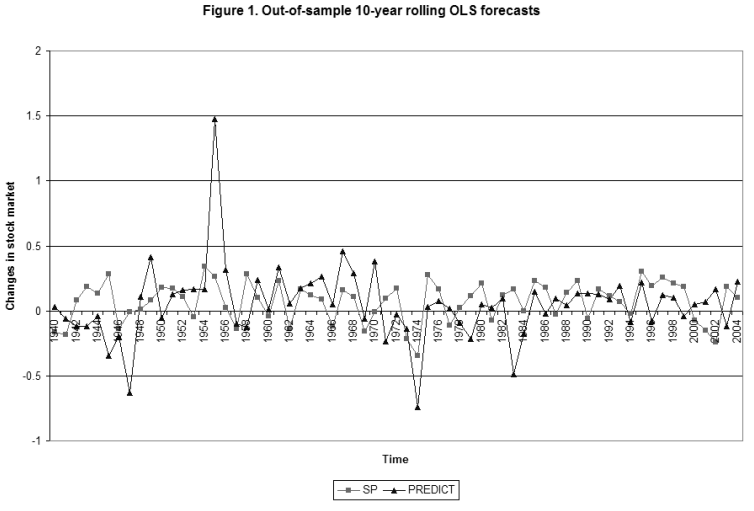


Fig. 3. Out-of-Sample Annual Point Forecasts (1940-2004)

Table 1. Quality comparison of annual forecasts (1940-2004)

Methods/Item	Absolute mean error	standard deviation	Accuracy ratio	Number of 0 accuracy
OLS	0.20572	0.18996	NA	NA
std dev. 95% confidence	0.723505	0.311973	0.125745	5
Min-Max interval	0.066643	0.040998	0.4617	0
Initial interval Fcast	0.073038	0.038151	0.385531	0
Interval Fcast	0.0516624	0.032238	0.641877	0

comparing methods. Compared with the point-based confidence interval forecasting, interval methods produce a much higher average accuracy ratio. The interval scheme with width adjustments further improves the overall forecasting quality of initial approximations in terms of the higher average accuracy ratio. All forecasts with interval computing have a positive accuracy ratio while a number of the point-based confidence intervals has zero accuracy.

6 Conclusion

In this paper, we model uncertainty as volatilities of a function. It is more reasonable to record volatile data as interval valued nodes of an interval function rather than point values. To apply classical least squares approximation with discrete interval valued nodes, we use interval arithmetic to obtain the normal equation. By using the midpoint approach, we calculate an inner approximation initially. Then, we adjust its width with computational heuristics. Although the

function to be approximated is unknown, we can still assess the quality of an interval approximation statistically with provided data. These quality indicators include absolute error, accuracy ratio, and their means and standard deviations.

Using this approach, in our case study, we performed annual forecasts for the S & P 500 index from 1940-2004 with real economical data. Although it is merely one of the initial attempts to use interval methods in financial forecasting, the empirical results provide astonishing evidence that interval least squares approximation may outperform traditional point approaches in terms of the overall less mean error and higher average accuracy ratio. Hence, interval methods have a great potential in dealing with uncertainty.

Acknowledgments

This research is partially supported by the US National Science Foundation grants CISE/CCF-0202042 and CISE/CCF-0727798.

The financial data was provided by Professor Ling T. He in Finance and Economics at the University of Central Arkansas. He initiated and actively participated in the case study.

Mr. Michael Nooner and anonymous referees have provided helpful comments for improving the quality of this paper.

References

1. Chatfield, C.: Calculating interval forecasts. Journal of Business and Economic Statistics 11, 121–144 (1993)
2. Chatfield, C.: Prediction intervals for time-series forecasting. In: Armstrong, J.S. (ed.) Principles of forecasting: Handbook for researchers and practitioners, Kluwer Academic Publisher, Boston (2001)
3. Chen, N., Roll, R., Ross, S.: Economic forces and the stock market. Journal of Business 59, 383–403 (1986)
4. Fama, E., French, K.: Industry costs of equity. Journal of Financial Economics 43, 153–193 (1997)
5. Gardner, E.: A simple method of computing prediction intervals for time series forecasts. Journal of Management Science 34, 541–546 (1988)
6. Granger, C.: Can we improve the perceived quality of economic forecasts? Journal of Applied Econometrics 11, 455–473 (1996)
7. Gooijer, J., Hyndman, R.: 25 years of time series forecasting. Journal of Forecasting 22, 443–473 (2006)
8. He, L.: Instability and predictability of factor betas of industrial stocks: The flexible least squares solutions. The Quarterly Review of Economics and Finance 45, 619–640 (2005)
9. He, L., Hu, C.: The stock market forecasting: an application of the interval measurement and computation. In: The 2nd Int. Conf. on Fuzzy Sets and Soft Comp. in Economics and Finance, St. Petersburg, Russia, pp. 13–22 (2006)
10. He, L., Hu, C.: Impacts of interval measurement on studies of economic variability: evidence from stock market variability forecasting. Journal of Risk Finance (to appear)

11. He, L., Hu, C.: Impacts of interval computing on stock market variability forecasting. Journal of Computational Economics (to appear)
12. Hu, C., He, L.: An application of interval methods to the stock market forecasting. Journal of Reliable Computing 13, 423–434 (2006)
13. Hu, C., et al.: An interval polynomial interpolation problem and its Lagrange solution. Journal of Reliable Computing 4(1), 27–38 (1998)
14. Hu, C., Xu, S., Yang, X.: A review on interval computation – software and applications. Journal of Computational and Numerical Analysis and Applications 1(2), 149–162 (2002)
15. Interval Computations, http://www.cs.utep.edu/interval-comp/main.html
16. Moler, C.: Numerical computing with MATLAB. SIAM (2004)
17. Moore, R.: Methods and applications of interval analysis. SIAM Studies in Applied Mathematics (1979)
18. Neumaier, A.: Interval methods for systems of equations. Cambridge University, Cambridge (1990)
19. Nooner, M., Hu, C.: A computational environment for interval matrices. In: NSF 2006 Workshop on Reliable Engineering Computing, Savannah, GA, pp. 65–74 (2006)
20. Ross, S.: The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360 (1976)

Interval Forecasting of Crude Oil Price

Shanying Xu[1], Xi Chen[2], and Ai Han[3]

[1] Chinese Academy of Sciences, Academy of Mathematics and Systems Science
xsy@iss.ac.cn
[2] Chinese Academy of Sciences, Academy of Mathematics and Systems Science
chenhbs@126.com
[3] Chinese Academy of Sciences, Academy of Mathematics and Systems Science
hanai06@mails.gucas.ac.cn

Summary. Uncertainty is the main obstacle in predicting crude oil price. Although there are various models and computational methods on crude oil price forecasting in literature, most of them do not effectively predict the variability of crude oil price due to uncertainty. Very recently, Hu and He [2] reported of using ILS (Interval Least Square) approach to forecast the stock market and obtained much better results than that obtained with traditional point methods, In this paper, we investigate if the ILS approach can forecast the relationship between commodity inventory levels and crude oil spot prices effectively. Our empirical study suggests that both the ILS method and the confidence interval method can produce comparable quality forecasts. While the computational result produced by the ILS method seems slightly worse than the 95% confidence intervals in two quality measurements, the differences are negligible. On a new forecasting quality measurement proposed in this paper, the ILS method produces results better than the 95% confidence intervals. Hence, interval method is a feasible alternative in crude oil price forecasting.

1 Introduction

1.1 Forecasting Crude Oil Price

As a strategic resource, crude oil and its trade have attracted extensive attention for a long time. Since the beginning of this century, international crude oil price has continued to rise rapidly. Therefore, the crude oil price forecasting has become a focus of economists and decision makers.

Recently, numerous studies have focused on the relationship between commodity inventory levels and spot prices. Specifically, in the crude oil market, previous work has explained a relationship between crude oil price and total inventories (crude plus products) of Organization for Economic Cooperation and Development (OECD) countries. For example, Pindyck [5] described the short-run dynamic relationship between commodity prices and inventories by using the petroleum markets for illustrations; Considine and Larson [1] studied the presence of risk premiums on crude oil inventories.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 353–363, 2008.
springerlink.com

Ye, Zyren, and Shore [6] developed a more practical simple model for short-term crude oil price forecasting. The model is easily interpreted and accepted, hence, it is intuitively appealing to decision-makers. The model also provides good in-sample and out-of-sample dynamic forecasts for the post-Gulf War time period. In addition to providing good forecasting results, a desirable feature of the model is that it can readily be implemented in a spreadsheet or other software package, with the variables easy to update.

Ye, Zyren, and Shore [7] also provided a nonlinear-term model, which improves upon the model that assumes only linear price adjustments to inventory. The model nonlinearity is based on the feature that inventory has a zero lower bound or a minimum operating requirement. Two nonlinear inventory variables are incorporated: one for the low inventory state and another for the high inventory state of the crude oil market. Compared with the linear forecast model, the nonlinear model demonstrates its superiority in both the model fitness and forecasting ability.

1.2 Interval Computing

Moore [3] proposed interval analysis in the late of 1950's. Since then, thousands of research articles and numerous books have appeared on the subject. The website of Interval Computations [8] provides comprehensive information on interval computations. An increasing amount of software resources for interval computations are available through the Internet. Interval computation, different from classical point arithmetic, has contributed a lot in many researching fields. For illustration, Hales and Ferguson solved the Kepler conjecture about the densest arrangement of spheres in space by using interval computing [8]. This work led them the 2007 Robbins Prize by the American Mathematical Society. In his response to this award, Hales explicitly thanked those who developed the tools of interval computations.

Interval computing does contribute unique merits compared with traditional point methods in the following aspects. First of all, interval computing can bound the errors of finite digit floating-point computing. Interval computations yield a pair of numbers, an upper and a lower bound, which are guaranteed to enclose the exact answer. Maybe we still don't know the truth, but at least we know how much we don't know. For example, interval arithmetic has been used in computers to cope with some of the continuous and infinite aspects of mathematics to fence the round-off errors, which might lead to inaccurate results. Secondly, interval computing plays an important role in dealing with uncertainty in computational modeling. In the real world, the value of data may not be precisely known due to certain forms of uncertainty. Hence, it is unreasonable to use point data for that may lead to inappropriate results especially when the inaccuracy is not negligible. One approach to dealing with such problem is to use intervals, which is not an approximate value of the de-sired data but is guaranteed to contain the unknown value. Thirdly, in reality, many variables are bounded by intervals for a given time. For instance, macroeconomic variables definitely vary within intervals during a given period and the point valued data does not reflect

the variability of these variables. Therefore, it makes more sense to use interval inputs instead of the point-based data for that they contain more in-formation. Last but not the least, if the predicted variable possesses the property mentioned above, it is more reasonable to provide interval forecasting outputs to decision makers than the point valued ones.

1.3 Motivation of This Work

In this paper, we propose to apply interval methods in monthly crude oil spot price forecasting for the following reasons.

First, as mentioned above, interval data contains more information than point data. In Ye et al. [7], two new variables were incorporated to capture the non-linear behavior of inventory and the forecasting ability has been improved. In other words, these new variables introduce more information to the former linear model in Ye et al [6]. In this paper, instead of adding new variables to the former model, we directly form interval input data to characterize the volatility of all the variables by which more information were included.

Secondly, variables such as crude oil spot price definitely vary frequently within an interval during a given month. Thus it is much more meaningful to supply interval valued forecasting outputs than the ordinary point valued ones.

Thirdly, Hu and He [2] have developed an interval least squares scheme and ap-plied it in forecasting the S & P 500 index. The annual interval forecasts they obtained are in much better quality than that obtained with point methods. Inspired by their work, we would like to investigate whether interval method is reasonable in forecasting other economical variables, as of crude oil spot price. To evaluate the forecasting quality, we not only apply the same measurements proposed by Hu and He, but also propose a new measurement which checks how many times that the forecasting outputs fail to encompass actual monthly average oil spot prices.

The rest of this paper is organized as follows. We introduce the model equations used in this study and review the interval least squares method (ILS) and rolling forecasting in Section 2. Then we describe the data and software for the empirical study in Section 3. We report the computational results in both numerical tables and graphical charts, and compare them with 95% confidence intervals (CI)[1] of point forecasts in Section 4. We conclude this paper with possible future work in Section 5.

2 The Model and Computational Methods

In this study, we use the oil price forecasting model established by Ye, Zyren and Shore [6]. We use this model mainly because of its broad acceptance in the literature. In this section, we briefly review this model and then introduce the interval least square method reported in [2]. Meanwhile, some computational details are provided.

[1] A confidence interval for a point forecast specifies a range of values within which the actual average monthly oil spot price may lie.

2.1 The Model

In Ye, Zyren and Shore [6], the observed level of petroleum inventory is decomposed into two components: the normal level and the relative level. The former is determined by historical seasonal movements and trends, and reflects the normal market demand and operational requirements. The later represents the difference between the observed and normal levels and reflects short-run market fluctuations. The relative inventory level (denoted by RIN) is determined as

$$RINt = INt - IN^*t; \quad IN^*t = a0 + b1T + \sum^{12} k = 2bkDk \tag{1}$$

where IN is the actual inventory level and IN^* is the normal inventory level. $Dk, k = 2, \cdots, 12$, are 11 seasonal dummy variables[2] and T is time (T begins with 1). $a0, b1$ and $bk(k = 2, 3, \cdots, 12)$ are parameters to be estimated. The crude oil price forecasting model is

$$WTIt = a + \sum_{i=0}^{3} biPINt - i + \sum_{j=0}^{5} cjDj911 + dLAPR99 + eWTIt - 1 + \varepsilon \tag{2}$$

In (2), WTI is the West Texas Intermediate crude oil spot price. Subscript t is for the t^{th} month; subscript i is for i^{th} month prior to the t^{th} month; $Dj911(j = 0, 1, \cdots, 5)$ referring to the 6 months from October 2001 to March 2002) is a set of single monthly variables[3] to account for market disequilibrium following the September 11, 2001 terrorist attacks in the United States; $LAPR99$ is a level-shifting variable[4] corresponding to the effect that OPEC quota tightening had on the petroleum market be-ginning in April 1999; a, bi, cj, d and e are coefficients to be estimated.

Relationships between WTI and the explanatory variables are time-varying. That is to say, the relationship may be valid for a limited time period. The relationships and the forecasts are associated with a time-window. By rolling the time window, one obtains a series of forecasts. The overall forecasting quality can be measured by the average performance of the series. We follow the method of rolling forecasting used by Hu and He [2]. In the paper, it is called an in-sample forecast if using the obtained coefficients in a time window to calculate the WTI for the last period in the time window. It is called an out-of-sample forecast if using the obtained coefficients in a time window to calculate the WTI for the first time period that immediately follows the time window.

2.2 Interval Least Square Method

As many economic variables are time-varying, it is more meaningful to analyze interval data rather than point data. This urges the emergence of interval

[2] Dummy variables are binary variables. For example, when it is February, $D2$ equals 1 and other $Dk(k = 3, 4, \cdots, 12)$ equals 0.

[3] $Dj911$ equals 1 at its corresponding month and equals 0 at all the other months.

[4] $LAPR99$ is also a dummy variable. It equals 0 before April 1999 and equals 1 after the month (including April 1999).

econometrics. This cross field is still at its infancy, facing computational difficulties and lack of theoretical basis. As for the simple linear regression for interval data, Hu and He [2] propose a computational scheme. We summarize this scheme here as Interval Least Square method.

In contrast to OLS, the inputs and outputs in ILS are intervals rather than points, though at this stage it still relies on OLS to estimate coefficients of the regression. The computational scheme is as follows:

(i) Read available interval data.
(ii) Find the midpoint of the intervals.
(iii) Solve the linear system of equations: $A\text{mid}x = b\text{mid}$ for x
(iv) Apply the obtained coefficient vector x to calculate the initial prediction with interval data using interval arithmetic
(v) Perform width adjustment for the initial prediction to make a forecast.

3 Data and Software for the Empirical Study

The major difference between this study and the traditional ones is the data type. In this section, we present the details about the data source, data preprocessing, and software used in our empirical study.

3.1 Data Source

The data we used in this study are monthly average West Texas Intermediate crude spot prices and the monthly OECD total inventory levels from the website of Energy Information Association (EIA). Average is commonly used in forecasting monthly oil price. Ye et al. [6, 7] use average as the dependent variable. The monthly data released by EIA official website is also the average of daily prices in a single month. Thus we use the average as the dependent variable and predictor in point model as well as in one of our error measurements which will be explained later. We substitute the OECD total inventory level for the OECD commercial inventory level used in [6] mainly because of that the OECD commercial inventory level is unavailable to access. The WTI spot price is in nominal dollars per barrel and the inventory level is measured in million barrels.

In [6, 7], the authors limit their study to the period from Jan 1992 to April 2003. We do of the same in this study. In [6], the process of generating out-of-sample forecasts begins by fitting the models for the Jan 1992 to Dec 1999 time period. We use the same length of this period, 96 months, as that of the time window for our rolling forecasting. As noticed, the rolling forecasting is interrupted by the event of September 11, 2001. The crude oil spot price fell and rose dramatically after this event. Therefore, six dummy variables are placed in the model. Consequently, because of the dummy variables, the model cannot produce appropriate forecasts for the six months after the event. Thus, the six months are not forecasted. The period of in-sample forecasts is from Dec 1999 to April 2003, and that of out-of-sample forecasts is from Jan 2000 to April 2003, excluding the period of Oct 2001 to March 2002.

3.2 Data Preprocessing

The monthly spot price interval is formed by the minimum and the maximum daily price in the month. The relative inventory level is calculated by using formula (1). Given the common sense that the inventories vary during a month, we transform the relative level into interval data by adding and subtracting a certain percentage. The percentage is determined by repeatedly experiments. In this study, as the percentage grows, the average Absolute Error and the average Accuracy Ratio also grow. We find that around 40%, the 95% confidence interval forecasts and the ILS interval forecasts are not so different to each other.

The percentage seems high, but that doesn't necessarily mean that the fluctuations are fierce, because it is the percentage of the "relative" inventory level rather than the observed inventory level. A high percentage against the "relative" level may only count for a small percentage against the whole inventory. Using the method of adding and subtracting a percentage of the relative level suggests that the widths of the intervals may vary significantly from each other. The interval is going to be large when the inventory is far away from the normal level, while to be small when the inventory is near the normal level. Thus, the widths are related to the fluctuations of the inventory level. Large input intervals are likely to produce large forecast intervals. When relative inventory level varies around zero, it is not likely to get large forecast intervals, that is, fierce fluctuations. In contrast, when the inventory level is far away from normal level, the forecast will exhibit bigger fluctuations and become more sensitive. This is consistent with common sense, when larger interval is interpreted as bigger fluctuations. Therefore, comparing with intervals with constant width, intervals formed in this way will cast non-linear effect on the widths of the price intervals.

3.3 Software

The software package we used in this study is developed and provided by Professor Chenyi Hu who visited the Academy in summer 2007. The package includes a col-lection of interval linear algebra subroutines (Nooner and Hu [4]) and an application program for interval least squares rolling forecasting. The package is written in C++.

4 Computational Results and Comparisons

Using the method discussed in Section 2 and the data discussed in Section 3, we obtain the empirical results that are presented in this section. Firstly, measurements to access the quality of forecasts are discussed, followed by comparisons of the empirical results that are showed in tables and diagrams.

4.1 The Error Measurements

Two error measurements are used to access the quality of the forecasts. One is the Absolute Error, the sum of absolute forecast errors of lower bound and

upper bound forecasts; the other is the Accuracy Ratio, the percentage of the intersection of the forecast interval and the actual interval against the union of the two intervals. The Accuracy Ratio should be between 0% and 100%. For more details, please refer to Hu and He [2].

4.2 Comparison with Actual Monthly Price Interval

Using model (2) and ILS method described above, we obtain the in-sample and the out-of-sample monthly interval forecasts. The forecast quality measurements are shown is Table 1.

Table 1. The forecast quality of ILS method; percentage: 40%

	in-sample	out-of-sample
Average Absolute Error	3.147	3.511
Average Accuracy Ratio	51.932	48.557

In order to compare them with the results of point forecasting, we also calculate the 95% confidence intervals of the point forecasts (using the same model and the same forecasting period). The forecast quality of confidence intervals is summarized in Table 2.

Table 2. The forecast quality of 95% confidence intervals

	in-sample	out-of-sample
Average Absolute Error	3.163	3.482
Average Accuracy Ratio	53.239	50.241

A comparison of the in-sample interval forecasts (ILS method) and the actual intervals are shown in Figure 1, and that of the out-of-sample (ILS method) is shown in Figure 2.

The two tables above show that the forecast capability of confidence interval method and ILS method are comparable. Confidence interval method is slightly better than ILS method. However, the difference is not huge. Although the new method does not outperform the tradition one as used in forecasting annual return of S&P index, it is comparable to the traditional one. Thus, it is reasonable to use interval method in this case as well.

Aside from forecast quality, direct interval forecast is more meaningful. Unlike confidence interval, which stands for the possible interval of the mean, interval out-put from ILS method stands for the lower and upper boundaries of the price. The meaning of interval output is clearer and more direct.

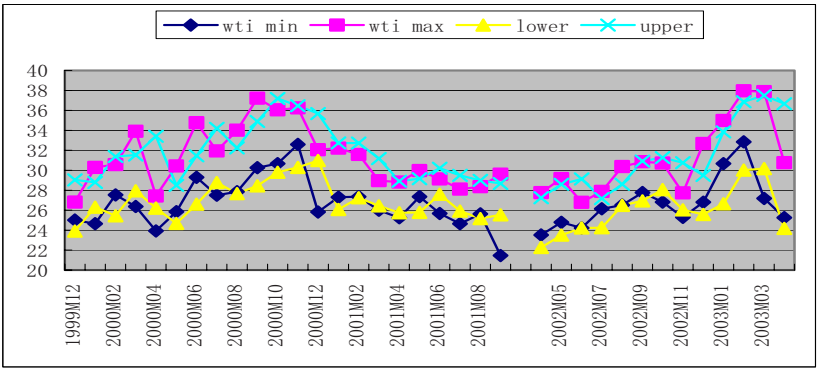


Fig. 1. In-sample forecasts comparing with the actual WTI prices

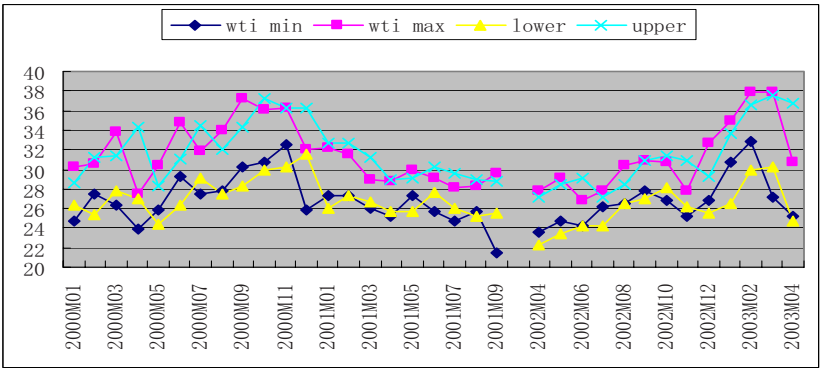


Fig. 2. Out-of-sample forecasts comparing with the actual WTI prices

4.3 Comparison with Actual Monthly Average Price

The following tables show the forecasting quality of the two methods CI and ILS by comparison them with the actual monthly average price of WTI crude oil.

Considering that decision makers who are used to interpret CI results may want to know if the intervals encompass the actual monthly average prices, especially in the case of out-of-sample forecast, we count the numbers of intervals which do not include the actual monthly average price of both methods. The results are represented in Table 3, and Figure 3 and Figure 4 are the related graphs. As is shown in Table 3, the number of fails of ILS method is smaller than that of CI method. The possibility of failure of ILS is 3% lower than that of CI method. That is, in this case, ILS method is slightly better than traditional CI method.

As is shown in Figures 3 and 4, forecasts of the two methods are not so different from each other, because they are based on the same model. Although they use

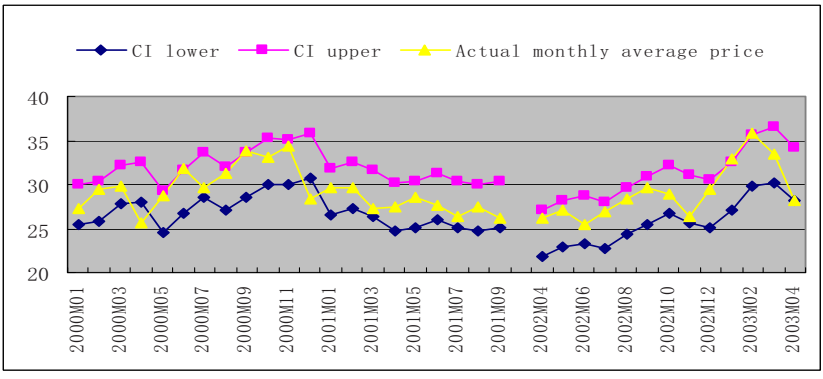


Fig. 3. Out-of-sample forecasts of CI method and monthly average oil spot price

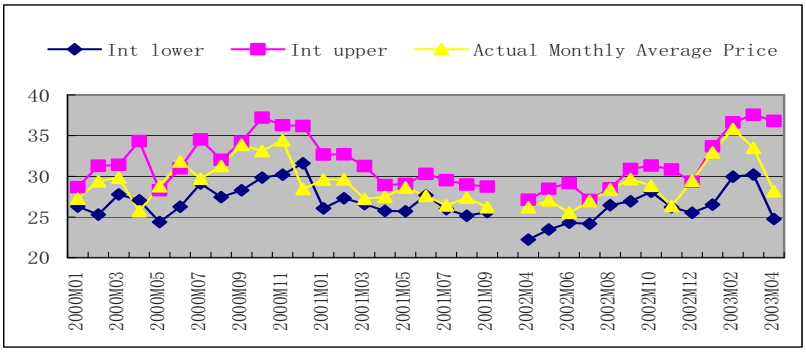


Fig. 4. Out-of-sample forecasts of ILS method and monthly average oil spot price

Table 3. Statistics of out-of-sample forecasts, which fail to encompass actual monthly average oil spot prices, of ILS method and CI method

	Number of fails	Prob.
ILS	6	17.65%
CI	7	20.59%

different approaches to produce the intervals, they share the same explanatory variables and the same lags. The model, due to its simple, linear form, may not perform well when the price fluctuates a lot. Yet, because we mainly focus on comparing the two approaches-ILS and CI, how to improve the ability of the model by, let us say, adding new variables, is not in our scope currently. However, since the model is made for point data, to make a model for interval data may improve overall forecast quality of ILS.

5 Conclusions and Future Work

In this work we study the potentials of interval arithmetic to forecast economic variables which manifest high uncertainty. Interval least square method is used to predict monthly WTI oil spot price. We use data ranging from January 1992 to April 2003 to build the relative stock model (RSTK) based on which we calculate the 95% confidence intervals for the rolling forecasts from December 1999 (in-sample) /January 2000 (out-of-sample) to April 2003. Then we apply the ILS approach to the RSTK model to obtain direct interval forecasts for the same period. The forecast results of traditional confidence interval method and ILS method are compared from two views, the view of researchers of interval arithmetic who emphasize intersections of intervals and the view of traditional decision makers who are familiar with average values. The empirical study shows that for predicting oil prices, the interval method is al-most as good as the tradition one. Thus, it is reasonable to use interval method in this case as well.

We believe that interval method is more meaningful in analyzing constantly changing economic variables and thus it is worth further studies following this one. We mention three here. First, asymmetric intervals of relative inventory level, for example, adding 40% and subtracting 60% to the point data, may improve forecast quality, since there are literatures arguing about asymmetric effects, like Ye et al. [7]. Second, the ILS method has the potential for improvement. For instance, we can use information about the width of intervals to gain better estimation of the coefficients. Since the estimation includes information of both the midpoint and the width, the relationship between variables may become more reliable. Finally, further efforts are needed to establish theoretical foundation for interval economic analysis.

Acknowledgments

Authors of this paper would like to express their appreciations to the anonymous referees for their insightful comments that helped us to improve the quality of this paper. Authors would also like to thank Professor Chenyi Hu for his technical contributions on the software that we used to perform the empirical experiments.

References

1. Considine, T.J., Larson, D.F.: Risk premiums on inventory assets: The case of crude oil and natural gas. The Journal of Futures Markets 21(2), 109–126 (2001)
2. Hu, C., He, L.T.: An application of interval methods to stock market forecasting. Reliable Computing 13(5), 423–434 (2007)
3. Moore, R.E.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
4. Nooner, M., Hu, C.: A computational environment for interval matrices. In: The NSF 2006 Workshop on Reliable Engineering Computing, Savannah, GA, pp. 65–74 (2006)

5. Pindyck, R.S.: The dynamics of commodity spot and futures markets: A primer. The Energy Journal 22(3), 1–29 (2001)
6. Ye, M., Zyren, J., Shore, J.: A monthly crude oil spot price forecasting model using relative inventories. International Journal of Fore-casting 21, 491–501 (2005)
7. Ye, M., Zyren, J., Shore, J.: Forecasting short-run crude oil price using high-and low-inventory variables. Energy Policy 34, 2736–2743 (2006)
8. Homepage of Interval Computations, http://www.cs.utep.edu/interval-comp/

Automatic Classification for Decision Making of the Severeness of the Acute Radiation Syndrome

Karin Mehr and Eberhard P. Hofer

[1] Institute of Measurement, Control, and Microtechnology, University of Ulm, 89069 Ulm, Germany
karin.mehr@uni-ulm.de
[2] Institute of Measurement, Control, and Microtechnology, University of Ulm, 89069 Ulm, Germany
eberhard.hofer@uni-ulm.de

Summary. Decision making of the severeness of the acute radiation syndrome is a major challenge due to the fact that the radiation dose is not known and cannot easily be reconstructed. Although radiation accidents are relatively rare, an automatic damage classification for individual medical prognosis of the health status of a patient is necessary. Under the threat of nuclear terroristic attacks the problem has received special attention.

Early classification of the severeness of the damage in case of the acute radiation syndrome allows for separation of patients with irreversibly damaged cell renewal systems from those with reversible damage. As a consequence, the available resources for the treatment of patients can be used most efficiently and, especially, therapeutic actions for patients with irreversible cell damage can be taken in time.

This paper concentrates on the damage to the hemopoietic system. Measurements from available patient data are represented by time series of cell counts of the relevant blood cell lines. Features extracted from reduced dynamic models form the basis for automatic damage classification where emphasis is on an early classification based on a time horizon not exceeding 10 days. Our newest results guarantee a generalization rate in the range of 80%.

1 Introduction

Management of acute radiation accidents requires early recognition of the medical severeness of the irradiation damage. It is known that damage to the hemopoietic system is a key indicator for classification [1, 3, 4, 11]. In the decision making approach grading codes H1 to H4 are extracted from blood cell counts of platelets, granulocytes, and lymphocytes of the individual patient. On this scale, H4 represents the most severe irreversible radiation damage which requires stem cell transplant. Consequently, a reliable and fast automatic classification is necessary to separate grading code H4 mainly from grading code H3.

V.-N. Huynh et al. (Eds.): Interval/Probabilistic Uncertainty, ASC 46, pp. 364–371, 2008.
springerlink.com

The classification results could be substantially increased [6, 7] by an improved new concept concerning the lymphocyte cell count in peripheral blood after radiation exposure [8]. For a reduced time horizon of 10 days the developed lymphopoietic modeling is a excellent starting point for the subsequent feature extraction. In this paper, an automatic classification based on feature extraction from patient data is proposed for early decision making of the severeness of the acute radiation syndrome.

2 Feature Extraction

Patient data from cell counts of platelets, granulocytes, and lymphocytes in the peripheral blood after acute radiation form the basis for feature extraction [5, 8]. Projections of the 4-dimensional feature space into two selected 2-dimensional feature spaces are shown in Figure 1. The plot indicates that there is no dominant clustering. Consequently an automatic classification procedure is a major challenge.

3 Automatic Classification

As a common basis for automatic classification we use the information of the given training data represented in the 4-dimensional feature space x in association with the expert classification ω.

Due to the fact that nuclear accidents are rare events only a limited number of sufficiently documented data records are available. For our classification we use 65 sets of patient data recorded at the WHO Ulm database together with their expert gradings (8 H1, 14 H2, 16 H3 and 27 H4). For training 50 sets are used and for test purposes the remaining 15 sets with expert gradings (2 H1, 3 H2, 4 H3 and 6 H4) have been selected. This leads to $5.49 \cdot 10^{12}$ combinations [7]. From these combinations 10,000 have been selected randomly for evaluation of the classification approach.

3.1 Bayesian Classification

The feature dependent probabilities $P(\omega_j|x)$ for the four damage classes ω_j with $j = 1 \ldots 4$, form the basis of the Bayesian classification [2, 10]. At first the classification for the class ω_i finally results from the probability distribution $P(\omega_i|x) > P(\omega_j|x)$ for all $j \neq i$.

Using the theorem of Bayes the probability

$$P(\omega_j|x) = \frac{p(x|\omega_j) \cdot P(\omega_j)}{\sum_j p(x|\omega_j) \cdot P(\omega_j)} \tag{1}$$

can be calculated from the a-priori probability for the damage class $P(\omega_j)$ and the class dependent probability density function $p(x|\omega_j)$ for the feature values x. The a-posteriori class dependent probability density function $p(x|\omega_j)$ is usually

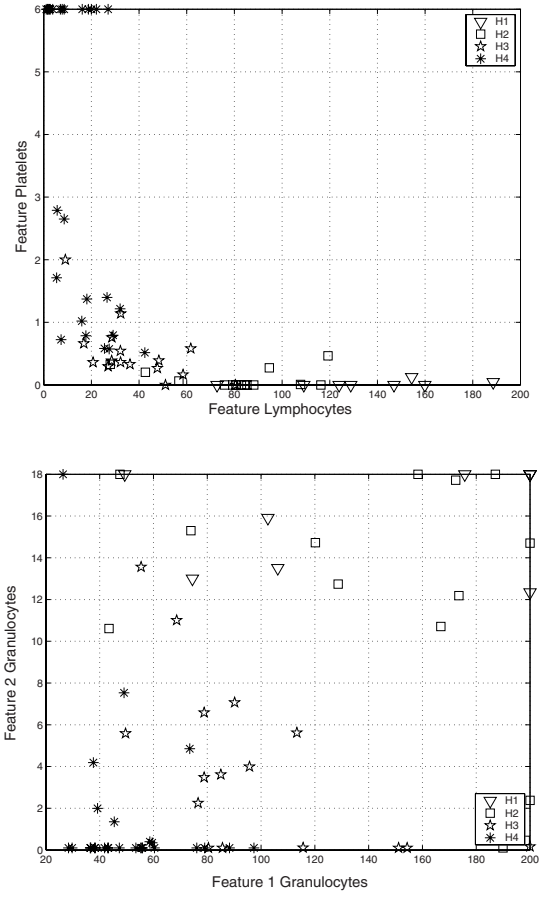


Fig. 1. Projections from the 4-dimensional feature space with feature platelets, feature lymphocytes, feature 1 granulocytes and feature 2 granulocytes in association with the given expert classifications H1 (mild) H2 (moderate), H3 (severe), and H4 (irreversibel damage)

determined from histograms in case of a low dimensional feature space and a large number of measurement data. In the case under consideration class dependent probability densities are estimated as parametric distributions using empirical mean value μ_j and covariance matrix Σ_j for each class ω_j.

In our approach the a-priori probabilities $P(\omega_j)$ for the four classes $j = 1 \ldots 4$ have been assumed to be equal, i.e., $P(\omega_j) = 0.25$. Using $P(\omega_j|x) \approx p(x|\omega_j)$ the automatic classification can be simplified considerably with

$$P(\omega_j|x) \approx p(x|\omega_j) \;=\; \frac{1}{|\Sigma_j|^{1/2} \cdot (2\pi)^{n/2}} \cdot e^{-\frac{1}{2}(x-\mu_j)^T(\Sigma_j)^{-1}(x-\mu_j)}. \qquad (2)$$

In a second step a decision rule to minimize the probability of false classification is introduced. The damage class ω_k with the smallest risk of error classification

$$R(\omega_{Ai}|x) = \sum_j \lambda(\omega_{Ai}|\omega_{Ej}) \cdot P(\omega_{Ej}|x)$$

will be selected. For the expert classification ω_{Ej} the 4×4 loss function matrix

$$\lambda(\omega_{Ai}|\omega_{Ej}) = \begin{cases} 0 & i = j \\ a & i < j \\ b & i > j \end{cases} \quad \text{for } i, j = 1, \ldots, 4$$

is linked with the possible automatic classification results ω_{Ai}. The loss function matrix allows to steer the automatic classification result.

If the automatic classification result corresponds to the expert classification with $i = j$ the loss function becomes zero. In all other cases $i \neq j$ special weightings are possible. In our application $a > b$ corresponds to a higher error probability of weak damage classes. Finally for the automatic classification this results in stronger damage classes. The influence of the weighting relation a/b on the generalization rate (accuracy) is shown in Figure 2.

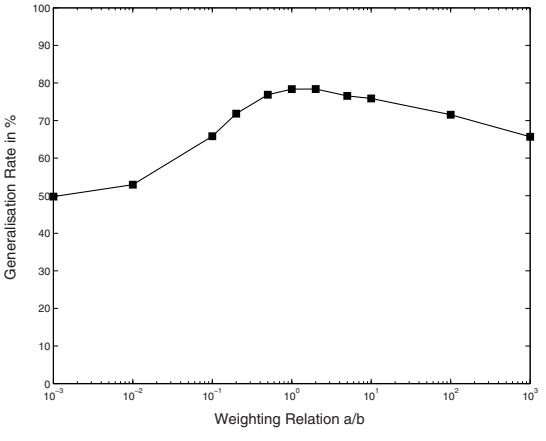


Fig. 2. Generalization rate using Bayesian classification

For a detailed evaluation the confusion matrix for the Bayesian classification is given in Table 1. The sensitivity indicates the percentage of correctly classified data for the respective expert classification. The precision corresponds to the probability of the correctness of the result. The element in the lower right corner corresponds to the accuracy of the correctly classified data records. As a result, on the average 78.4% of the unknown test data are classified correctly.

Table 1. Confusion matrix of automatic classification results with Bayesian classification using the weighting relation $a/b = 2$

		Expert Classification				
		H1	H2	H3	H4	Precision
Test Result	H1	**1.3687**	0.6486	–	–	0.6785
	H2	0.6313	**1.7883**	0.2619	–	0.6669
	H3	–	0.5630	**3.0737**	0.4752	0.7475
	H4	–	0.0001	0.6644	**5.5248**	0.8926
	Sensitivity	0.6844	0.5961	0.7684	0.9208	**0.7837**

3.2 Classification with Parzen Windows

A nonparametric method to calculate the necessarily class dependent probability density function $p(x|\omega_j)$ for the feature values x is based on specific kernel or window functions [2, 9]. In our approach kernel functions $\varphi(u)$ are gaussians with localization at every feature space value x_i and variable specific univariat standard deviation h_j. The number c_j includes all training data sets with the expert classification ω_j

$$p(x|\omega_j) = \frac{1}{c_j} \cdot \sum_{i=1}^{c_j} \frac{1}{h_j} \varphi\left(\frac{x - x_i}{h_j}\right). \tag{3}$$

Figure 3 shows the characteristic dependence of the generated probability density function from the used kernel parameter h_j. The computed probability density function for the kernel parameter $h_j = 0.15$ is very similar to the Bayesian case.

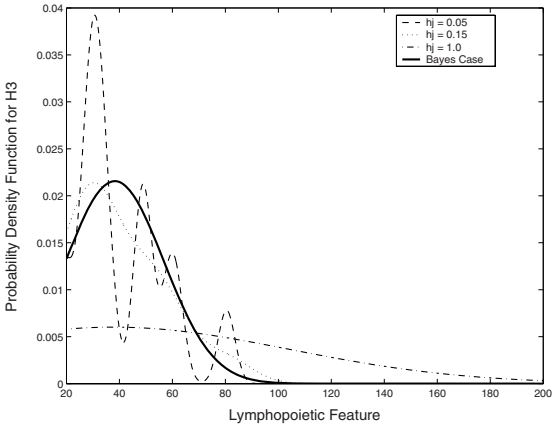


Fig. 3. Probability density function for different kernel parameters in relation to the Bayesian case for the lymphopoietic feature of all available data sets for H3

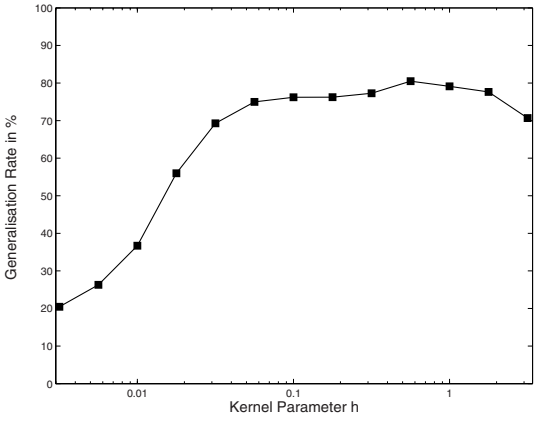


Fig. 4. Generalization rate using classification with Parzen windows

The relation between class specific kernel parameter h_j and general kernel parameter h is given by

$$h_j = \frac{h}{\sqrt{c_j}}. \tag{4}$$

Details of the evaluation of the classification with Parzen windows for the 15 test cases are given by the confusion matrix in Table 2, analogously to the Bayesian classification in Table 1. Best results for the accuracy are obtained for the kernel parameter $h = 0.56$ with an average of 80.5% correctly classified unknown test data.

Table 2. Confusion matrix of automatic classification results with Parzen windows method using the kernel parameter $h = 0.56$

		Expert Classification				
		H1	H2	H3	H4	Precision
Test Result	H1	**1.4686**	0.3907	–	–	0.7899
	H2	0.5314	**2.0048**	0.3045	–	0.7057
	H3	–	0.6045	**3.1852**	0.5850	0.7281
	H4	–	–	0.5103	**5.4150**	0.9139
	Sensitivity	0.7343	0.6683	0.7963	0.9025	**0.8049**

4 Reliability of the Diagnosis Result

Both methods presented here are suitable for practical use. In every case an accuracy of 78,4% (Bayes) to 80,5% (Parzen) is reached for all damage classes H1 to H4. For both cases the false classification is at maximum the neighboring damage class. Finally, the sensitivity for the irreversibel damage class H4 reaches

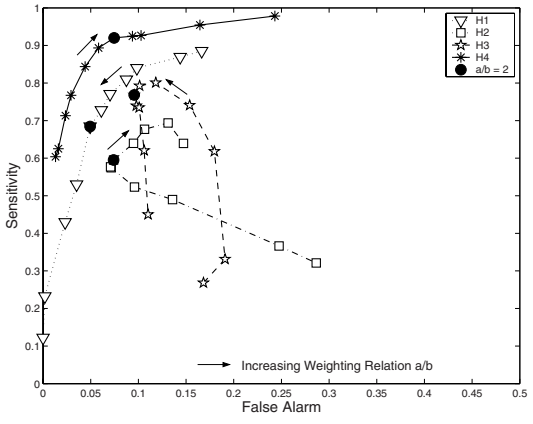


Fig. 5. Receiver operating characteristic using Bayesian classification

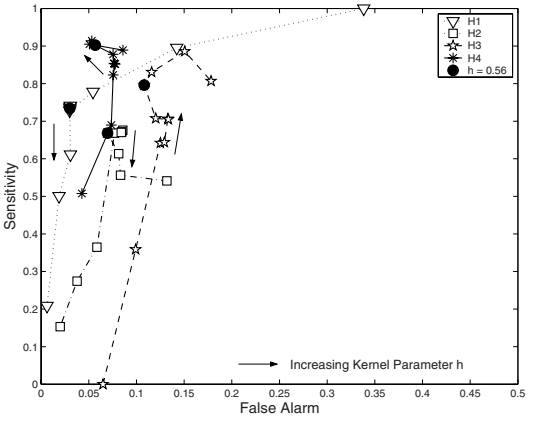


Fig. 6. Receiver operating characteristic using classification with Parzen windows

92,1% (Bayes) and 90,3% (Parzen) in connection with a false alarm rate of 7,4% (Bayes) and 5,7% (Parzen). Figures 5 and 6 show the receiver operating characterstics for the described automatic classification methods.

The individual identification of the irreversible damaged hemopoietic system is extremely important. Using our method it is possible to separate reversible damage (H1, H2 and H3) and irreversible damage (H4) within 10 days after exposition very reliable . The likelihood ratio LR^+ for the positive test (H4) results in 14,45 (Bayes) and 15,84 (Parzen), respectively. The likelihood ratio LR^- for the negative test (not H4) results in 0,085 (Bayes) and 0,102 (Parzen), respectively.

5 Conclusion and Outlook

In this paper, the acute radiation syndrome has been treated under the aspect of early damage classification. An automatic classification procedure using Bayesian classification as well as classification with Parzen windows lead to best results. In summary, the two presented automatic classification methods are excellent exclusion tests as well as diagnostic tests for an irreversible damage of the hemopoietic system. This form take basis to develop a very reliable diagnostic tool for decision making. A further goal must be to reduce the time horizon to a few days after radiation exposure.

References

1. Bond, V.P., Fliedner, T.M., Archambeau, J.O.: Mammalian Radiation Lethality - A Disturbance In Cellular Kinetics. Academic Press, New York (1965)
2. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. John Wiley & Sons, Inc., Wiley-Interscience (2001)
3. Fliedner, T.M., Friesecke, I., Beyrer, K.: Medical Management of Radiation Accidents. British Institute of Radiology, London (2001)
4. Fliedner, T.M.: Nuclear terrorism: The role of hematology in coping with its health consequences. Current Opinion in Hematology 13, 436–444 (2006)
5. Grässle, D.H., et al.: Classification of the Individual Medical Severeness of Radiation Accidents within Short Time. In: Shimemura, E., Fujita, M. (eds.) Proc. The 10th Japanese - German Seminar on Nonlinear Problems in Dynamical Systems - Theory and Applications, Hakui, Ishikawa, Japan, pp. 127–136 (2002)
6. Hofer, E.P., Mehr, K.: Decision Support for Treatment of Radiation Accidents. In: Noguchi, H., Ishii, H., Iuniguchi, M. (eds.) Proc. The 7th Czech - Japan Seminar on Data Analysis and Decision Making under Uncertainty, Hyogo, Japan, pp. 69–72 (2004)
7. Hofer, E.P., Mehr, K.: An Early Damage Classification for the Acute Radiation Syndrome. In: Kroupa, T., Vejnarova, J. (eds.) Proc. The 8th Czech - Japan Seminar on Data Analysis and Decision Making under Uncertainty, Trest, Czech Republic, pp. 48–54 (2005)
8. Mehr, K., Hofer, E.P.: Feature Extraction in Lymphopoietic Modeling for Radiation Damage Classification on a Short Time Horizon. In: Proc. The 9th Czech - Japan Seminar on Fuzzy Systems and Innovational Computing, Kitakyushu City, Japan, pp. 357–360 (2006)
9. Parzen, E.: On estimation of a probability density function and mode. Annals of Mathematical Statistics 33(3), 1065–1076 (1962)
10. Schürmann, J.: Pattern Classification - A Unified View of Statistical and Neural Approaches. John Wiley & Sons, New York (1996)
11. Waselenko, J.K., et al.: Medical Management of the Acute Radiation Syndrome: Recommendations of the Strategic National Stockpile Radiation Working Group. Annals of Internal Medicine 140(12), 1037–1051 (2004)

Index