ADVANCES IN SOFT COMPUTING 42

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

Springer

Advances in Soft Computing 42

Editor-in-Chief: J. Kacprzyk

Advances in Soft Computing

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

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

Abraham Ajith, Yasuhiko Dote, Takeshi Furuhashi, Mario Köppen, Azuma Ohuchi, Yukio Ohsawa (Eds.)
Soft Computing as Transdisciplinary Science and Technology, 2005
ISBN 978-3-540-25055-5

Barbara Dunin-Keplicz, Andrzej Jankowski, Andrzej Skowron, Marcin Szczuka (Eds.)
Monitoring, Security, and Rescue Techniques in Multiagent Systems, 2005
ISBN 978-3-540-23245-2

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

Patrica Melin, Oscar Castilo, Eduardo Gomez Ramirez, Janus Kacprzyk, Witold Pedrycz (Eds.)
Analysis and Design of Intelligent System 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

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

Springer

Editors

Prof. Oscar Castillo
Tijuana Institute of Technology
Department of Computer Science,
Tijuana, Mexico
Mailing Address
P.O. Box 4207
Chula Vista CA 91909, USA
ocastillo@tectijuana.mx

Prof. Patricia Melin
Tijuana Institute of Technology
Department of Computer Science,
Tijuana, Mexico
Mailing Address
P.O. Box 4207
Chula Vista CA 91909, USA
pmelin@tectijuana.mx

Prof. Oscar Montiel Ross
CITEDI-IPN
Av. del Parque, Tijuana, Mexico
oross@citedi.mx

Prof. Roberto Sepúlveda Cruz
CITEDI-IPN
Av. del Parque, Tijuana, Mexico
rsepulve@citedi.mx

Prof. Witold Pedrycz
University of Alberta,
Dept. Electrical and Computer
Engineering
Edmonton, Alberta T6J 2V4,
Canada
pedrycz@ece.ualberta.ca

Prof. Janusz Kacprzyk
Polish Academy of Sciences,
Systems Research Institute,
Newelska 6 01-447 Warszawa
Poland
kacprzyk@ibspan.waw.pl

Library of Congress Control Number: 2007926021

ISSN print edition: 1615-3871
ISSN electronic edition: 1860-0794
ISBN-10 3-540-72433-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72433-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: by the authors and SPS using a Springer LaTeX macro package

Printed on acid-free paper SPIN: 12060313 89/SPS 5 4 3 2 1 0

Preface

This book comprises a selection of papers from IFSA 2007 on theoretical advances and applications of fuzzy logic and soft computing. These papers were selected from over 400 submissions and constitute an important contribution to the theory and applications of fuzzy logic and soft computing methodologies. Soft Computing consists of several computing paradigms, including fuzzy logic, neural networks, genetic algorithms, and other techniques, which can be used to produce powerful intelligent systems for solving real-world problems. The papers of IFSA 2007 also make a contribution to this goal.

This book is intended to be a major reference for scientists and engineers interested in applying new fuzzy logic and soft computing tools to achieve intelligent solution to complex problems. We consider that this book can also be used to get novel ideas for new lines of research, or to continue the lines of research proposed by the authors of the papers contained in the book.

The book is divided in to sixteen main parts. Each part contains a set of papers on a common subject, so that the reader can find similar papers grouped together. Some of these parts are comprised from the papers of organized sessions of IFSA 2007 and we thank the session's organizers for their incredible job on forming these sessions with invited and regular paper submissions.

In Part I, we have two papers on "Intuitionistic Fuzzy Sets and their Applications" from a session organized by Eulalia Szmidt and Janusz Kacprzyk. These papers show important theoretical results, as well as novel applications of intuitionistic fuzzy logic. The area of intuitionistic fuzzy logic has also become a potential area of promissory results for the future of fuzzy logic.

In Part II, we have a collection of papers on the topic of "The application of fuzzy logic and Soft Computing in Flexible Querying" from a session organized by Guy DeTre and Slawek Zadrozny. These papers show important theoretical results and applications of fuzzy logic and soft computing in achieving flexible querying for database systems. The area of flexible querying has become an important subject for achieving intelligent interfaces with human users and for managing large databases.

In Part III, we have a collection of papers on "Philosophical and Human Scientific Aspects of Soft Computing" from a session organized by Vesa A. Niskanen. These papers show the interesting relationships between the philosophical aspects of soft computing and the formal-scientific aspects of soft computing. Papers on this subject are very important because they help in understanding the area of soft computing, and also enable proposing new theories and methods in this area.

In part IV, we have a collection of papers on "Search Engine and Information Processing and Retrieval" from a special FLINT Session organized by Masoud Nikravesh.

These papers describe important contributions on search engines for the web, summarization, computing with words and granular computing, for information processing and retrieval. Papers on these subjects are very important theoretically as well as in the applications because of the importance of web search for documents and images.

In Part V, we have a set of papers on “Perception Based Data Mining and Decision-Making” from a Session organized by Ildar Batyrshin, Janusz Kacprzyk, and Ronald R. Yager. These papers constitute an important contribution to data mining and linguistic summarization using fuzzy logic. Papers on these subjects are very important because data mining and building summaries are necessary in managing large amounts of data and information.

In Part VI, we have a set of papers on “Soft Computing in Medical Sciences” from a session organized by Rudolf Seising and Christian Schuh. These papers describe important contributions on the use of different soft computing methodologies for solving problems in medicine. Papers on this area are particularly important due to wide variety and complexity of the problems addressed in the medical sciences.

In Part VII, we have a collection of papers on “Joint Model-Based and Data-Based Learning: The Fuzzy Logic Approach” from a session organized by Joseph Aguilar-Martin and Julio Waissman Vilanova. These papers describe important contributions to solving the problems of learning in different types of models using fuzzy logic. Also, the new learning methods are applied to different applications. Learning from data and models is very important for solving real-world problems.

In Part VIII, we have a group of papers on “Fuzzy/Possibilistic Optimization” from a session organized by Weldon Lodwick. These papers describe important theoretical results and applications of fuzzy optimization. The optimization problem is considered from the point of view of fuzzy logic, which gives better results than traditional approaches.

In Part IX, we have a set of papers on “Algebraic Foundations of Soft Computing” from a session organized by Irina Perfilieva and Vilem Novak. These papers represent a significant contribution to the state of the art in the mathematical foundations of different areas in soft computing. The theoretical results will be the basis for future developments in soft computing.

In Part X, we have a group of papers on the subject of “Fuzzy Trees” from a session organized by Ziheng Huang and Masoud Nikravesh. These papers show important theoretical results and applications of fuzzy trees. The use of fuzzy trees is very important as a model of human decision making and for this reason can have many real-world applications.

In Part XI, we have a group of papers on “Soft Computing in Petroleum Applications” from a session organized by Leonid Sheremetov and Masoud Nikravesh. These papers describe important theoretical results and applications of soft computing in real-world problems in the petroleum industry. The point of view of soft computing for this type of problems gives better results than traditional approaches.

In Part XII, we have a group of papers on “Fuzzy Logic and Soft Computing in Distributed Computing” from a session organized by Lifeng Xi and Kun Gao. These papers describe important theoretical results and applications of fuzzy logic and soft computing to problems in distributed computing. The proposed methods using fuzzy logic and soft computing give better results than traditional approaches

In Part XIII, we have a collection of papers on “Fuzzy Logic Theory” describing different contributions to the theory of fuzzy logic. These papers show mainly theoretical results on fuzzy logic that can help advance the theory and/or provide fundamental tools for possible solutions to real-world problems.

In Part XIV, we have a group of papers on “Fuzzy Logic Applications” that show a wide range of applications of fuzzy logic theory. The papers describe in detail important real-world problems that have solved satisfactorily with fuzzy systems. Also, the fuzzy solutions are shown to be better than traditional solutions to these problems.

In Part XV, we have a collection of papers on “Neural Networks” that comprise theoretical contributions on neural networks and intelligent systems developed with neural models, as well as real applications of these areas. The papers represent an important contribution to the state of the art in both theory and applications of neural networks.

In Part XVI, we have a collection of papers on “Soft Computing” comprising important contributions in this field. These papers show theoretical results and important applications of soft computing methodologies. Also, there are papers on hybrid intelligent systems, that combine several soft computing techniques.

We end this preface of the book by giving thanks to all the people who have help or encourage us during the compilation of this book. We would like to thank our colleagues working in Soft Computing, which are too many to mention each by their name. Of course, we need to thank our supporting agencies in our countries for their help during this project. We have to thank our institutions, for always supporting our projects.

Prof. Dr. Oscar Castillo
Tijuana Institute of Technology,
Mexico

Prof. Dr. Patricia Melin
Tijuana Institute of Technology,
Mexico

Prof. Dr. Oscar Montiel
CITEDI-IPN, Tijuana,
Mexico

Prof. Roberto Sepúlveda
CITEDI-IPN, Tijuana,
Mexico

Prof. Dr. Witold Pedrycz
University of Alberta,
Canada

Prof. Dr. Janusz Kacprzyk
Polish Academy of Sciences,
Poland

Organization

IFSA 2007 is organized by the International Fuzzy Systems Association (IFSA), the Hispanic American Fuzzy Systems Association (HAFSA), and the Division of Graduate Studies of Tijuana Institute of Technology, with support of DGEST and CONACYT of Mexico.

Executive Committee

Honorary Chair

Lotfi Zadeh,
University of California at Berkeley

General Chair

Oscar Castillo,
Tijuana Institute of Technology

Program Chair

Patricia Melin,
Tijuana Institute of Technology

Advisory Committee

Janusz Kacprzyk
Polish Academy of Sciences

Witold Pedrycz
University of Alberta

Ronald R. Yager
Iona College, USA

Panels Chair

Masoud Nikravesh
University of California at Berkeley

Workshop Chair

Eduardo Gomez-Ramirez,
La Salle University, Mexico

International Program Committee

Ildar Batyrshin, Mexican Petroleum Institute, Mexico
Z. Zenn Bien, KAIST, Korea
Ulrich Bodenhofer, Johannes Kepler University, Austria
Piero Bonissone, GE Global Research, USA
Patrick Bosc, IRISA/ENSSAT, France
Bernadette Bouchon-Meunier, University Pierre Marie Curie, France
Christer Carlsson, IAMSR/Abo Akademi University, Finland
Joao P. B. Carvalho, INESC, TU Lisbon, Portugal
Oscar Castillo, Tijuana Institute of Technology, Mexico
Carlos Coello, CINVESTAV, Mexico
Oscar Cordon, European Centre for Soft-Computing, Spain
Guoqing Chen, Tsinghua University, China
Bernard De Baets, Ghent University, Belgium
Hai-Bin Duan, Beihang University, China
Pablo Estevez, University of Chile, Chile
Dimitar Filev, Ford Research Center, USA
Takeshi Furuhashi, Nagoya University, Japan
Eduardo Gomez-Ramirez, La Salle University, Mexico
Fernando Gomide, State University of Campinas, Brazil
Madan M. Gupta, University of Saskatchewan, Canada
Mustafa Gunes, King Abdulaziz University, Saudi Arabia
Janos Grantner, Western Michigan University, USA
Kaouru Hirota, Tokyo Institute of Technology, Japan
Hani Hagras, University of Essex, UK
Robert John, De Montfort University, UK
Cengiz Kahraman, Istanbul Technical University, Turkey
Janusz Kacprzyk, Polish Academy of Sciences, Poland
Jim Keller, University of Missouri, USA
Etienne Kerre, Ghent University, Belgium
George Klir, Binghamton University-SUNY, USA
Lazlo Koczy, Budapest University of Technology and Economics, Hungary
Rudolf Kruse, Universitaet Magdeburg, Germany
Vladik Kreinovich, University of Texas, El Paso, USA
Reza Langari, Texas A & M University, USA
Vincenzo Loia, Universita di Salerno, Italy
Luis Magdalena, European Centre for Soft Computing, Spain
E. Mamdani, University of London, UK
Patricia Melin, Tijuana Institute of Technology, Mexico
Jerry Mendel, University of Southern California, USA
Evangelia Micheli-Tzanakou, Rutgers University, USA
K. C. Min, Korea,
Ying Mingsheng, Tsinghua University, China
Masoud Nikravesh, University of California, Berkeley, USA
V. Novak, University of Ostrava, Czech Republic
Jose A. Olivas-Varela, Universidad de Castilla La Mancha, Spain

Gabriella Pasi, Universita degli Studi di Milano, Italy
Witold Pedrycz, University of Alberta, Canada
Frank Chung-Hoon Rhee, Hanyang University, Korea
Rajkumar Roy, Cranfield University, UK
Edgar N. Sanchez, CINVESTAV-Guadalajara, Mexico
Pilar Sobrevilla, Universidad Politecnica de Cataluna, Spain
Joao M. C. Sousa, IDMEC, TU Lisbon, Portugal
Thomas Sudkamp, Wright State University, USA
Michio Sugeno, Doshisha University, Japan
Ricardo Tanscheit, Catholic University of Rio de Janeiro, Brazil
Tsu-Tian Lee, National Taipei University of Technology, Taiwan
Enric Trillas, European Centre for Soft Computing, Spain
I. Burhan Turksen, Economics and Technology University, Turkey, University of Toronto, Canada
Jose L. Verdegay, University of Granada, Spain
Michael Wagenknecht, Univ. of Applied Sciences Zittau, Germany
Ronald R. Yager, Iona College, USA
Liu Yingming, Sichuan University, China
Hans Zimmerman, Aachen University of Technology, Germany
Jacek Zurada, University of Louisville, USA

Local Organizing Committee

Luis Tupak Aguilar Bustos,
CITEDI-IPN, Tijuana
Arnulfo Alanis,
Tijuana Institute of Technology, Tijuana
Oscar Castillo
Tijuana Institute of Technology, Tijuana
Carlos A. Coello Coello,
CINVESTAV, Mexico
Mario Garcia V.
Tijuana Institute of Technology, Tijuana
Pilar Gómez Gil,
Universidad de las Americas Puebla, Puebla
Eduardo Gómez Ramírez,
Universidad de La Salle, México
Miguel Angel Lopez
Tijuana Institute of Technology, Tijuana
Gerardo Maximiliano Méndez,
Instituto Tecnológico de Nuevo León, Monterrey
Alejandra Mancilla
Tijuana Institute of Technology, Tijuana
Patricia Melin
Tijuana Institute of Technology, Tijuana
Oscar Montiel Ross
CITEDI-IPN, Tijuana

Antonio Rodríguez Díaz,
Univ. Autónoma de Baja California, Tijuana
Edgar N. Sánchez,
CINVESTAV, Guadalajara
Roberto Sepúlveda Cruz
CITEDI-IPN, Tijuana

List of Additional Reviewers

Troels Andreassen, Denmark
Gloria Bordogna, Italy
Jesús Campana, Spain
Eliseo Clementini, Italy
Oscar Cordon, Spain
Fabio Crestani, Switzerland
Guy De Tre, Belgium
Marcin Detyniecki, France
Axel Hallez, Belgium
Enrique Herrera-Viedma, Spain
Janusz Kacprzyk, Poland
Donald Kraft, USA
Zongmin Ma, China
Luis Mancilla, Mexico
Milan Mares, Czech Republic
Nicolas Marin, Spain
Christophe Marsala, France
Maria-Jose Martin-Bautista, Spain
Tom Matthe, Belgium
Oscar Montiel, Mexico
Adam Niewiadomski, Poland
Gabriella Pasi, Italy
Frederick E. Petry, USA
Olivier Pivert, France
Olga Pons, Spain
Ellie Sanchez, France
Roberto Sepúlveda, Mexico
Eulalia Szmidt, Poland
Joerg Verstraete, Belgium
Manolis Wallace, Greece
Slawomir Zadrozny, Poland

Contents

Part XIV: Fuzzy Logic Applications

Part I

Intuitionistic Fuzzy Sets and Their Applications

An Intuitionistic Fuzzy Graph Method for Finding the Shortest Paths in Networks

M.G. Karunambigai[1], Parvathi Rangasamy[2], Krassimir Atanassov[3], and N. Palaniappan[4]

[1,2] Department of Mathematics, Vellalar College for Women, Erode, TN, India – 638 009
gkaruns@yahoo.co.in, paarvathis@rediffmail.com

[3] CLBME – Bulgarian Academy of Sciences, P.O. Box 12, Sofia – 1113, Bulgaria
krat@bas.bg

[4] Department of Mathematics, Alagappa University, Karaikudi, TN, India – 630 003
palaniappan1950@yahoo.com

Abstract. The task of finding shortest paths in graphs has been studied intensively over the past five decades. Shortest paths are one of the simplest and most widely used concepts in networks. More recently, fuzzy graphs, along with generalizations of algorithms for finding optimal paths within them, have emerged as an adequate modeling tool for imprecise systems. Fuzzy shortest paths also have a variety of applications. In this paper, the authors present a model based on dynamic programming to find the shortest paths in intuitionistic fuzzy graphs.

Keywords: Index Matrix (IM), Intuitionistic Fuzzy Graphs (IFGs), Shortest path, Dynamic programming (DP).

1 Introduction

Over the past several years a great deal of attention has been paid to mathematical programs and mathematical models that can be solved through the use of networks. There has been a plenty of articles on network programming and several significant advances have been made. Through these advances efficient algorithms have been developed for even large scale programs. Posing problems on networks not only yields computational advantages, it also serves as a means for visualizing a problem and for developing a better understanding of the problem.

The aim of this paper is to concentrate on the most basic network problem, the shortest path problem. The fuzzy shortest path problem was first analyzed by Dubois and Prade [7]. However the major drawback to this problem is the lack of interpretation. To overcome this situation, a new model based on shortest paths in intuitionistic fuzzy graphs is presented. An algorithm for this model based on DP recursive equation approach is also developed and checked with the local telecommunication department map.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 3–10, 2007.
springerlink.com

In this paper, Section 2 provides preliminary concepts required for analysis. In Section 3 and 4, the shortest path problem in intuitionistic fuzzy graphs is introduced and new model has been developed. A case study work is carried out in Section 5 considering a local city's telecommunications department map as an IFG. Section 6 concludes the paper.

2 Preliminaries

Definition 2.1 [6]**.** A *graph* consists of a structure $G = (V, E)$, where V is a set of *vertices*, and the predicate $E \subseteq V \times V$ is a set of *edges*.

Definition 2.2 [4]**.** Let E be a non-empty set. An *Intuitionistic Fuzzy Set* (IFS) A in E is defined as an object of the form $A=\{(x, \mu_A(x), \gamma_A(x)): x \in E\}$ where the fuzzy sets $\mu_A : E \to [0,1]$ and $\gamma_A : E \to [0,1]$ denote the membership and non-membership functions of A respectively, and $0 \le \mu_A(x) + \gamma_A(x) \le 1$ for each $x \in E$.

Definition 2.3. Let I be a fixed set of indices and R be the set of all real numbers. By an *IM* with index sets K and L (K, L $\subset$ I), we mean the object [2]

$$[\,K, L, \{a_{k_i,l_j}\}] = \begin{array}{c|cccc} & l_1 & l_2 & \cdots & l_n \\ k_1 & a_{k_1,l_1} & a_{k_1,l_2} & \cdots & a_{k_1,l_n} \\ k_2 & a_{k_2,l_1} & a_{k_2,l_2} & \cdots & a_{k_2,l_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ k_m & a_{k_m,l_1} & a_{k_m,l_2} & \cdots & a_{k_m,l_n} \end{array}$$

where K = $\{k_1, k_2, \cdots, k_m\}$, L = $\{l_1, l_2, \cdots, l_n\}$, for $1 \le i \le m$, and for $1 \le j \le n : a_{k_i,l_j} \in R$. For the IMs A = [K, L, $\{a_{k_i,l_j}\}$] and B = [P, Q, $\{a_{k_i,l_j}\}$] the usual matrix operations "addition" and "multiplication" are defined[4].

Definition 2.4. Let E_1 and E_2 be two universes and let $A = \{\langle x, \mu_A(x), \gamma_A(x)\rangle : x \in E_1\}$, $B = \{\langle x, \mu_B(x), \gamma_B(x)\rangle : x \in E_2\}$ be two IFSs ; A – over E_1 and B – over E_2. Now define [3]

$$A \times_1 B = \{\langle\langle x, y\rangle, \mu_A(x).\mu_B(y), \gamma_A(x).\gamma_B(y)\rangle : \langle x, y\rangle \in E_1 \times E_2\}$$

$$A \times_2 B = \left\{\begin{array}{l}\langle\langle x, y\rangle, \mu_A(x) + \mu_B(y) - \mu_A(x).\mu_B(y), \gamma_A(x).\gamma_B(y)\rangle : \\ \langle x, y\rangle \in E_1 \times E_2\end{array}\right\}$$

$$A \times_3 B = \left\{\begin{array}{l}\langle\langle x, y\rangle, \mu_A(x).\mu_B(y), \gamma_A(x) + \gamma_B(y) - \gamma_A(x).\gamma_B(y)\rangle : \\ \langle x, y\rangle \in E_1 \times E_2\end{array}\right\}$$

$$A \times_4 B = \left\{ \begin{array}{l} \langle \langle x, y \rangle, \min(\mu_A(x), \mu_B(y)), \max(\gamma_A(x), \gamma_B(y)) \rangle : \\ \langle x, y \rangle \in E_1 \times E_2 \end{array} \right\}$$

$$A \times_5 B = \left\{ \begin{array}{l} \langle \langle x, y \rangle, \max(\mu_A(x), \mu_B(y)), \min(\gamma_A(x), \gamma_B(y)) \rangle : \\ \langle x, y \rangle \in E_1 \times E_2 \end{array} \right\}.$$

It must be noted that $A \times_i B$ is an IFS, but it is an IFS over the universe $E_1 \times E_2$, where "$\times_i$" is one of the five Cartesian products above and "$\times$" is the classical Cartesian product on ordinary sets (E_1 and E_2).

Definition 2.5 [5]**.** Let X and Y are arbitrary finite non-empty sets. Intuitionistic Fuzzy Relation (IFR) is an IFS R $\subset X \times Y$ of the form : R$=\{\langle \langle x, y \rangle, \mu_R(x, y), \gamma_R(x, y) \rangle : x \in X, y \in Y\}$, $\mu_R : X \times Y \to [0,1]$ and $\gamma_R : X \times Y \to [0,1]$ are the degrees of membership and non-membership as the ordinary IFSs or degrees of validity and non-validity of the relation R; and for every $\langle x, y \rangle \in X \times Y :\ 0 \le \mu_R(x, y) + \gamma_R(x, y) \le 1$.

3 Main Results

Let the oriented graph G = (V,A) be given, where V is a set of vertices and A is a set of arcs. Every graph arc connects two graph vertices.

In [5] an approach for introducing of an IFG is given. Here we will modify it in two directions on the basis of some ideas generated from IFS-theoretical and from IFS-decision making points of view. We shall start with the oldest version of the concept.

Let operation $\times$ denote the standard Cartesian product operation, while operation $\circ \in \{\times_1, \times_2, \times_3, \times_4, \times_5\}$.

Following [3] we shall note that the set $G^* = \{\langle \langle x, y \rangle, \mu_G(x, y), \gamma_G(x, y) \rangle : \langle x, y \rangle \in V \times V\}$ is called an o-IFG (or briefy, an IFG) if the functions $\mu_G : V \times V \to [0,1]$ and $\gamma_G : V \times V \to [0,1]$ denote the respective degrees of membership and non-membership of the element $\langle x, y \rangle \in V \times V$. These functions have the forms of the corresponding components of the o-Cartesian product over IFSs, and for all $\langle x, y \rangle \in V \times V :\ 0 \le \mu_G(x, y) + \gamma_G(x, y) \le 1$.

This approach supposes that the given set V and the operation o are choises and fixed previously and they will be used without changes.

On the other hand, following the IFS-interpretations in decision making, we can construct set V and values of functions μ_G and γ_G in the current time, for example on the basis of expert knowledge and we can change their forms on the next steps of the process of IFG's use.

Now, we shall introduce a definition of a new type of an IFG.[2]

Let E be a universe, containing fixed graph-vertices and let V $\subset$ E be a fixed set. Construct the IFS $V = \{(x, \mu_V(x), \gamma_V(x)) : x \in E\}$ where the fuzzy sets $\mu_V : E \to [0,1]$

and $\gamma_V : E \to [0,1]$ determine the degree of membership and the degree of non-membership to set V of the element (vertex) $x \in E$, respectively, and for every $x \in E$ such that $0 \le \mu_V(x) + \gamma_V(x) \le 1$.

Now, we shall use the idea for an IFS over universe that is an IFS over another universe [2] and will define the set $G^* = \{\langle\langle x, y\rangle, \mu_G(x, y), \gamma_G(x, y)\rangle : \langle x, y\rangle \in G \times G\}$ is called an o-Generalized IFG (or briefly, an GIFG) if the functions $\mu_G : V \times V \to [0,1]$ and $\gamma_G : V \times V \to [0,1]$ denote the respective degrees of membership and non-membership of the element (the graph arc) $\langle x, y\rangle \in V \times V$. As above, these functions have the forms of the corresponding components of the o-Cartesian product over IFSs, and for all $\langle x, y\rangle \in V \times V$ such that $0 \le \mu_G(x, y) + \gamma_G(x, y) \le 1$.

Definition 3.1. An i*ntuitionistic fuzzy path* in an IFG is a sequence of vertices and edges such that either one of the following conditions are satisfied:

(i) $\mu_G(x, y) > 0$ or (ii) $\mu_G(x, y) = 0$ and $\gamma_G(x, y) < 1$, for all $\langle x, y\rangle \in P$.

Example 3.2. Let $V = \{v_1, v_2, v_3, v_4\}$. Consider the following index matrix for the IFG G.

$$\begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{bmatrix} \langle 0,0\rangle & \langle 0.2,0.5\rangle & \langle 0.5,0.2\rangle & \langle 0,0\rangle \\ \langle 0.3,0.6\rangle & \langle 0,0\rangle & \langle 0.4,0.5\rangle & \langle 0.7,0\rangle \\ \langle 0,0\rangle & \langle 0,0\rangle & \langle 0.2,0.7\rangle & \langle 0.8,0.1\rangle \\ \langle 0,0\rangle & \langle 0.7,0\rangle & \langle 0,0\rangle & \langle 0,0\rangle \end{bmatrix} \end{array}$$

The graph corresponding to this IM is given in Fig. 1.

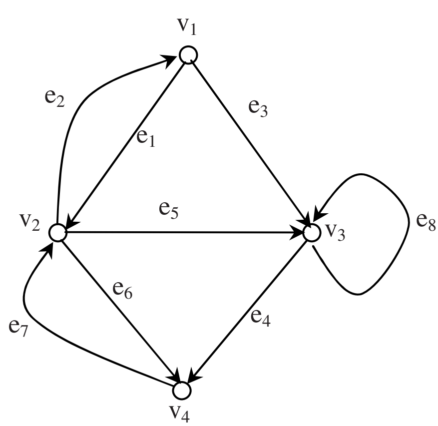


Fig. 1. Path in an IFG G

Here $v_1e_1v_2e_6v_4$, $v_1e_3v_3e_4v_4$, $v_1e_1v_2e_5v_3e_4v_4$ are intuitionistic fuzzy paths from v_1 to v_4.

4 Background

The models to be considered are an extension of a general network. A network is generally depicted by a graph and the terms will be used interchangeably. Let a graph, denoted by *G (V, E),* be a set of points *V*, and a set of pairs of these points *E*. The set V refers to the vertices of the graph and the set *E* refers to the edges of the graph. An edge is denoted by a pair of vertices {i, j}.

If *E* is changed to a set of ordered pairs of distinct elements of *V*, then *G(V, E)* is a directed graph and *E* is the set of ordered pairs (i, j). The ordered pairs (i, j) are referred to as arcs or edges and an arc goes from vertex i to vertex j. An arc (i, i) is referred to as a loop.

A path from a vertex s to a vertex t is a sequence of arcs of the form (s, i_1), (i_1, i_2),…(i_k, t). In other words, vertex t can be reached from vertex s. A path from s to t is denoted as an (s, t) path. An (s, t) path is open if s $\neq$ t and is closed if s = t. A cycle is a closed path (s, s) in which no vertices are repeated except s and there exists at least one arc. A graph that contains no cycles is called acyclic.

In a given graph both a source vertex and a sink vertex can be designated. These are interpreted as terminal vertices at which some activity begins and ends. In an acyclic directed graph with N vertices, the source can be labeled as vertex 1 and the sink as vertex N and all other vertices can be labeled such that for any arc (i,j), $i < j$.

A special type of an acyclic directed graph is a layered graph. This is an acyclic directed graph in which the vertex set *V* can be partitioned into M subsets, $V_1, \ldots, V_M$, such that if $|V_k| > 1$, the vertices in V_k are sequentially numbered and there does not exist an arc (i, j) for $i, j \in V_k$. For a layered graph, generally V_1 is the source vertex and V_M is the sink vertex.

If each arc (i, j) has an associated weight or length c_{ij}, then an (s, t) path has an associated weight or length equal to the sum of the weights of the arcs in the path. This in turn gives rise to the shortest path problem, which is to find the path with minimal weight between two vertices s and t. There are a variety of ways to find the shortest path for a network [8]. Some of them are general methods such as the labeling algorithm follow from DP. It is assumed that the graphs for the models to be presented are directed acyclic graphs. As any graph that has no cycles of negative weight can easily be converted to a directed acyclic graph [8], this is not a major restriction.

Assume the following in order to use the hybrid dynamic programming method for the shortest route problems and other network problems: (i) The network is directed and acyclic; (ii) the network is layered.

It should be noted that many applications naturally take on a layered network form [1]. For example, activity networks such as PERT are always directed and acyclic. It is also generally true that edges have positive lengths. This is especially true in terms of transportation related problems.

The multi-criteria DP recursion that will be used is

$$f(N) = (1, 1, 1\ .\ .\ .\ .\ .\ 1),$$
$$f(i) = \operatorname{dom}(e_{ij} \tilde{+} f(i)), \tag{1}$$

where e_{ij} is an R-tuple associated with each arc or edge (i, j) or the path from i to j. This R-tuple consists of the membership and non-membership grades of arc (i, j) or the membership and non-membership grades of the paths from i to j in the respective intuitionistic fuzzy sets associated with the possible lengths, 1 through R. Hence,

$$e_{ij} = (\mu_{e_{ij}}, \gamma_{e_{ij}})$$

where $\mu_{e_{ij}} = (\mu_1(i,j), \mu_2(i,j), ...\mu_R(i,j))$ and $\gamma_{e_{ij}} = (\gamma_1(i,j), \gamma_2(i,j), ...\gamma_R(i,j))$. The operator $\tilde{+}$ represents the combinatorial sum and dom is the domination operator.

The combinatorial sum for fuzzy shortest paths is defined as follows. Recall that we assume there are M layers and the possible lengths of an edge are 1 through R. Therefore, the shortest a path could be in length is M - 1 and the longest a path could be in length is (M - 1). R. To find the paths of possible lengths, combinations of the possible lengths must be considered. To find the possible length I of a path, the lengths that can be used in combination are 1 , 2 , . . . ,I - M + 1. The combinatorial sum of two tuples is then defined as follows. Let Z $= \min\{\mu_x(j,k), \mu_y(k,q)\}$ for membership values and Z $= \max\{\gamma_x(j,k), \gamma_y(k,q)\}$. Then $e_{j,q} = e_{j,k} \tilde{+} e_{k,q}$ where the membership and non-membership grades of i-th element of the R-tuple $e_{j,q}$ is given by $m(e_{j,q})^i = \max_{x+y=i} (Z) = \mu_{\mathrm{i}}(j,q)$ and

$$nm(e_{j,q})^i = \min_{x+y=i} (Z) = \gamma_{\mathrm{i}}(j,q).$$

The recursive equation (1) will yield the set of non-dominated paths from source 1 to N. We then define

$$m(\tilde{P}_{1,N}) = \left\{1/\max_i(\mu_1^{\mathrm{i}}(1.\mathrm{N}), \cdots, \mathrm{K}/\max_i(\mu_{\mathrm{K}}^{\mathrm{i}}(1.\mathrm{N}), \cdots, \mathrm{R}/\max_i(\mu_{\mathrm{R}}^{\mathrm{i}}(1.\mathrm{N}))\right\}$$

and

$$nm(\tilde{P}_{1,N}) = \left\{1/\min_i(\gamma_1^{\mathrm{i}}(1.\mathrm{N}), \cdots, \mathrm{K}/\min_i(\gamma_{\mathrm{K}}^{\mathrm{i}}(1.\mathrm{N}), \cdots, \mathrm{R}/\min_i(\gamma_{\mathrm{R}}^{\mathrm{i}}(1.\mathrm{N}))\right\}$$

where $\mu_K^i(s,t)$ and $\gamma_K^i(s,t)$ represent the membership and non- membership grades in the intuitionistic fuzzy set K of the path from vertex s to vertex t given by the i-th non-dominated R-tuple. Now, intuitionistic fuzzy shortest path length, denoted by $P_{1,N}$, is equal to nondom $\{m(\tilde{P}_{1,N}), nm(\tilde{P}_{1,N})\}$.

5 Case Study

Consider the map of telecommunication department of Erode city in India. An IFG is formed with 5 layers (Fig. 2).

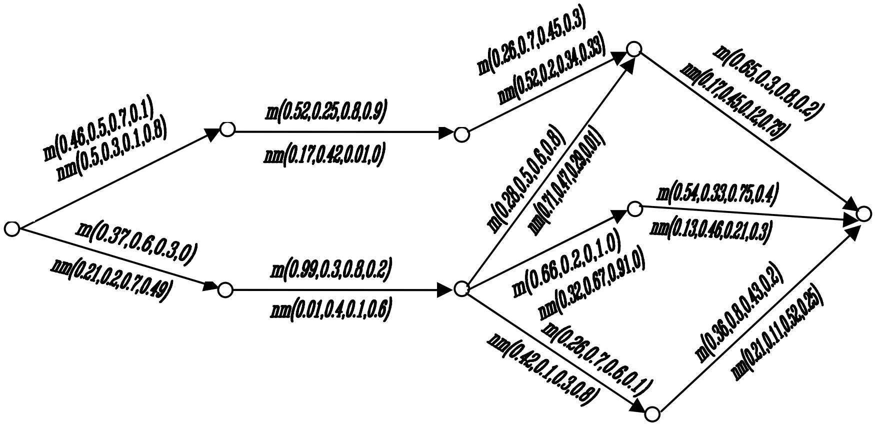


Fig. 2. A directed acyclic IFG

Assume that each edge can take a length of 1, 2, 3 or 4. The four tuple associated with each edge gives the membership value of the edge in each of the fuzzy sets 1, 2, 3and 4. In the graph there are 5 layers. Hence, M - 1 = 4 and the shortest possible path is 4 units and the most a path could be is 16 units since R = 4. If we apply the recursion (membership) given by equation (1), we obtain

f(9) = (1, 1, 1 . . . . . 1),
f(6) = (0.65, 0.3, 0.8, 0.2, 0, 0, …,0),
f(7) = (0.54, 0.33, 0.75, 0.4, 0, 0, …, 0),
f(8) = (0.36, 0.8, 0.43, 0.2, 0, 0,…, 0),
f(4) = $e_{46} \tilde{+} f(6)$ = {(0, 0.26, 0.65, 0.45, 0.7, 0.45, 0.3, 0.2, 0, 0,…,0),
f(5) = {0, 0.54, 0.5, 0.7, 0.65, 0.6, 0.8, 0.4, 0, 0,…,0}.
f(2) = {0, 0, 0.26, 0.52, 0.45, 0.65, 0.65, 0.7, 0.7, 0.45, 0.3, 0.2, 0, 0,…,0}.
f(1) = {0, 0, 0, 0.37, 0.54, 0.5, 0.6, 0.6, 0.65, 0.65, 0.7, 0.7, 0.45, 0.3, 0.2, 0.1}.

Similarly, for the non-membership function, we have

f(9) = (0, 0, 0,…, 0),
f(6) = (0.17, 0.45, 0.12, 0.73, 0, 0, …,0),
f(7) = (0.13, 0.46, 0.21, 0.3, 0, 0, …, 0),
f(8) = (0.21, 0.11, 0.52, 0.25, 0, 0,…, 0),
f(4) = {(0, 0.52, 0.2, 0.34, 0.2, 0.34, 0.33, 0.73,0, 0,…,0),
f(5) = {0, 0.32, 0.21, 0.11, 0.3, 0.25, 0.3, 0.8, 0, 0,…,0},
f(2) = {0, 0.17, 0.42, 0.01, 0, 0.2, 0.2, 0.2, 0.2, 0.17, 0.17, 0.01, 0.01, 0.01,0.01, 0.01},

f(3) ={0, 0.01, 0.32, 0.1, 0.11, 0.13, 0.11, 0.12,0.25, 0.01, 0.01, 0.01, 0.01,0.01, 0.01,0.01},
f(1) = {0, 0.21, 0.2, 0.2, 0.21, 0.2, 0.2, 0.2, 0.2, 0.2, 0.21, 0.2, 0.2, 0.2, 0.2, 0.2}.

The intuitionistic fuzzy shortest path length is then given by
$P_{1,9}$ = {1/0, 2/0, 3/0.2, 4/0.2, 5/0.21, 6/0.2, 7/0.2, 8/0.2, 9/0.2, 10/0.2, 11/0.21, 12/0.2, 13/0.2, 14/0.2, 15/0.2, 16/0.1}.

Therefore it is determined that the shortest possible path has length 4. This path corresponds to the path 1-2-4-6-9 in the original network and is formed by the general backtracking techniques of DP. Note that this model represents intuitionistic fuzzy shortest path. Therefore, it can be used as a decision tool due to the maintenance of the underlying structure. Also note that the procedure presented is a naive DP approach and can be easily made more efficient.

6 Conclusion

In this paper we have presented new formulation for intuitionistic fuzzy shortest path problems. This method is developed because the classical fuzzy shortest path problem yields a fuzzy length with no actual path associated with it [8]. The formulation presented circumvents this problem. Dynamic programming based method is developed to solve the new problems and a numerical example is given for better understanding. This type of analysis may have potential in developing new formulations for general fuzzy mathematical programming, or for analyzing current formulations.

Acknowledgments

The author Parvathi Rangasamy would like to thank University Grants Commission, NewDelhi, India for its financial support to the Minor Research Project MRP 686/05(UGC-SERO) dated Feb2005.

References

1. Adamo, J H.: Fuzzy decision trees. Fuzzy Sets and Systems. 4 (1980) 207-219
2. Atanassov, K.: Index matrix representation of intuitionistic fuzzy graphs. Fifth Scientific session of the Mathematical Foundations of Artificial Intelligence Seminar, Sofia. Oct.5 (1994) 36–41
3. Atanassov, K, A. Shannon.: On a generalization of intuitionistic fuzzy graphs. Notes on Intuitionistic Fuzzy Sets. 8(2002) 73-78
4. Atanassov, K.: Intuitionistic fuzzy Sets. Notes on Intuitionistic Fuzzy Sets. 4(1998) 59-61
5. Atanassov, K.: Intuitionistic fuzzy Sets. Springer Physica-Verlag, Berlin.(1999)
6. Bondy, J.A. Murty, U.S.:Graph Theory with Appliations. American Elsevier Publishing Co. Inc. NewYork (1997)
7. Dubois, D. Prade, H.: Fuzzy Sets and Systems. Academic Press, New York (1980)
8. Klein, C.M.: Fuzzy Shortest Paths. Fuzzy Sets and Systems. 39 (1991) 27-41

On Imprecision Intuitionistic Fuzzy Sets & OLAP – The Case for KNOLAP

Ermir Rogova and Panagiotis Chountas

Harrow School of Computer Science, Data & Knowledge Management Group,
University of Westminster, Watford Road, Northwick Park, HA5 3TP, London, UK
{rogovae,chountp}@wmin.ac.uk

Abstract. Traditional data repositories are typically focused on the storage and querying of crisp-precise domains of data. As a result, current commercial data repositories have no facilities for either storing or querying imprecise-approximate data. However, when considering scientific data (i.e. medical data, sensor data etc) value uncertainty is inherited to scientific measurements. In this paper we revise the context of "value uncertainty", and examine common models related to value uncertainty as part of the OLAP model. We present our approach for extending the OLAP model to include treatment of value uncertainty as part of a multidimensional model inhabited by flexible date and non-rigid hierarchical structures of organisation.

1 Introduction

In this paper we introduce the semantics of the Intuitionistic Fuzzy cubic representation in contrast to the basic multidimensional-cubic structures. The basic cubic operators are extended and enhanced with the aid of Intuitionistic Fuzzy Logic [1], [2].

Since the emergence of the OLAP technology [3] different proposals have been made to give support to different types of data and application purposes. One of this is to extend the relational model (ROLAP) to support the structures and operations typical of OLAP. Further approaches [4], [5] are based on extended relational systems to represent data-cubes and operate over them. The other approach is to develop new models using a multidimensional view of the data [6].

Nowadays, information and knowledge-based systems need to manage imprecision in the data and more flexible structures are needed to represent the analysis domain. New models have appeared to manage incomplete datacube [7], imprecision in the facts and the definition of fact using different levels in the dimensions [8].

Nevertheless, these models continue to use inflexible hierarchies thus making it difficult to merge reconcilable data from different sources with some incompatibilities in their schemata. These incompatibilities arise due to different perceptions-views about a particular modelling reality.

In addressing the problem of representing flexible hierarchies we propose a new multidimensional model that is able to treat with imprecision over conceptual hierarchies based on Intuitionistic Fuzzy logic. The use of conceptual hierarchies enables us to:

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 11–20, 2007.
springerlink.com

- define the structures of a dimension in a more perceptive way to the final user, thus allowing a more perceptive use of the system.
- query information from different sources or even use information or preferences given by experts to improve the description of hierarchies, thereby getting more knowledgeable query results. We outline a unique way for incorporating "kind of" relations, or conceptual imprecise hierarchies as part of a Knowledge based multidimensional analysis (KNOLAP).

2 Semantics of the IF-Cube in Contrast to Crisp Cube

In this section we review the semantics of Multidimensional modeling and Intuitionistic Fuzzy Logic and based on these we propose a unique concept named as Intuitionistic Fuzzy Cube (IF-Cube). The IF-Cube is the basis for the representation of flexible hierarchies and thus flexible facts.

2.1 Principles of Intuitionistic Fuzzy Logic

Each element of an Intuitionistic fuzzy [1], [2] set has degrees of membership or truth (μ) and non-membership or falsity (ν), which don't sum up to 1.0 thus leaving a degree of hesitation margin (π).

As opposed to the classical definition of a fuzzy set given by $A' = \{< x, \mu_{A'}(x) > | x \in X\}$ where $\mu_A(x) \in [0, 1]$ is the membership function of the fuzzy set A′, an intuitionistic fuzzy set A is given by:

$$A = \{< x, \mu_A(x), v_A(x) > | x \in X\}$$

where: $\mu_A : X \to [0, 1]$ and $v_A : X \to [0, 1]$ such that $0< \mu_A(x) + v_A(x)<1$ and $\mu_A(x)$ $v_A(x) \in [0, 1]$ denote a degree of membership and a degree of non-membership of $x \in A$, respectively.

Obviously, each fuzzy set may be represented by the following Intuitionistic fuzzy set

$$A=\{<x, \mu_A' (x), (x), 1- \mu_A' (x)>| x \in X\}$$

For each intuitionistic fuzzy set in X, we will call $\pi_A (x) = 1 - \mu_A(x) - v_A(x)$ an intuitionistic fuzzy index (or a hesitation margin) of $x \in A$ which expresses a lack of knowledge of whether x belongs to A or not. For each $x \in A$ $0<\pi_A (x)<1$.

2.2 Overview of the Cube Model

A logical model that influences the database design and the query engines is the *multidimensional-cubic* view of data in the warehouse. In a multidimensional data model, there is a set of *numeric measures* that are the objects of analysis. Examples of such measures are sales, budget, etc. Each of the numeric measures depends on a set of *dimensions,* which provide the context for the measure. The attributes of a dimension may be related via a hierarchy of relationships. In the above example, the product name is related to its category and the industry attribute through a hierarchical relationship, see "Fig.1".

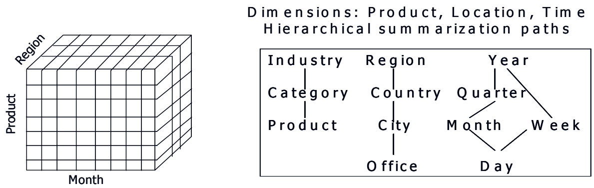


Fig. 1. "Cube "Sales" – Rigid Hierarchies for Product, Location Time Dimensions"

According to [6] a cube structure is defined as a 4-tuple, <D, M, A, f> where the four components indicate the characteristics of the cube. These characteristics are:

a set of n dimensions $D = \{d_1, d_2, \ldots, d_n\}$ where each d_i is a dimension name, extracted from a domain $dom_{dim(i)}$. A set of k measures $M = \{m_1, m_2, \ldots, m_k\}$ where each m_i is a measure name, extracted from a domain $dom_{measure(i)}$. The set of dimension names and measures names are disjoint; i.e., $D \cap M = 0$. A set of t attributes $A = \{a_1, a_2, \ldots, a_t\}$ where each a_i is an attribute name, extracted from a domain $dom_{attr(i)}$. A one-to-many mapping $f : D \rightarrow A$, i.e. there exists, corresponding to each dimension, a set of attributes.

2.3 Semantics of the IF-Cube

In contrast an **IF-Cube** is an abstract structure that serves as the foundation for the multidimensional data cube model. Cube C is defined as a five-tuple *(D, l, F, O, H)* where:

- *D* is a set of dimensions
- *l* is a set of levels $l_1,\ldots, l_n,$
- A dimension $d_i = (l \leq O, l_\perp, l_\top)$ $dom(d_i)$ where $l = l_i$ $i=1\ldots n$.
  l_i is a set of values and $l_i \cap l_j = \{\}$,
  $\leq O$ is a partial order between the elements of *l*.
  To identify the level *l* of a dimension, as part of a hierarchy we use *dl.*
  $l_\perp$: base level $l_\top$: top level
  for each pair of levels l_i and l_j we have the relation
  $\mu_{ij} : l_i \times l_j \rightarrow [0,1]$ $\quad v_{ij} : l_i \times l_j \rightarrow [0,1]$ $\quad 0 < \mu_{ij} + v_{ij} < 1$
- *F* is a set of fact instances with schema $F = \{<x, \mu_F(x) , v_F(x)>\mid x\in X \}$, where $x=<att_1, \ldots,att_n>$ is an ordered tuple belonging to a given universe *X*, $\mu_F(x)$ *and* $v_F(x)$ are the degree of membership and non-membership of *x* in the fact table *F* respectively.
- *H* is an object type history that corresponds to a cubic structure*(l, F, O, H′)* which allows us to trace back the evolution of a cubic structure after performing a set of operators i.e. aggregation.

The example below provides a sample imprecise cube (*D, l, F, O, H) i.e. sales* and a conceptual non-rigid hierarchy product with reference to milk consisting of $l_i,\ldots, l_n$ levels with respective levels of membership and non membership $< \mu_{ij}\ v_{ij}, >$.

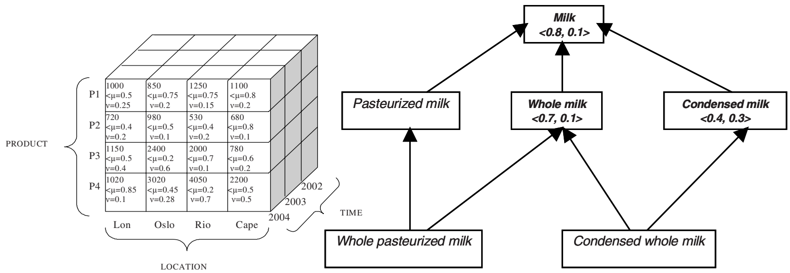


Fig. 2. “Imprecise Cube ‘Sales’ – Conceptual – Ontological, IF Hierarchy ‘Milk’ “

The defined IF OLAP Cube and the proposed OLAP operators allow us to:

- accommodate imprecise facts
- utilise *conceptual hierarchies* used for aggregation purposes in the cases of roll-up and roll down operations.
- offer a unique feature such as keeping track of the history when we move between different levels of a hierarchical order.

In the next section, the fundamental cubic operators are defined and explained with the aid of examples. The examples make use of cubic slices commonly known as fact tables. Each operator is presented in the following format: the operator’s name, symbol, textual description, input, output, mathematical description and an example of the operator.

3 Cubic Operators

Selection (Σ): The selection operator selects a set of fact-instances from a cubic structure that satisfy a predicate (θ). A predicate (θ) involves a set of atomic predicates (θ_1, …, θ_n) associated with the aid of logical operators *p (* i.e. ∧, ∨, etc.) . The set of possible facts (cubic instances) that satisfy the θ should carry a degree of membership μ and non-membership v expressed as

$$F = \{<x, min(\mu_F(x), \mu(\theta(x))), max(v_F(x), v(\theta(x)))> \mid x \in X\} \tag{1}$$

This guaranties a resulting cube populated with fact instances that satisfy the predicate (θ) either completely or to some degree of certainty.

Input: $C_i = (D, l, F, O, H)$ and the predicate θ
Output: $C_o = (D, l, F_o, O, H)$ where $F_o \subseteq F$ and $F_o = \{f \mid (f \in F) \wedge (f$ satisfies $\theta)$
Mathematical notation: $\sum_\theta (C_i) = C_o$

Example: Find the sales amount of 1000 with membership of greater than 0.4 and non membership of less than 0.3 for all products in all cities during 2004

$$\mathbf{\Sigma_{(amount>1000 \wedge (\mu>0.4 \wedge v<0.3) \wedge year=2004)}(Sales)=C_{Result}}$$

Cubic Product (⊗): This is a binary operator $C_{i1} \otimes C_{i2}$. It is used to relate two cubes C_{i1} and C_{i2} assuming that $D_1 \subseteq D_2$ and O_1 , O_2 are reconcilable partial orders. Thus, l_1, l_2 could lead to l_o being a ragged hierarchy.

Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$ where

$$D_o = D_1 \cup D_2,\quad l_o = l_1 \cup l_2,\ O_o = O_1 \cup O_2\quad H_o = H_1 \cup H_2,\ F_o = F_1 \text{ X } F_2$$

$$F_o = \{<<x, y>, \min(\mu_{f1}(x), \mu_{f2}(y)), \max(\nu_{f1}(x), \nu_{f2}(y),)> | <x, y> \in X \times Y\}$$

Mathematical notation: $C_{i1} \otimes C_{i2} = C_o$
Example: Consider the two cubes we want to relate, C_{i1}: C_{Sales} and C_{i2}: $C_{Discounts}$. $C_{Discounts}$ has the same dimensions as C_{Sales} except the measure amount is not sale but is a discount. In that case the cubic product would be:

$$C_{Sales'} \otimes C_{Discounts} = C_{Result}$$

ProdID	StoreID	Amount	<μ, ν>
P1	S1	10	.7, .2
P2	S2	15	.5, .5

⊗

ProdID	StoreID	Discount	<μ, ν>
P2	S1	2	.5, .5
P3	S3	5	.3, .3

=

S.ProdID	S.StoreID	S.Amount	D.ProdID	D.StoreID	D.Discount	<μ, ν>
P1	S1	10	P2	S1	2	.5, .5
P1	S1	10	P3	S3	5	.3, .3
P2	S2	15	P2	S1	2	.5, .5
P2	S2	15	P3	S3	5	.3, .5

Fig. 3. Fact-Sales', Fact-Discounts and Fact-Result

Join (Θ): It can be expressed using Cubic Product operator. $C_{i1} = (D_1, l_1, F_1, O_1, H_1))$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$ are candidates to join if $D_1 \cap D_2 \neq 0$,
Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$

Mathematical notation: $C_{i1}\ \Theta\ C_{i2} = \sigma_p(C_{i1} \otimes C_{i2})$

Union (∪): The union operator is a binary operator that finds the union of two cubes. C_{i1} and C_{i2} have to be union compatible. The operator also coalesces the value-equivalent facts using the minimum membership and maximum non-membership.
Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$ where $D_o = D_1 = D_2$, $l_o = l_1 = l_2$, $O_o = O_1 = O_2$, $H_o = H_1 = H_2$, $F_o = F_1 \cup F_2 = \{ < x, \max(\mu F_1(x), \mu F_2(x)), \min(\nu F_1(x), \nu F_2(x)) > | x \in X \}$
Mathematical notation: $C_{i1} \cup C_{i2} = C_o$
Example: Consider the two cubes we want to relate, C_{i1}: C_{Sales_North} and C_{i2}: C_{Sales_South}, in that case the union of these two cubes would be:

$$C_{Sales\_North} \cup C_{Sales\_South} = C_{Result}$$

ProdID	StoreID	Amount	<μ, ν>
P1	S1	10	.7, .2
P2	S2	15	.5, .5

∪

ProdID	StoreID	Amount	<μ, ν>
P1	S1	10	.5, .5
P3	S3	5	.3, .3

=

S.ProdID	S.StoreID	S.Amount	<μ, ν>
P1	S1	10	.7, .2
P2	S2	15	.5, .5
P3	S3	5	.3, .3

Fig. 4. Fact-Sales_North, Fact-Sales_South and Fact-Result

Difference (-):. The difference operator removes the portion of the cube C_{i1} that is common to both cubes. C_{i1} and C_{i2} have to be union compatible

Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$ where $D_o=D_1=D_2$, $l_o=l_1=l_2$, $O_o=O_1=O_2$, $H_o=H_1=H_2$,
$F_o = F_1 \cap F_2 = \{ < x, \min(\mu F_1(x),\mu F_2(x)), \max(\nu F_1(x),\nu F_2(x)) > \mid x \in X \}$
Mathematical notation: $C_{i1} - C_{i2} = C_o$
Example: Consider the two cubes we want to relate, C_{i1}: C_{Sales_North} and C_{i2}: C_{Sales_South}, in that case the difference between North and South sale cubes would be:

$$C_{Sales\_North} - C_{Sales\_South} = C_{Result}$$

ProdID	StoreID	Amount	<μ, ν>
P1	S1	10	.7, .2
P2	S2	15	.5, .5

-

ProdID	StoreID	Amount	<μ, ν>
P1	S1	10	.5, .5
P3	S3	5	.3, .3

=

S.ProdID	S.StoreID	S.Amount	<μ, ν>
P1	S1	10	.5, .5
P2	S2	15	.5, .5

Fig. 5. Fact-Sales_North, Fact-Sales_South and Fact-Result

3.1 Extended Operators

Aggregation (A): An aggregation operator *A* is a function *A(G)* where $G = \{<x, \mu_F(x), \nu_F(x)> \mid x \in X\}$ where $x=<att_1, \ldots, att_n>$ is an ordered tuple belonging to a given universe *X*, $\{att_1, \ldots, att_n\}$ is the set of attributes of the elements of *X*, $\mu_F(x)$ *and* $\nu_F(x)$ are the degree of membership and non-membership of *x*. The result is a bag of the type $\{<x', \mu_F(x'), \nu_F(x')> \mid x' \in X\}$. To this extent, the bag is a group of elements that can be duplicated and each one has a degree of μ and ν.

Input: $C_i = (D, l, F, O, H)$ and the function *A(G)*
Output: $C_o = (D, l_o, F_o, O_o, H_o)$

The definition of the extended group operators allows us to define the extended group operators ***Roll up (Δ), and Roll Down (Ω).***

Roll up (Δ): The result of applying Roll up over dimension d_i at level dl_r using the aggregation operator A over a datacube $C_i = (D_i, l_i, F_i, O, H_i)$ is another datacube

$C_o = (D_o, l_o, F_o, O, H_o)$.
Input: $C_i = (D_i, l_i, F_i, O, H_i)$
Output: $C_o = (D_o, l_o, F_o, O, H_o)$

An object of type history is a recursive structure *H* =
- ω *is the initial state of the cube*
- *(l, D, A, H')* is the state of the cube after performing an operation on the cube

The structured history of the datacube allows us to keep all the information when applying *Roll up* and get it all back when *Roll Down* is performed. To be able to apply the operation of *Roll Up* we need to make use of the IF_{SUM} aggregation operator.

Roll Down (Ω): This operator performs the opposite function of the *Roll Up* operator. It is used to roll down from the higher levels of the hierarchy with a greater degree of generalization, to the leaves with the greater degree of precision. The result of applying *Roll Down* over a datacube $C_i = (D, l, F, O, H)$ having $H=(l', D', A', H')$ is another datacube $C_o = (D', l', F', O, H')$.

Input: $C_i=(D, l, F, O, H)$
Output: $C_o=(D',l', F', O, H')$ where *F'* → *set of fact instances defined by operator A.*

To this extent, the *Roll Down* operative makes use of the recursive history structure previously created after performing the *Roll Up* operator.

The definition of aggregation operator points to the need of defining the IF extensions for traditional group operators [9], such as *SUM, AVG, MIN and MAX.* Based on the standard group operators, we provide their IF extensions and meaning.

***IF_{SUM}* :** The IF_{sum} aggregate, like its standard counterpart, is only defined for numeric domains. Given a fact F defined on the schema X $(att_1, \ldots, att_n)$, let att_{n-1} defined on the domain $U=\{u_1, \ldots, u_n\}$. The fact F consists of fact instances F_i *with* $1 \le i \le m$. The fact instances F_i are assumed to take Intuitionistic Fuzzy values for the attribute att_{n-1} for $i = 1$ *to* m we have $F_i[att_{n-1}] = \{<\mu_i(u_{ki}), \nu_i(u_{ki})>/\, u_{ki} \mid 1 \le k_i \le n\}$. The IF_{sum} of the attribute att_{n-1} of the fact table F is defined by:

$$IF_{SUM}((att_{n-1})(F)) =$$

$$\{<u>/\, y \mid ((\, u = \min_{i=1}^{m} (\mu_i(u_{ki}), \nu_i(u_{ki})) \wedge (y = \textstyle\sum_{ki=k1}^{km} u_{ki}\,)\,(\,\forall_{k1, \ldots km} : 1 \le k1, \ldots km \le n))\}$$

Example: $IF_{SUM}((Amount)(ProdID))$
={<.8,.1>/10}+{(<.4,.2>/11),(<.3,.2>/12)}+{(<.5,.3>/13),(<.5,.1>/12)}
={(<.8∧.4,.1∧.2>/10+11),(<.8∧.3, .1∧.2>/10+12)}+{<.5, .3>/13, <.5, .1>/12}
={(<.4, .2>/21), (<.3, .2>/22)} + {<.5, .3>/13, <.5, .1>/12}
={(<.4∧.5,.2∧.3>/21+13),(<.4∧.5,.2∧.1>/21+12),(<.3∧.5,.2∧.3>/22+13),(<.3∧.5, .2∧.1>/22+12) = {(<.4, .3>/34), (<.4, .2>/33), (<.3, .3>/35), (<.3, .2>/34)}
={(<.3, .3>/34), (<.4, .2>/33), (<.3, .3>/35)}

***IF_{AVG}* :** The IF_{AVG} aggregate, like its standard counterpart, is only defined for numeric domains. This aggregate makes use of the IF_{SUM} that was discussed previously and the standard *COUNT.* The IF_{AVG} can be defined as:

$$IF_{AVG}((att_{n-1})(F) = IF_{SUM}((att_{n-1})(F)) \,/\, COUNT((att_{n-1})(F))$$

Example: $IF_{AVG}((Amount)(ProdID))$
= $IF_{SUM}((Amount)(ProdID)) / COUNT((Amount)(ProdID))$
= {(<.3, .3>/34), (<.4, .2>/33), (<.3, .3>/35)} / 3
= {(<.3, .3>/11.33), (<.4, .2>/11), (<.3, .3>/11.66)}

***IF_{MAX}* :** The IF_{MAX} aggregate, like its standard counterpart, is only defined for numeric domains. The IF_{sum} of the attribute att_{n-1} of the fact table F is defined by:

$$IF_{MAX}((att_{n-1})(F)) =$$

$$\{<u>y \mid ((u = \min_{i=1}^{m} (\mu_i(u_{ki}), \nu_i(u_{ki})) \wedge (y = \max_{i=1}^{m} (\mu_i(u_{ki}), \nu_i(u_{ki})))(\,\forall_{k1,\ldots km} : 1 \le k1, \ldots km \le n))\}$$

Example: $IF_{MAX}((Amount)(ProdID))$
IF_{MAX}={<.8,.1>/10},{(<.4,.2>/11),(<.3,.2>/12)},{(<.5,.3>/13),(<.5, 0.1>/120)}
={(<.8∧.4,.1∧.2>/max(10,11)),(<.8∧.3,.1∧.2>/max(10,12)},{<.5,.3>/13,<.5, .1>/12}
={(<.4,.2>/11),(<.3,.2>/12)},{<.5,.3>/13, <.5, .1>/12} ={(<.4∧.5,.2∧.3>/max(11,13)),
(<.4∧.5,.2∧.1>/max(11,12)),(<.3∧.5,.2∧.3>/max(12,13)),(<.3∧.5,.2∧.1>/max(12,12))
={(<.4,.3>/13),(<.4,.2>/12),(<.3,.3>/13), (<.3,.2>/12)} = **{(<.3,.3>/13),(<.3,.2>/12)}**

IF_{MIN} **:** The IF_{MIN} aggregate, like its standard counterpart, is only defined for numeric domains. Given a fact F defined on the schema X $(att_1, \ldots, att_n)$, let att_{n-1} defined on the domain $U=\{u_1, \ldots, u_n)$. The fact F consists of fact instances f_i *with* $1 \le i \le m$. The fact instances f_i are assumed to take Intuitionistic Fuzzy values for the attribute att_{n-1} for $i = 1$ *to* m we have $f_i[att_{n-1}] = \{<\mu_i(u_{ki}), \nu_i(u_{ki})>/ u_{ki} \mid 1 \le k_i \le n\}$. The IF_{sum} of the attribute att_{n-1} of the fact table F is defined by:

$$IF_{MIN}((att_{n-1})(F)) =$$
$$\{<u>/ y | ((u = \min_{i=1}^{m} (\mu_i(u_{ki}),\nu_i(u_{ki})) \wedge (y = \min_{i=1}^{m} (\mu_i(u_{ki}),\nu_i(u_{ki})))(\forall_{k1,\ldots km} : 1 \le k1,\ldots km \le n))\}$$

We can observe that the IF_{MIN} is extended in the same manner as IF_{MAX} aggregate except for replacing the symbol **max** in the IF_{MAX} definition with **min**. Once we have defined our Intuitionistic Fuzzy multidimensional model and have defined the IF cubic-algebra, the concept of knowledge based OLAP is introduced. Ideally, in a Knowledge based OLAP environment for summarizing purposes it is desirable to use Intuitionistic Fuzzy hierarchies like milk see “Fig.2” instead of rigid hierarchies like Product in “Fig.1”.

4 The Case for Knowledge Based OLAP-KNOLAP

Concepts are used to describe how the data is organized in the data sources and to map such data to the concepts described in the Domain Ontology. These definitions are used to apply more extensively the business semantics described in the Domain Ontology, to support the rewrite of queries’ conditions and to combine OLAP features in this process. These semantics support the automatic recommendation of analysis according to the context of users’ explorations in order to guide the decision making, feature inexistent in current analytical tools.

With respect to the Intuitionistic Fuzzy hierarchy milk, we try to express different ontological semantics, or “kind of” relations such as to what extent:

- Condensed whole milk is a “kind-of” Whole milk?
- Condensed whole milk is a “kind-of” Condensed milk?
- Pasteurised whole milk is a “kind-of” Whole milk?
- Pasteurised whole milk is a “kind-of” Pasteurised milk?
- Pasteurised milk is ”kind-of” milk? Etc.

It is obvious from the above examples that if we wish to summarise the sales, for example, of products of “Pasteurised milk” we need to take into account as well the fact that “Whole Pasteurised milk” may also be treated as “Pasteurised milk” when applying i.e. the ***IF_{SUM}.***

These observations led us to introduce the concept of closure of an Intuitionistic fuzzy set over a universe that has a hierarchical structure, which is a developed form defined on the whole hierarchy. Intuitively, in the closure of this Intuitionistic fuzzy set, the "kind of" relation is taken into account by propagating the degree associated with an element to its sub-elements more specific elements in the hierarchy. For instance, in a query, if the user is interested in the element Milk, we consider that all kinds of Milk, Whole milk, Pasteurized milk, etc. are of interest. On the opposite, we consider that the super-elements (more general elements) of Milk in the hierarchy i.e. "Milk" are too general to be relevant for the user's query.

Let us consider the Intuitionistic fuzzy set M defined as: {Milk<0.8,0.1>, Whole-Milk<0.7,0.1>, Condensed-Milk<0.4,0.3>}} which is presented in "Fig.6". Then the next step is to calculate the <μ, ν> values for "Pasteurized milk", "Whole Pasteurized milk " and "Condensed whole milk."

- If the hierarchical IF structure expresses preferences in a query, the choice of the maximum values for μ and minimum value ν from the pairs of values <μ, ν> from the parent elements to the sub elements allows us not to exclude any possible answer (high possibility necessity degrees). In real cases, the lack of answers to a query generally makes this choice preferable, because it consists of widening the query answer rather than restricting it.
- If the hierarchical IF represents an ill-known concept, the choice of the maximum value for μ and minimum value ν allows us to preserve all the possible values, but it also makes the answer less specific. In a way, it also participates in enlarging the query, as a less specific datum may share more common values with the query (the possibility degree of matching can thus be higher, although the necessity degree can decrease).

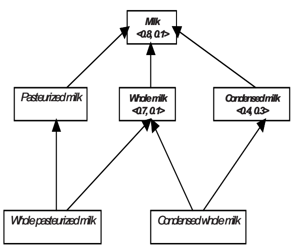


Fig. 6. "IF Hierarchy 'Milk' "

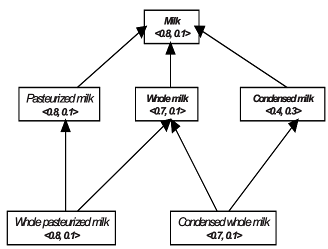


Fig. 7. "Fully weighted Hierarchy 'Milk' "

"Fig.7" is a fully weighted Hierarchy after applying the maximum values for μ and minimum value ν from the pairs of values <μ, ν> from the parent elements to the sub elements, i.e. from (whole-milk, condensed-milk) to (condensed-whole-milk), from (milk) to (pasteurized milk), and from (whole-milk, pasteurized milk) to (pasteurized-whole-milk).

The complete study of the hierarchical IF requires the formal definition of the IF hierarchical closure. We will further need to formally define the containment of an IF hierarchical set to another.

Furthermore if one wishes to consider multiple versions of evolving IF hierarchies, the similarity between different versions of IF hierarchical set in the geometrical framework introduced by [10], [11] needs to be examined as well.

5 Conclusions

In this paper we have presented a new multidimensional-cubic model named as the IF-Cube. The main contribution of this new model is that is able to operate over data with imprecision in the facts and the summarisation hierarchies. Classical models imposed a rigid structure that made the models present difficulties when merging information from different but still reconcilable sources. We introduce the automatic recommendation of analysis according to the context of users' explorations in order to guide the decision making with the aid of Intuitionistic fuzzy set over a universe that has a hierarchical structure and the corresponding hierarchies.

These features are inexistent in current OLAP tools. Furthermore we notice that our IF cube can be used for the representation of Intuitionistic fuzzy linguistic terms.

There is a need to formally define the closure of Intuitionistic fuzzy set over a universe that has a hierarchical structure as well the containment between different versions of these sets. We also need to study the impact of imprecision with respect to star and snowflake data warehouse conceptual structures. Finally, a graphical way needs to be developed to represent the results of the operations in order to get a more intuitive way to read the information obtained.

References

[1] K., Atanassov (1999). *Intuitionistic Fuzzy Sets*, Springer-Verlag, Heidelberg

[2] K., Atanassov *Intuitionistic Fuzzy Sets,* Fuzzy Sets and Systems, 20, 87–96, (1986)

[3] R. Kimball, *The Data Warehouse Toolkit.* New York: John Wiley & Sons, 1996.

[4] S., Chaudhuri U. Dayal V., Ganti. Database Technology for Decision Support Systems. In: Computer, Vol. 34, p. 48-55, 2001

[5] M., Jarke et al. Fundamentals of data warehouses. Springer, London, 2002

[6] H. Thomas & A., Datta, A Conceptual Model and Algebra for On-Line Analytical Processing in Decision Support Databases. Information Systems Research 12: 83-102, 2001

[7] C. Dyreson, Information retrieval from an incomplete data cube, VLDB, Morgan Kaufman Publishers, pp. 532-543, 1996.

[8] T. Pedersen, C. Jensen, and C. Dyreson, A foundation for capturing and querying complex multidimensional data, Information Systems, vol. 26, pp. 383-483, 2001.

[9] E. Rundensteiner L. Bic. Aggregates in posibilistic databases, VLDB'89, pp. 287-295, 1989

[10] E., Szmidt & J. Kacprzyk Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, Vol. 114, No.3,505–518, 2000

[11] E., Szmidt & J. Kacprzyk A new concept of similarity measure for Intuitionistic fuzzy sets and its use in group decision making LNAI 3558, pp272–282, Springer, 2005

Part II

The Application of Fuzzy Logic and Soft Computing in Flexible Quering

Algorithm for Interpretation of Multi-valued Taxonomic Attributes in Similarity-Based Fuzzy Databases

M. Shahriar Hossain and Rafal A. Angryk

Department of Computer Science, Montana State University,
Bozeman, MT 59717-3880, USA
{mshossain, angryk}@cs.montana.edu

Abstract. In this work we are analyzing our ability to discover knowledge from multi-valued attributes (often referred in literature on fuzzy databases as collections [1-3]), that have been utilized in fuzzy relational database models [4-7] as a convenient way to represent uncertainty about the data recorded in the data tables. We present here implementation details and extended tests of a heuristic algorithm, which we used in the past [8-11] to interpret non-atomic values stored in fuzzy relational databases. In our evaluation we consider different data imprecision levels, as well as diverse shapes of fuzzy similarity hierarchies.

Keywords: fuzzy relational databases, fuzzy collections, non-atomic symbolic values, data mining.

1 Introduction

The dynamic growth of data mining research and applications [12] carries significant importance for researchers working in areas related to fuzzy databases. Successful future of unconventional database models becomes more and more dependent on the development of consistent and time-efficient mechanisms for mining the data such databases are capable to store and process.

Fuzzy databases let the user reflect uncertainty about inserted information via the insertion of multiple descriptors in every column of the data table. At the same time, majority of currently available data mining algorithms allow the user to deal only with atomic attribute values. This leads to the challenge of developing a mechanism allowing consistent interpretation of the non-atomic values stored in the fuzzy databases. This process can be interpreted as mapping (i.e. defuzzification) of fuzzy tuples (containing collections of values) from tables in fuzzy relational databases into sets of atomic records, which are compatible with 1NF definition and can be analyzed using regular data mining techniques. We need to explain, that the term "defuzzification" used in this paper should be interpreted in a broader context than typically used in the fuzzy sets' community (i.e. transformation of a fuzzy set to a single, crisp value). In this work we will focus on problem of interpretation of multi-valued, nonnumeric entries in the similarity-based fuzzy relational databases. We have decided to choose this topic for our research as these types of entries appear in large number of applications, starting from multiple online surveys about customers' preferences (e.g. *mark*

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 23–35, 2007.
springerlink.com

types of food you like), and finishing on reports from medical examinations (e.g. *identify areas of pain, describe its severity, etc.*). In this paper we will present a heuristics allowing to transfer non-atomic symbolic values to the singleton forms, that can be then interpreted by many regular data mining algorithms. At the very beginning of this article, however, we want to emphasize that the point of this paper is not to argue about accuracy or efficiency of the approach that has been used, but rather to raise awareness of the problem, and to show that a defuzzification of uncertain data in similarity-based fuzzy databases is achievable. In the next section, we provide a brief review of the necessary background that covers the following areas: (1) fuzzy database model incorporating usage of non-atomic values to reflect uncertainty, (2) taxonomic symbolic variables and some of their properties, (3) our heuristics for interpretation of non-atomic entries in the fuzzy tuples. In the section 3, we present implementation details of our approach, discuss the artificial data sets we created for testing, and finally discuss results we obtained. Finally, in section 4, we briefly summarize conclusions coming from our investigation and point out new directions of future research.

2 Background

2.1 Fuzzy Database Model

There are two fundamental properties of fuzzy relational databases, proposed originally by Buckles and Petry [4-5] and extended further by Shenoi and Melton [6-7]: (1) utilization of non-atomic attribute values to characterize features of recorded entities we are not sure of, and (2) ability of processing the data based on the domain-specific and expert-specified fuzzy relations applied in place of traditional equivalence relations.

In fuzzy database model it is assumed that each of the attributes has its own fuzzy similarity table, which contains degrees of similarity between all values occurring for the particular attribute. Such tables can be also represented in the form of the similarity hierarchies, named by Zadeh [13] *partition trees* that show how the particular values merge together as we decrease similarity level, denoted usually by α. Examples of partition trees for the domains *Country* and *Food-Type* are presented in the figures 1 and 2, respectively.

2.2 Taxonomic Symbolic Variables and Their Analysis

In [14], Bock and Diday included a comprehensive collection of recent works related to the extraction of statistical information from symbolic data. Based on the classification presented in the book, the data type discussed in this paper should be characterized as *generalized taxonomic symbolic variables*. These types of data entries are expected to carry no quantitative meaning, but yet the values can be nonlinearly ordered in a form of rooted, hierarchical tree, called by Diday [14] a *taxonomy*. Such variables are called *taxonomic* or *tree-structured* variables. Zadeh's partition trees [13], as presented in the figures 1 and 2, represent such taxonomies, allowing us to incorporate fuzzy similarity relations not only to data querying, as it has been used in the past, but also to the data analysis.

In [14], chapter six [15] has been devoted almost entirely to the problem of derivation of basic description statistic (in particular: medians, modes and histograms) from multi-valued symbolic data. The authors propose transformation of non-atomic values to collection of pairs in the format of (ξ , φ), where ξ stands for a singleton (i.e. atomic) symbolic value, and φ represents ξ's observed frequency. This is achieved by extending the classical definition of frequency distribution. For a multi-valued variable, the new definition states that the number reflecting the *observed frequency* (i.e. count of ξ's appearances in the data set) can be a positive real, instead of a positive integer, as is the case for each single-valued variable.

The observed frequency distribution of a multi-valued variable *Z*, for a single data point, can be now defined as the list of all values in the finite domain of *Z*, together with the percentage of instances of each of the values in this data point. In [15], the authors mention that other definition of frequency distributions can be proposed, suggesting taking into account the natural dependencies between the symbolic values, as reflected by the provided attribute's taxonomy. This observation provided us with motivation for the work presented below.

2.3 Similarity-Driven Vote Distribution Method for Interpretation of Non-atomic Values

Following the rationale presented in the previous section, we attempted to develop a simple method to transfer non-atomic values in the fuzzy records to the collection of pairs including atomic descriptors and their fractional *observed frequencies*. In our wok, however, we wanted to utilize background knowledge about attributes' values, which is stored in fuzzy databases in form of fuzzy similarity relations [13]. In this section we present a simple example to introduce our approach.

We want a reader to assume for a moment that he/she needs to find a drugs' dealer who, as a not-confirmed report says (i.e. our fuzzy tuple), was recently seen in *{Canada, Colombia, Venezuela}*. The most trivial solution would be to split the count of observed frequency equally among all inserted descriptors, that is to interpret the entry as the following collection *{Canada|0.333, Colombia|0.333, Venezuela|0.333}*. This approach however does not take into consideration real life dependencies, which are reflected not only in the number of inserted descriptors, but also in their similarity (represented by a taxonomy of attribute values, which reflects our pre-defined fuzzy similarity relation/table).

In our work we used a simple heuristics [8-11] letting us to replace the even distribution of a vote with a nonlinear spread, dependent both on the similarity of inserted values and on their quantity. Using the partition tree built from the fuzzy similarity table (grey structure in Fig. 1), we can extract from the set of the originally inserted values those concepts which are more similar to each other than to the remaining values. We call them *subsets of resemblances* (e.g. *{Colombia, Venezuela}* from the above example). Then we use them as a basis for calculating a distribution of a database record's fractions. An important aspect of this approach is extraction of the *subsets of resemblances* at the lowest possible level of their common occurrence, since the nested character of fuzzy similarity relation guarantees that above this α-level they are going to co-occur regularly.

Our heursitc algorithm is pretty straightforward. Given (1) a set of values inserted as a description of particular entity's attribute, and (2) a hierarchical structure reflecting Zadeh's partition tree [13] for the attribute; we want to extract a table, which includes (a) the list of all subsets of resemblances from the given set of descriptors, and (b) the highest level of α-proximity of their common occurrence. We then use the list to fairly distribute fractions of the original fuzzy database record.

Our algorithm uses preorder recursive traversal for searching the partition tree. If any subset of the given set of descriptors occurs at the particular node of the concept hierarchy we store the values that were recognized as similar, and the corresponding value of α. An example of such a search for subsets of resemblances in a tuple with the values *{Canada, Colombia, Venezuela}* is depicted in Fig. 1. Numbers on the links in the tree represent the order in which the particular subsets of similarities were extracted.

After extracting the *subsets of resemblances*, we apply a summarization of α values as a measure reflecting both the frequency of occurrence of the particular attribute values in the *subsets of resemblances*, as well as the abstraction level of these occurrences. Since during the search the country *Canada* was reported only twice, we assigned it a grade *1.4* (i.e. *1.0+0.4*). For *Colombia* we get: *Colombia|(1.0 + 0.8 + 0.4) = Colombia|2.2*, and for the last value: *Venezuela|(1.0 + 0.8 + 0.4) = Venezuela|2.2*.

At the very end we normalize grades assigned to each of the entered values: *Canada |(1.4/5.8) = Canada |0.24, Colombia |(2.2/5.8) = Colombia |0.38,Venezuela |(2.2/5.8) = Venezuela |0.38*. This leads to the new distribution of the record's fractions, which, in our opinion, more accurately reflects real life dependencies than a linear-split approach.

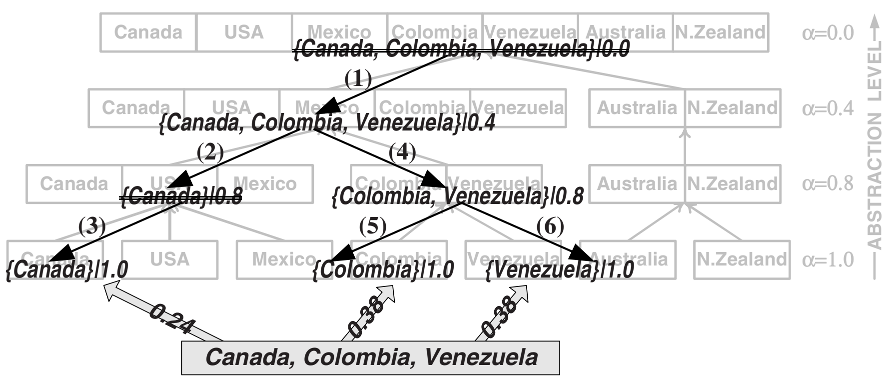


Fig. 1. Subsets of Resemblances extracted from the Partition tree

3 Implementation of the Defuzzification Algorithm

The implementation of our defuzzification system is based on a class named `TreeNode`, which contains only a single node of the fuzzy similarity hierarchy, pointers to its immediate descendants, and some public methods to access them. Each

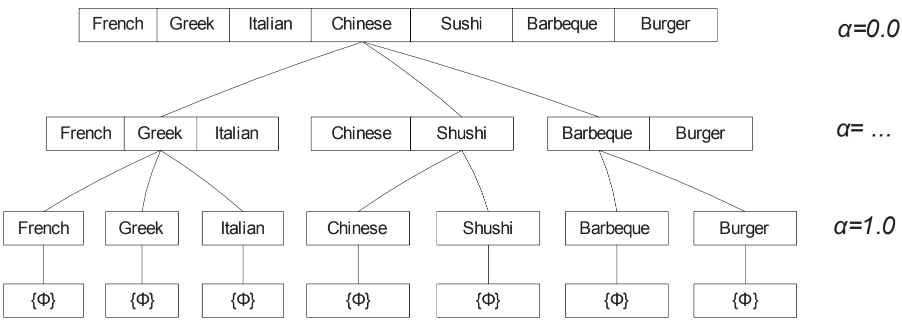


Fig. 2. Object representation of a partition tree

descendant is an instance of the same class `TreeNode`. Hence, a `TreeNode` contains pointers to descendent `TreeNode`. Descendents that do not point to any other child node basically points to null `TreeNode` (*{Φ}*). This is depicted in Fig. 2.

The `TreeNode` class is given below for better illustration.

```
public class TreeNode {
  int childCount = 0;
  Vector nodeData = new Vector();
  Vector childrenNodes = new Vector();
  /* THIS IS NOT A BINARY TREE. SO IT CAN CONTAIN ANY NUMBER
OF          CHILDREN. childrenNodes CONTAINS ELEMENTS OF TYPE
TreeNode */
  public TreeNode (Vector colon) {
    /* CREATES A NEW INSTANCE OF TreeNode*/
    nodeData = (Vector) colon.clone();
    childrenNodes = new Vector();   }
  public void addChild(TreeNode childNode){
    childrenNodes.add(childNode);
    childCount++;   }
  public int getTotalChildren(){
    /* e.g., IN THE CASE OF A NODE WITH THREE DESCENDENTS,
THIS FUNCTION        WOULD RETURN 3.*/
    return childCount;    }
  public Vector getNodeVector (){
    /* RETURNS THE PARENT NODE */
    return nodeData;    }
  public void setNodeAsLeaf (){
    childrenNodes.clear();      }
  public Vector getChildrenVector (){
    /* RETURNS THE CHILDREN NODES AS A VECTOR*/
    return childrenNodes;    }
  public TreeNode getChild (int i){
    /* RETURNS i-th CHILD WHERE EACH CHILD ITSELF IS A
TreeNode*/
    return ( (TreeNode) childrenNodes.get(i) );    }
}//END OF public class TreeNode
```

The basic algorithm uses preorder recursive tree traversal (depth-first search, DFS) for searching matching subsets in the partition tree. Our goal is to find at each node of the partition tree the largest matching subset. The high level outline of the algorithm is portrayed in the following code. `List` and `TreeNode` parameters of the `partitionTreeTraversal` function are passed by reference where other parameters are passed by value. The method `clone()` of an object returns entirely a new instance of the cloned object to avoid changes to a referenced parameter while the referenced parameter is necessary to be kept intact for the upper levels of recursion. A `Vector` can store a collection of entered attribute values, e.g. *{French, Greek, Italian}* and it can be used as an instance of `searchVector` of the algorithm. Besides, a `Vector` can also be a collection of other `Vectors`. On the other hand, a `List` is a collection of `Vectors` e.g., during the first call of `partitionTreeTraversal` the parameter `sList` should contain a sorted (descending order, based on the size of subsets) list of all the possible subsets of `searchVector` except the empty subset (*{Φ}*). The operation denoted by "–" in the algorithm below is considered a regular *SetDifference* operation.

```
public Vector partitionTreeTraversal(List sList, TreeNode
masterNode,                                              Vector
searchVector, int level){
  List subsetList = sList.clone();
  int totalChildren = masterNode.getTotalChildren();
  Vector masterVector = masterNode.getNodeVector();
  Vector subsetVector = GET THE LONGEST SUBSET FROM subset-
List THAT                          IS FOUND IN THE ROOT OF
masterNode;

  if (subsetVector.size()==0){
    return {Φ};
  }
  else if (subsetVector.size()==searchVector.size()){
    UPDATE THE CORRESPONDING ENTRY FOR EACH ELEMENT OF
searchVector or      subsetVector WITH CORRESPONDING α–VALUE
OF level–TH LEVEL ;
  }
  else{ //IF PARTIALLY AVAILABLE
    ADD CORRESPONDING α-VALUE OF level–TH LEVEL TO THE
CORRESPONDING      ENTRY FOR EACH ENTITY OF subsetVector ;
  }

  Vector resultVector = searchVector.clone();
  for (int i=0; i<totalChildren; i++){
    Vector temp =
    partitionTreeTraversal( subsetList, master-
Node.getChild(i),                              subsetVector,
level+1 );
    subsetList = subsetList – (ALL SUBSETS THAT HAVE AT LEAST
ONE                                 ENTITY OF temp);
    resultVector = resultVector – temp;
    if (resultVector == {Φ} )      //NOTE:[1]
```

[1] Do not traverse other branches because all entities are already found in the previously traversed branches.

```
        break ;
    }
    if (totalChildren==0)     /*IF masterNode IS A LEAF*/
      return subsetVector;
  } //END OF public void partitionTreeTraversal
```

3.1 Analysis of the Algorithm

Let us assume that the average branching factor of the partition tree is *b* and the number of abstraction levels in the tree is *d*. If each node has *b* descendents, then the root (*level 1*) has 1 node, *level 2* has *b* nodes, *level 3* has b^{3-1} nodes, …., *level d* has b^{d-1} nodes. Hence the total number of nodes, *N* in the tree is: $N = \sum_{k=1}^{d} b^{k-1}$, leading to $Nb = \sum_{k=1}^{d} b^{k}$, where *k* reflects the current *level* of the `partitionTreeTraversal` algorithm. The last equation produces the following result: $Nb - N = N(b-1) = b^d - 1$, thus $N = \frac{b^d - 1}{b-1}$. Therefore, in the worst case when the `searchVector` is dispersed in all the leaf nodes of the partition tree at d^{th} level, `partitionTreeTraversal` has to visit a total of $\frac{b^d - 1}{b-1}$ nodes which would result in a time complexity of $O(b^d)$ which should not be alarming despite of exponential growth, as typically $b \ll N$ and $d \ll N$ in real-life data mining. The worst case search scenario occurs when the imprecise attribute of the fuzzy record, stored in `searchVector`, contains every element of the root node of the partition tree; in our example – if the `searchVector` is *{French, Greek, Italian, Chinese, Sushi, Barbeque, Burger}*. In such case the traversal would occur to every subtree of the hierarchy of Fig. 2. Such situation, however, should appear rather rarely as this type of entries is used in fuzzy relational database only if total lack of knowledge concerning an attribute value occurs.

Obviously, the best case is when the entire `searchVector` is always found in the first descendent of every node along the traversal trail starting from the root. This case is possible only when the `searchVector` is atomic (as for example, *{French}*, considering the partition tree of Fig. 2) causing a traversal of a total of *d* nodes and would result in a time complexity of *O(d)*, where d denotes the depth of the tree. A heuristics of having the longest node (containing maximum number of elements in a certain level) always first may increase the probability that a concept is always discovered in the left subtree.

The function `partitionTreeTraversal` is written in such a manner that it avoids unnecessary branches while traversing the tree. For example, if the top level `searchVector` is *{French, Sushi, Barbeque}* then the traversal would follow the path depicted in Fig. 3 where the DFS traversal is marked by numbers 1 to 10. The partition tree is shaded in Fig. 3 and the search vector is drawn in black. From this, it becomes evident that the average case time complexity of the algorithm is highly dependent on the dispersed characteristic of the `searchVector`. At *level 1* the whole `searchvector` is found; the algorithm then searches the entire `searchVector` in the left-most branch of *level 1*, but it gets only a subset of `searchVector`, this traversal is propagated upto *level 4* to update the summarization of corresponding α-value for only relevant part of the original `searchVector`, i.e. for *{French}*.

Level 3 contains pointers to children that are *{Φ}*, hence a traversal of *level 4* indicates search success of an atomic concept in *level 3*. The algorithm then searches for *{Sushi, Barbeque}* in the middle branch at *level 2*, where again a subset *{Sushi}* is detected. The search fails in *level 3*, in the left branch as it contains only *{Chinese}*, the algorithm immediately returns to *level 2* for the next branch and succeeds for *{Sushi}* that goes upto *level 4*. The algorithm then continues the searching taking *{Barbeque}* alone, at the right most branch of *level 1* in the same way. It is evident from Fig. 3 that the unnecessary branches are always omitted by the `partitionTreeTraversal` algorithm. So the performance of the algorithm depends on the distribution of the `searchVector` in the partition tree. Average case is dominated by the probability of the existence of a `searchVector` in a certain node at certain level.

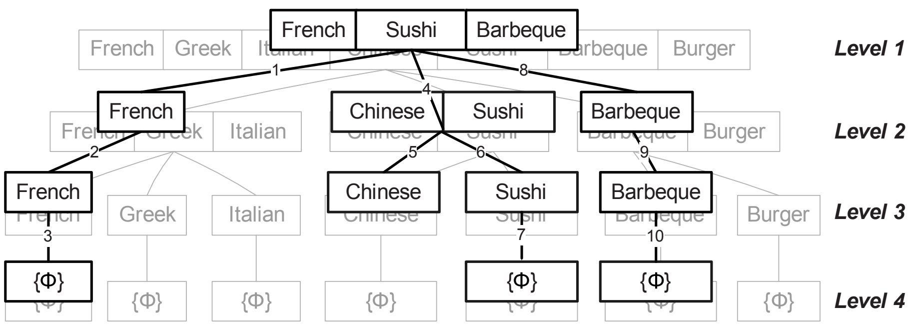


Fig. 3. Path traversed when the top level `searchVector` is *{French, Sushi, Barbeque}*. The partition tree is drawn in light gray. The DFS traversal is marked with numbers from 1 to 10.

3.2 Impact of the Character of the Partition Tree on the Algorithm Performance

The performance of the `partitionTreeTraversal` algorithm is influenced by two factors: (1) degree of imprecision reflected by the number of entered attribute values (i.e. by their number, and their similarity), and (2) the character of the partition tree specified for this attribute. This section includes some tests on the performance of the proposed algorithm when applied with different types of similarity trees.

Data Set and Partition Trees. The experiments have been conducted using some artificial datasets and different types of partition trees. Data were picked randomly from the domain (domain size = 32 symbolic values), where the domain is defined as the set of all values specified at the lowest (i.e. α=*1.0*) level of the partition tree. Number

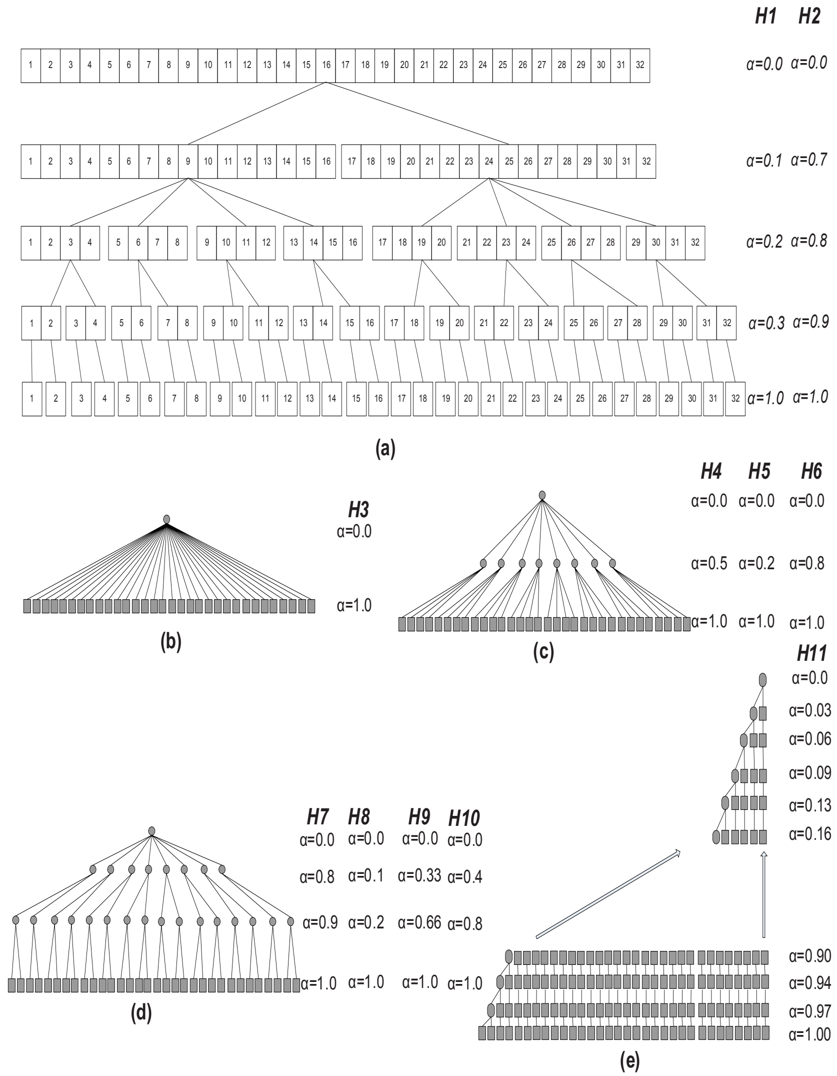


Fig. 4. Hierarchies used for testing. The hierarchies can be grouped as **(a)** *H1*, *H2*, **(b)** *H3,* **(c)** *H4*, *H5*, *H6*, **(d)** *H7, H8, H9, H10* and **(e)** *H11*. Each group possesses a look-alike partition tree, but with different α-values.

of values in a fuzzy record (stored during defuzzification in `searchVector`), considered as 75% imprecise, was kept at the 75% the domain size (i.e. 24 values) to allow for the randomness in choice of values, when maintaining consistent number of values. The number of records in each imprecise data file was 30000, so that the random characteristic could be evenly distributed among the partition tree as the `partitionTreeTraversal` algorithm omits the unnecessary branches. Hence the huge number of records ensures an average case traversal for the comparison purpose in this experiment. Algorithm run times for different percentage of imprecise records are recorded and plotted for different hierarchies.

The numeric values in the partition tree of Fig. 4 (a) are symbols to represent non-numeric descriptive values like all other partition trees of the previous sections. Fig. 4 (a) contains two hierarchies with different α-values (*H1* and *H2*) although their level-wise construction is common. *H1* is a distribution where the concepts are nested at the higher levels and *H2* is a distribution where concepts are nested at the lower levels of the hierarchy according to the α-values. The same domain is used for different types of hierarchies in Fig. 4 (b), (c), (d) and (e) although the numeric symbols are not explicitly presented in the trees. Fig. 4 portrays a total of eleven similarity hierarchies (*H1, H2, …, H11*) which can be grouped in five categories and for each of the category only a single similarity tree is drawn. *H1* and *H2* have a total of five levels in their hierarchy but they differ in the α-values. Similarly, *H4*, *H5* and *H6* have three levels where *H7* to *H10* contain four levels in their hierarchies. The hierarchy *H3* of (b) has only two levels (the root and the leaves), whereas the hierarchy *H11* of (e) is a dendrogram with a total of 32 levels. Some levels in the hierarchy of Fig. 4 (e) have been omitted due to the limited space of this presentation. Obviously, Fig. 4 (e) presents the worst case possible among all the cases of Fig. 4 where every hierarchy contains 32 concepts in the lowest level of the hierarchy. The performance behavior is explained in the following subsection. Behavior of `partitionTreeTraversal` Algorithm. The `partitionTreeTraversal` algorithm is applied to different types of hierarchies of Fig. 4 with different percentage of imprecision of data. The time required for the algorithm to traverse the partition tree is proportional to the percentage of imprecision of the data. Fig. 5 illustrates timing with two look-alike hierarchies, *H1* and *H2* with different nested characteristic depending on the α-values of Fig. 4 (a). It is evident from Fig. 5 that the structure of these two hierarchies are almost the same because the `partitionTreeTraversal` algorithm traverses the same nodes while storing and updating reported existence of concepts in the hierarchy. This is despite the fact that the generated vote's fractions may be different, due to different propagation of α's within the tree. The algorithm is tested using nine other hierarchies depicted in Fig. 4. Fig. 6 (a) is a plot of time at different percentage of imprecision of data with the hierarchies *H3* to *H10* of Fig. 4(a). It is evident from Fig. 6(a) that the distribution of α-values among the levels does not have significant impact on the `partitionTreeTraversal` algorithm. Because the algorithm traverses the partition tree discretely depending on the number of levels rather than traversing in a continuous domain of α-values. A different algorithm that traverses depending on the continuous domain of α-values (rather than depending on the number of levels) better performs for the hierarchies that have nesting at the top levels for a downward search in the partition tree as it discovers most of the levels at the beginning of the domain of α-values. The timing plots of *H4*, *H5* and *H6* follow almost the same trend in the graph. Similarly, *H7*, *H8*, *H9* and *H10* also produce similar graphs that are closely aligned in Fig. 6 (a). The trend of *H3* is separate from the other lines as it is a different hierarchy and performs the best as it has least number of levels in its partition tree. As `partitionTreeTraversal` algorithm is a level dependent depth-first search, it performs the same both for the hierarchies nested at high levels and hierarchies nested at low levels. This behavior is reflected in Fig. 6 (a) because

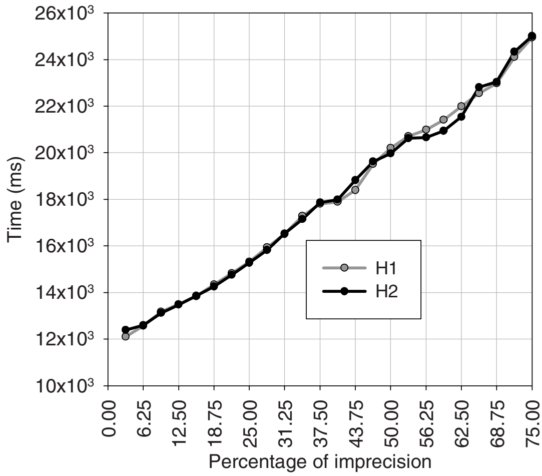


Fig. 5. Comparison of timing with two hierarchies that have the same distribution of concepts but different α-values. The hierarchies *H1* and *H2* are drawn in Fig. 4(a).

the hierarchies with the same tree are clearly grouped in the plot. The plot has three groups because it includes three types of partition trees, (b), (c) and (d) of Fig. 4.

The three types of hierarchies are well separated in the plot of Fig. 6 (a).

Fig. 6 (b) shows that the hierarchy *H11* takes the worst possible time compared to all other hierarchies (*H3* to *H10*). From the experiment, it becomes apparent that `partitionTreeTraversal` is a number of levels-dependent search, where the performance decreases with increase of the number of levels in the partition tree which suggests flattering the partition trees (by making them more bushy), or inclusion of new concepts without increment in the number of levels.

Impact of Conceptualization at Higher Levels and Lower Levels. Let us consider the partition trees of Fig. 7 (a) and (b). The hierarchy of (a) starts splitting at high levels of the tree generating low α-values whereas (b) splits at the comparatively low

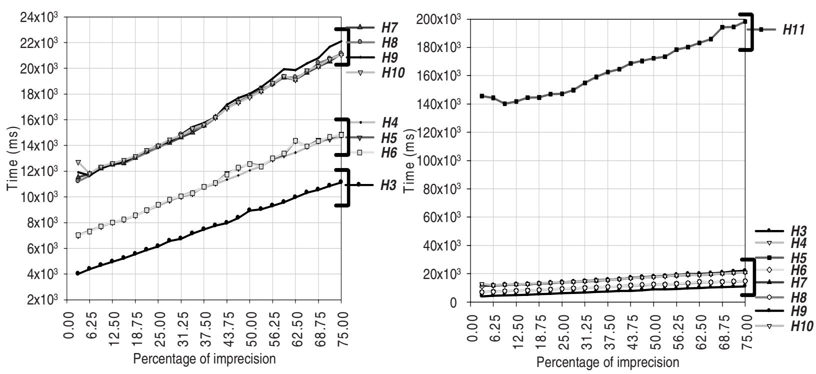


Fig. 6. (a) Plot for hierarchy *H3* to *H10* of Fig. 4 (b). Plot for hierarchy *H3* to *H11* of Fig. 4.

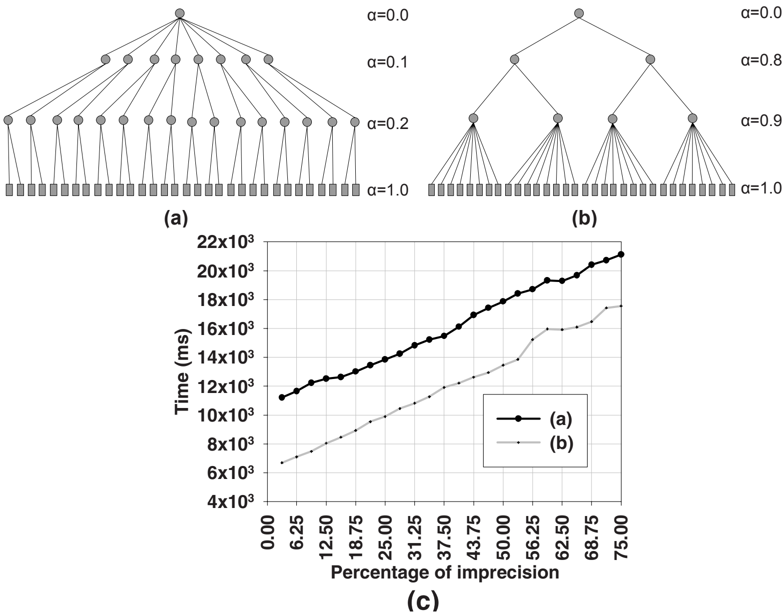


Fig. 7. **(a)** A partition tree where the abstracts start to split at high levels. **(b)** A partition tree where the abstracts split at low level. **(c)** Plot of runtime behaviors of `partitionTreeTraversal` algorithm with the hierarchies of (a) and (b) at different percentage of imprecision of data.

level. Although both the hierarchies have 32 leaf nodes at the bottom of the tree, in (b), they are concentrated fast at the bottom of the tree. In contrary, the other hierarchy possesses more concepts relatively at the top of the tree resulting in more intermediate nodes at every level. The `partitionTreeTraversal` algorithm would take more time in the case of Fig. 7 (a) compared to (b) as the algorithm needs to traverse more nodes for traversing the hierarchy of (a). The resulting runtime behaviors with these two hierarchies with respect to percentage of imprecision of data are plotted in Fig. 7 (c). The figure shows that the `partitionTreeTraversal` algorithm performs better with the hierarchy of Fig.7 (b) than that of (a). This suggests low branching factor near the root of the partition tree and comparatively higher branching factor at the bottom.

4 Conclusions

The work presented in this paper shows that the fuzzy collections can be transferred to the atomic values in the efficient way. The heuristic algorithm presented here allows for transfer of fuzzy records that reflect uncertainty to the form that allows analysis of such data via majority of regular data mining algorithms. Although, the presented heuristic for counting fractions of records as frequencies of observed atomic values might be questioned by statisticians, the technique in our opinion can be treated as a “proof of concept” that imprecise data does not has to be disregarded, but can be mined in a regular matter. The task of discovery of the most appropriate data defuzzification algorithms remains as the topic for future research.

References

[1] J.C. Cubero, N. Marín, J..M. Medina, O. Pons, M.A. Vila, "Fuzzy object Management in an Object-Relational Framework", Proc. of 10th Intl. Conf. of Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2004, pp.1767—1774.

[2] Z. Ma (Ed.), “Advances In Fuzzy Object-oriented Databases: Modeling and Applications”, Idea Group Publishing, Hershey, PA, USA, 2004.

[3] F. Berzal, N. Marín, O. Pons, M.A. Vila, “Development of applications with fuzzy objects in modern programming platforms”, International Journal of Intelligent Systems, 20 (11), 2005, pp. 1117 – 1136.

[4] B.P. Buckles & F.E. Petry, “A fuzzy representation of data for relational databases”, Fuzzy Sets and Systems, 7(3), 1982, pp. 213-226.

[5] F.E. Petry, “Fuzzy Databases: Principles and Applications”, Kluwer Academic Publishers, Boston, MA, 1996.

[6] S. Shenoi and A. Melton, “Proximity Relations in the Fuzzy Relational Database Model”, International Journal of Fuzzy Sets and Systems, 31(3), 1989, pp. 285-296.

[7] S. Shenoi, A. Melton, and L. T. Fan, “Functional Dependencies and Normal Forms in the Fuzzy Relational Database Model”, Information Sciences, 60(1-2), 1992, pp. 1-28.

[8] R. Angryk, “Similarity-driven Defuzzification of Fuzzy Tuples for Entropy-based Data Classification Purposes”, Proc. of 15th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE ’06), Vancouver, Canada, July 2006, pp. 1490-1498.

[9] R. Angryk, “On Interpretation of Non-Atomic Values and Induction of Decision Rules in Fuzzy Relational Databases”, Proc. of 8th Intl. Conf. on Artificial Intelligence and Soft Computing, Zakopane, Poland, June 2006, in Series: Lecture Notes in Artificial Intelligence (LNAI), Vol. 4029, Springer-Verlag, 2006, pp. 170-181.

[10] R. Angryk, F. Petry, R. Ladner, “Mining Generalized Knowledge from Imperfect Data”, Proc. of 10th Intl. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU ’04), Perugia, Italy, July 2004, pp. 739-746.

[11] R. Angryk, F. Petry, “Discovery of Abstract Knowledge from Non-Atomic Attribute Values in Fuzzy Relational Databases,” in: B. Bouchon-Meunier, G. Coletti, R. Yager (Eds.), Modern Information Processing, From Theory to Applications, Elsevier, 2006, pp. 171-182.

[12] M. Kantardzic, J. Zurada, “New Generation of Data Mining Applications”, IEEE Press and John Wiley, 2005.

[13] L.A. Zadeh, “Similarity relations and fuzzy orderings”, Information Sciences, 3(2), 1970, pp. 177-200.

[14] H.–H. Bock, E. Diday (Eds.), “Analysis of Symbolic Data, Exploratory Methods for Extracting Statistical Information from Complex Data”, Springer, 2000.

[15] Bertrand, P. & Goupil, F. (2000), “Descriptive statistics for symbolic data”, in Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data (eds. H.-H. Bock and E. Diday), Berlin, Springer-Verlag, pp 103-124.

Flexible Location-Based Spatial Queries

Gloria Bordogna[1], Marco Pagani[1], Gabriella Pasi[2], and Giuseppe Psaila[3]

[1] CNR-IDPA unità Dalmine, via Pasubio 5, 24044 Dalmine (BG) Italy
gloria.bordogna@idpa.cnr.it,marco.pagani@idpa.cnr.it
[2] Univ. degli Studi di Milano Bicocca, Disco, via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
gabriella.pasi@itc.cnr.it
[3] Univ. degli Studi di Bergamo, Facoltà di Ingegneria, via Marconi 5, 201446 Dalmine (BG) Italy
psaila @unibg.it

Abstract. A model for representing and evaluating flexible Location-Based Spatial Queries (LBSQ) is proposed. In a LBSQ the selection condition is generally a constraint on the distance of the objects in the database (instances) from the user location. Such queries are becoming more and more useful in location-based services such as those provided by cell-phones, Wireless LAN and GPS technologies. However their usefulness is limited by the inability of current systems to represent and manage the imprecision often characterizing the knowledge of the user's and instances' locations. In this contribution we propose a fuzzy model of flexible LBSQs in which either the user location, or the instances locations or the selection condition itself or even all of them are imprecise. To define a unifying approach in all cases of imprecision we generalize the notion of the Minkowski sum within fuzzy sets and apply it to combine the (imprecise) user location with the (soft) query condition. This way we derive the actual soft constraint with respect to the user's location. The instances relevant to the query are those whose locations are included in the actual soft constraint representation to some extent.

Keywords: location-based queries, imprecise locations, Minkowski sum.

1 Introduction

During the last years the outstanding growing market for positioning technologies such as Global Positioning System, Radio Frequency Identification Systems, and Global Systems for Mobile communications has raised the research interest for more efficient and effective location-based services [16,21].

Among location-based services, the evaluation of Location-Based Spatial Queries (LBSQs) that retrieves information based on the current locations of users is a crucial task, due to the mobility of users and, in some applications, also of the objects in the database (instances) [8,18]. For example, an airplane pilot may ask for "the nearest airplane crossing his route". In order to avoid possible collisions, the answer to this query should not depend on the location of the airplanes when the pilot issued the request but on the locations where the pilot receives the query results, which are generally not precisely known at the time the query is issued. In this example, both the user and the instances in the database (the other airplanes) are moving, thus their

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 36–45, 2007.
springerlink.com

locations are known with imprecision. The need to take into account the location of the user and/or of the instances at the time the query results are received by the user is necessary in many applications of robotics. Generally this is useful whenever the speed of the user and/or of the instances is so high to determine a situation that makes the query results sensible with respect to the current positions of the considered items at the time the user receives the query results.

These are not the only cases of imprecision in LBSQs. Imprecision on the user location can derive by several causes such as measurement errors or limited resolution of the device used to detect the location coordinates, or insufficient network speed. In some cases imprecision can be introduced on purpose to mask the exact user location for preserving the privacy of the user [1,13]. Finally, LBSQs can involve imprecision also in the condition specification such as in the query “find the taxi cabs that are very close”. Generally the uncertainty in location data has been modeled by means of probability distributions on the spatial domain [18], and the research in this respect mainly focused on efficiency issues [8, 9, 18, 19].

In our proposal, we model in the unifying framework of fuzzy set theory the representation and evaluation mechanism of flexible LBSQs by taking ideas from the approaches to flexible querying in fuzzy databases [5,7,9,11,17,20]. In this contribution by flexible LBSQs we intend soft range queries against possibly ill-defined location information admitting degrees of satisfaction. A range query specifies a selection condition that consists in a bounding box centered at the user location. With the terms “soft range queries” we mean LBSQs specifying a vague range condition. This is expressed by a linguistic term such as *close* defined as a soft constraint on the spatial domain [3]. Imprecision may affect user location, instances location and the range condition alone or in any combination one another. Imprecision on location data is represented by means of possibility distributions [5]. The soft constraint specifying the vague range condition is defined with a membership function that decreases with the distance from the coordinates’ origin. The notion of Minkowski sum is generalized to fuzzy sets and is used to generate the actual soft constraint with respect to the possibly imprecise user location. Finally, the degree of satisfaction of a soft range query is computed as the fuzzy inclusion degree of the instances’ locations in the fuzzy set representing the actual soft constraint [4,7]. In the next section an overview of the literature on LBSQs is briefly introduced. In section 3 the formalization of Flexible LBSQ evaluation is defined. Finally, in the conclusion the main results are summarized.

2 Related Works

The research on LBSQs focused mainly on efficiency issues such as the investigation of new ways of indexing and caching spatial data to support the processing of LBSQs including point query, window query, nearest neighbor (NN) search, k nearest neighbor search [14,23,24,25]. Another issue was the management of LBSQs in a distributed way so as to achieve efficiency while allowing complex queries based on the use of location-dependent operators [12,15]. In this respect, another direction is the study of LBSQs involving some uncertainty in location data. This is also the focus or our proposal. Uncertainty on the user location can derive by several causes as outlined in the introduction, or can be introduced on purpose to mask the exact location for preserving the privacy of the user [1,13]. In this respect the amount of uncertainty required to meet both privacy and the requirement on the service quality has been studied [8]. Generally the location uncertainty has been modeled by means of probability distributions on the spatial domain, basically bi-dimensional Gaussian functions or uniform distributions within a window or neighborhood [18]. The research on this topic faced the evaluation of probabilistic queries, such as probabilistic range queries [22]. Probabilistic queries evaluate uncertain location information and provide plausible answers in the form of probabilities. Imprecise LBSQs have also been studied to model imprecision affecting the instances location, such as in the case of moving objects [8, 18,19]. To our knowledge no one has yet considered that both the location of the user and of the objects in the spatial database and the selection condition itself can be imprecise such as in the query "find the closest airplanes to my path ".

In our proposal we consider these three situations alone and in any combinations one another, and provide a modeling within fuzzy set theory, by taking ideas from flexible querying in fuzzy databases [5,7,11,17,20]. Imprecision on location data is represented by means of possibility distributions, which are easier to define than probability distributions since they do not need to satisfy the normalization constraint. The vague selection condition, that is a vague range condition, is represented by a soft constraint defined with respect to the coordinates' origin. The generation of the soft constraint with respect to the user's location is dynamically determined by computing the fuzzy Minkowski sum [2], that is defined in this contribution to this purpose. This way, we generate a new soft constraint that the instances' locations, possibly imprecise, must satisfy to some extent in order to be retrieved. Similarly to what happens in fuzzy databases, a matching function between fuzzy sets is defined to compute the degrees of satisfaction of the instances, which in our context is a fuzzy inclusion function between spatial distributions [4,7].

3 Flexible Location-Based Spatial Queries

In this section we classify the kinds of flexible LBSQs considered in this contribution and introduce their formal representations and evaluation functions. In Table 1 the types of LBSQs are characterized, based on the imprecision affecting their information units, i.e., the user location, the instances locations, and the selection condition. For simplicity we restrict our analysis to soft range queries and define a mechanism to compute degrees of satisfaction of the instances. We consider the linguistic expression *close* as example of specification of the vague range condition. Notice that *close* could be replaced by any other linguistic term defining a soft constraint on the distance from a position such as "*very close*", "*not too far*", etc.; thus, the vague range condition can be regarded as the specification of a nearest neighbor search condition.

3.1 Type 1: Crisp Query

This is the usual crisp range query of the kind "find instances located at a maximum distance [$\pm\Delta x$, $\pm\Delta y$] from my location". In this LBSQ both the location data and the range condition are precisely represented by their coordinates (x,y) on the spatial domain and the bounding box $[\pm\Delta x, \pm\Delta y]$ centered at the user location (x_u,y_u). Only the instances whose coordinates (x_i,y_i) fall within the limits $x_u\pm\Delta x$ and $y_u\pm\Delta y$ completely satisfy the range condition:

$$\text{Select } i \mid \{(x_i,y_i)\} \subseteq \text{box}((x_u-\Delta x,\ y_u-\Delta y),(x_u+\Delta x,\ y_u+\Delta y))$$

Table 1. Types of flexible LBSQ

Query type	User location	Instances location	Query range condition
1	(x_u,y_u)	(x_i,y_i)	$[\pm\Delta x, \pm\Delta y]$
2	*Around*(x_u,y_u)	(x_i,y_i)	$[\pm\Delta x, \pm\Delta y]$
3	(x_u,y_u)	*Around*(x_i,y_i)	$[\pm\Delta x, \pm\Delta y]$
4	*Around*(x_u,y_u)	*Around*(x_i,y_i)	$[\pm\Delta x, \pm\Delta y]$
5	(x_u,y_u)	(x_i,y_i)	*close*
6	*Around*(x_u,y_u)	(x_i,y_i)	*close*
7	(x_u,y_u)	*Around*(x_i,y_i)	*close*
8	*Around*(x_u,y_u)	*Around*(x_i,y_i)	*close*

3.2 Type 2: Crisp Query with Imprecise User Location

In this type of LBSQ, the user location is affected by imprecision *Around*(x_u,y_u) while the instances' locations and the selection conditions are precise. This is for example the case of a moving robot whose location varies in time looking for some stable resources close to his current location, that is around (x_u,y_u). Evaluating this kind of queries implies having a representation of the imprecise user location. We represent an imprecise user location *Around*(x_u,y_u) by means of a possibility distribution π_u: $X\times Y\rightarrow[0,1]$ on the bi-dimensional spatial domain. The form of π_u depends on the specific application.

Examples of definition of π_u

In the case of imprecision introduced on purpose, π_u can be defined as a uniform distribution within a box or a circle centered at (x_u,y_u), i.e., $\pi_u(x,y)=u\in[0,1]$ $\forall x,y$ with $|(x,y)-(x_u,y_u)| < r$, $\pi_u(x,y)=0$ otherwise. In the case of a moving user such as a robot, π_u can be built based on a data driven approach by monitoring the robot, and by determining its speed and travel direction. Given two subsequent locations of the robot, i.e., the user, (x_0,y_0) and (x_1,y_1) at time t_0 and t_1 respectively, we can build π_u by considering the robot's speed and most possible position (x_t,y_t) at the answer time t:

$$\pi_u(x,y)=\begin{cases} e^{-\left[\frac{1}{2}k((x-x_t)^2+(y-y_t)^2)\right]} for \quad |(x,y)-(x_t,y_t)|<|(x_1,y_1)-(x_t,y_t)| \quad with \begin{cases} x_t=x_1+\dfrac{x_1-x_0}{t_1-t_0}(t-t_1) \\ y_t=y_1+\dfrac{y_1-y_0}{t_1-t_0}(t-t_1) \end{cases} \end{cases} \quad (1)$$

In this case π_u decreases smoothly with the distance from the most possible location (x_t, y_t) and becomes null outside the circle of radius $r = \sqrt{(x_1, y_1)^2 - (x_t, y_t)^2}$.

-°-

Based on the user's location representation π_u and the precise range condition, we derive a representation of the region delimiting candidate instances that satisfy the query to some extent. This is achieved by computing the Minkowski sum $\oplus$ of the range condition and the at least $\propto$-possible locations of the user with $\propto>0$. We represent the range condition $Z=[\pm\Delta x, \pm\Delta y]$ as a box centered in (0,0) and displacements $\pm\Delta x$ and $\pm\Delta y$ on the x and y axis respectively:

$$S=\textit{At least} \propto\textit{-possible}(\pi_u(x,y))=\{(x,y) \mid (\pi_u(x,y)> \propto)\}$$

The Minkowski sum $\oplus$ of the two polygons S and Z on the Euclidean spatial domain is defined as [2]:

$$S \oplus Z = \{ s + z \mid s \in S \text{ and } z \in Z \} \quad (2)$$

in which s and z are points on the spatial domain. The Minkowski sum is defined as the union of all the translations of Z by a point s located in S. For **type 2** queries the Minkowski sum can be interpreted as the union of all the range queries by considering all possible positions of the user who is located somewhere inside S. Clearly, only the instances whose location (x_i,y_i) is within the region $S\oplus Z$ satisfy the query. Then, the evaluation of this range query with user location imprecision corresponds to select the instances that satisfy the topological relationship inclusion (x_i,y_i) in $S\oplus Z$, i.e.:

$$\text{select } i \mid \{(x_i,y_i)\} \subseteq (\textit{At least 0-possible}(\pi_u \oplus [\pm\Delta x, \pm\Delta y])$$

Note that we can generalize the evaluation of the **type 2** query with imprecise user location considering any desired $\propto$-possible location of the user with $0<\propto\leq 1$. To compute degrees of satisfaction for the instances we define the generalized fuzzy Minkowski sum $\oplus_F$ that combines two fuzzy sets and determines a fuzzy set as a result.

Definition of the generalized Fuzzy Minkowski sum
Given two fuzzy sets S and Z defined on a spatial domain X, the generalized Fuzzy Minkowski sum $S\oplus_F Z$ is defined as the fuzzy union (max) of all the translations of Z by every element s belonging to some extent to the fuzzy set S:

$$S \oplus_F Z = \{ \mu_{S\oplus_F Z}(r) / r \mid r = s + z \text{ and } s \in S,\ z \in Z\} \quad (3)$$

where $\mu_{S\oplus_F Z}(r) = \max_{\forall s\in S, \forall z\in Z \mid s+z=r}(\min(\mu_S(s), \max(\mu_S(s+z), \mu_Z(z))))$

Example: let us consider a simple example in a one-dimensional spatial domain:

S:={0.5/2, 1./3, 1./4, 0.7/5} Z:={0.2/-2, 0.8/-1, 1./0, 0.8/1, 0.2/2}

R=S $\oplus_F$ Z={0.2/0, 0.5/1, 0.8/2, 1./3, 1./4, 0.8/5, 0.7/6, 0.2/7}

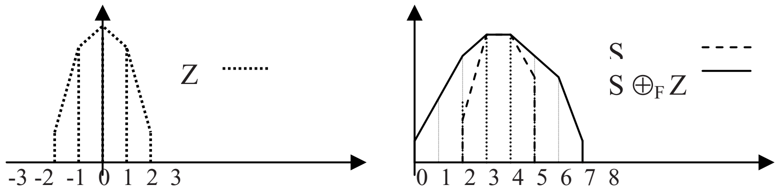


Fig. 1. Representation of the Fuzzy Minkowski sum in a one-dimensional domain

It can be proved that the fuzzy set R representing the result of Generalized Fuzzy Minkowski sum $R=S\oplus_F Z$ includes S, i.e., $\mu_{S\oplus_F Z}(x) \geq \mu_S(x)\ \forall x\in X$, and that it reduces to the crisp Minkowski sum in the case in which both S and Z are classic sets:

$$\mu_{S\oplus_F Z}(x)=1\ \forall x\in S\oplus_F Z \text{ and } \mu_{S\oplus_F Z}(x)=0 \text{ otherwise.}$$

-°-

In the context of the evaluation of **type 2** queries with imprecise user location $S=\pi_u$ and crisp range condition $Z=[\pm\Delta x, \pm\Delta y]$, we assume that $\mu_Z(z)=1$ for $-\Delta x\leq x_z\leq\Delta x$ and $-\Delta y\leq y_z\leq\Delta y$, while $\mu_Z(z)=0$ otherwise. In this case the generalized fuzzy Minkowski sum $\pi_u\oplus_F[\pm\Delta x,\pm\Delta y]$ identifies the union of all the translations of the range Z by any point $s=(x_s,y_s)$ belonging to some extent to the possible user location π_u. It can be interpreted as the fuzzy union of all the range queries by considering all possible positions of the user who is located somewhere inside S. The instances whose location (x_i,y_i) is within the support of the fuzzy set $\pi_u\oplus_F[\pm\Delta x, \pm\Delta y]$ satisfy the query. For these instances, we can compute a satisfaction degree *degree(i)* for ranking the instances to the range query based on the evaluation of the fuzzy inclusion of their precise location (x_i,y_i) in $\pi_u\oplus_F(\pm\Delta x, \pm\Delta y)$ as follows:

$$degree(i)= \text{degree}\,(\{(x_i,y_i)\}\subseteq_F (\pi_u\oplus_F[\pm\Delta x, \pm\Delta y]))= \mu_{\pi_u\oplus_F[\pm\Delta x, \pm\Delta y]}(x_i,y_i) \tag{4}$$

3.3 Type 3: Crisp Range Query with Imprecise Instances' Locations

This type of LBSQ is dual with respect to the previous one. In this case, the instances' locations are imprecisely known while the user position is precise. For example, this is the case of moving objects, such as taxicabs with the user being located at a taxi station. To evaluate this kind of queries, first we represent the instances locations by means of possibility distributions on the spatial domain $Around(x_i,y_i)=\pi_i$. As in the previous case, we can adopt a data driven approach to generate π_i by exploiting collected information on previous positions of the instances.

To evaluate this kind of queries, we can adopt two alternative procedures.

Procedure A

With this procedure, we first build the crisp Minkowski sum $(x_u,y_u)\oplus[\pm\Delta x, \pm\Delta y]$ of the precise user location with the range condition. Then, we derive the fuzzy set $(x_u,y_u)\oplus_F[\pm\Delta x, \pm\Delta y]$ from the crisp Minkowski sum as follows:

$$\mu_{(x_u,y_u)\oplus_F[\pm\Delta x, \pm\Delta y]}(r)=1\ \ \forall r\in(x_u,y_u)\oplus[\pm\Delta x,\pm\Delta y],\ \mu_{(x_u,y_u)\oplus_F[\pm\Delta x,\pm\Delta y]}(r)=0 \text{ otherwise}$$

Finally, to compute the degrees of satisfaction of the LBSQ by the *N* instances, for each instance *i,* we evaluate the degree of fuzzy inclusion of its imprecise location π_i in the fuzzy set $(x_u,y_u)\oplus_F[\pm\Delta x, \pm\Delta y]$ as follows:

$$degree(i) = \text{degree}\,(\pi_i \subseteq_F ((x_u, y_u)\oplus_F[\pm\Delta x, \pm\Delta y]))$$

Definition of fuzzy inclusion degree between fuzzy sets

The fuzzy inclusion between two fuzzy sets *A* and *B* on a spatial domain *X* can be defined based on the cardinality ∫ of the fuzzy sets and the proportion of *A* included in *B* [3]:

$$degree(i) = \text{degree}\,(A \subseteq_F B) = \frac{\int_X (A \cap_F B)\mathrm{d}x}{\int_A A} = \frac{\int_X min(\mu_A(x), \mu_B(x))dx}{\int_X \mu_A(x)dx} \quad (5)$$

where $\cap_F$ is the intersection of fuzzy sets and $\int \mu_A(x)\mathrm{d}x$ is the integral (for X continuous) or sum (for X discrete) of the membership values of the fuzzy set *A*. Notice that formula (5) reduces to formula (4) in the particular case in which the fuzzy set *A* is a single point of the spatial domain as in the case of **type 2** queries.

-°-

In the case of **type 3** queries, formula (5) reduces to compute the proportion of π_i included in the support of the crisp Minkowski sum *support*$((x_u,y_u)\oplus_F[\pm\Delta x, \pm\Delta y])$:

$$degree(i) = \text{degree}(\pi_i \subseteq_F ((x_u,y_u)\oplus_F[\pm\Delta x,\pm\Delta y])) = \frac{\int_{support((x_u, y_u)\oplus_F[\pm\Delta x, \pm\Delta y])} \pi_i(x), dx}{\int_X \pi_i(x)dx} \quad (6)$$

This procedure computes just once the crisp Minkowski sum based on (2) and evaluates *N* fuzzy inclusion degrees between fuzzy sets by applying formula (6).

Procedure B

By adopting this procedure, the evaluation of the LBSQ of **type 3** corresponds to evaluate *N* range queries of **type 2**, in which we exchange the user location with the instances location. We first build the *N* fuzzy Minkowski sums $(\pi_i\oplus_F[\pm\Delta x,\pm\Delta y])$, with i=1,..,*N* of the imprecise instance location π_i with the crisp range condition $[\pm\Delta x, \pm\Delta y]$. Then, if the precise user location falls within each *N* region, we retrieve the corresponding instance. Also in this case, for each instance we can compute a satisfaction degree *degree(i)* of the query based on the evaluation of the fuzzy inclusion of the precise user location (x_u,y_u) in the fuzzy Minkowski sum $(\pi_i\oplus_F[\pm\Delta x, \pm\Delta y])$ by means of formula (4):

$$degree(i) = \text{degree}\,(\{(x_u, y_u)\} \subseteq_F (\pi_i \oplus_F[\pm\Delta x, \pm\Delta y])) = \mu_{\pi_i \oplus_F [\pm\Delta x, \pm\Delta y]}(x_u,y_u)$$

With respect to ***procedure A***, in this case the computation of the fuzzy inclusion degree is much simpler than (6) since it reduces to formula (4), but we have the increased cost of computing *N* fuzzy Minkowski sums instead of just a crisp one. The decision on which procedure to adopt depends on efficiency reasons, i.e. the costs of computation of N fuzzy Minkowski sums (based on def. (3)) versus the cost of a crisp Minkowski sum plus N fuzzy inclusions between fuzzy sets (6).

3.4 Type 4: Crisp Range Query with Both Imprecise User's and Instances' Locations

This situation is the one in which both the user and instances locations are imprecise, such as in the airplane example of the introduction. In this case we adopt ***procedure A*** described for **type 3** queries. We have the increased complexity of computing a fuzzy Minkowski sum $\pi_u \oplus_F[\pm\Delta x, \pm\Delta y]$ of the imprecise user location, and the crisp range condition based on definition (3) instead of a simpler crisp Minkowski sum. In constrast, ***procedure B*** is much more inefficient than ***procedure A***. In fact, since we have N imprecise instances' locations π_i, by adopting ***procedure B*** we would have to compute N fuzzy Minkowski sums $\pi_i \oplus_F[\pm\Delta x,\pm\Delta y]$. Further, being also the user location imprecise, π_u, we would have also to evaluate N fuzzy inclusions of the fuzzy set $A=\pi_u$ in the fuzzy sets $B=\pi_i \oplus_F(\pm\Delta x,\pm\Delta y)$, with i=1,..$N$, by applying formula (5) so as to derive the N degrees to rank the instances:

$$degree(i)= \pi_u \subseteq_F (\pi_i \oplus_F[\pm\Delta x,\pm\Delta y]).$$

3.5 Type 5, 6, 7 and 8: Soft Range Queries with Possible Imprecise Location

All these types of LBSQs specify a vague range condition by means of a linguistic predicate such as *close*. In the context of fuzzy databases, vague conditions are defined as soft constraints on the domains of attributes [4,5,6,17,20]. We retain this representation and define a vague range condition like *close* as a soft constraint on the spatial domain $X\times Y$ with the membership function μ_{close} that decreases with the distance from the coordinate origin (0,0) (see Figure 2): $\mu_{close}(x,y)\rightarrow[0,1]$ $\forall x,y \in X\times Y$. The shape of μ_{close} can be either a box with vague boundaries or a symmetric function decreasing with the distance from the origin. This way the vague range condition is defined in an absolute way. It is during the evaluation of the query that the representation of the actual soft constraint is generated with respect to either the users' or the instances locations.

To evaluate this kind of queries according to definition (3), we first compute the fuzzy Minkowski sum of the user location $location_u$, that can be either precise (for **type 5** and **7** queries) or imprecise π_u (**type 6** and **8** queries), and the fuzzy set representing the soft range condition *close*:

$$close_u := location_u \oplus_F \mu_{close}.$$

This way we generate $close_u$ that represents the actual soft range condition defined with respect to the user location. $close_u$ must be satisfied to some extent by the location of the i-th instance, $location_i$, in order to retrieve the instance.

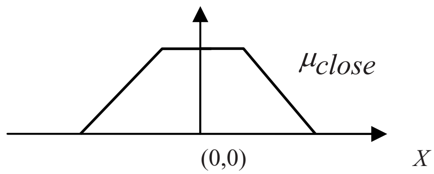


Fig. 2. Vague range condition *close* in a one dimensional doman *X*

The second step computes the degrees of the fuzzy inclusion of each instance location $location_i$ in the fuzzy set $close_u$. This step has a distinct complexity if we are evaluating **type 5 - 6** queries with respect to **type 7 – 8** queries. For **type 5 - 6** queries, since the instances' locations are precise (x_i,y_i), we just take, as satisfaction degree of an instance *i*, the value computed by formula (4):

$$degree(i) = \mu_{location_u \oplus \mu_{close}}(x_i, y_i)$$

In the case in which $location_i = \pi_i$ (imprecise instance location) as in **type 7 - 8** queries, we have to compute the fuzzy inclusion degree of the fuzzy sets $A=\pi_i$ into the fuzzy set $B=close_u$ for each instance *i* by applying definition (6):

$$degree(i) = \text{degree}(\pi_i \subseteq_F (location_u \oplus_F \mu_{close}))$$

4 Conclusions

Providing effective and efficient mechanisms to support LBSQs affected by imprecision is useful in many application fields. In this contribution, a model for evaluating flexible LBSQs with imprecise locations and vague selection condition is proposed. The model is based on the fuzzy generalization of the Minkowski sum between crisp sets. The fuzzy Minkowski sum is defined on two fuzzy subsets of the spatial domain, so as to produce a fuzzy set as a result. We also apply the notion of fuzzy inclusion of fuzzy sets to compute a degree of satisfaction of a LBSQ in the case of imprecise instances' locations. As far as we know, there is not up to date a proposal in this respect within the fuzzy context. The proposals based on probability distributions faced just some cases in which imprecision affects either user location, or instance locations but never both of them at the same time with vague range conditions. Our proposal has the advantage with respect to the probabilistic approach of formalizing all situations of imprecision in LBSQs evaluation within a unifying framework.

References

1. Beresford A., Stajano F.: Location Privacy in Pervasive Computing. IEEE Pervasive Computing, 2(1). (2003) 46-55.
2. Berg M., Kreveld M., Overmars M., Schwarzkopf O.: Computational geometry, Algorthms and Applications, Springer Verlag, (2000).

3. Bordogna G., Pagani M., Psaila, G.: SQL spatial with Customisable soft selection conditions, in “Flexible Databases supporting Imprecision and Uncertainty”, series Studies in fuzziness and soft computing, Springer Verlag,, 203, (2006).
4. Bordogna G., Pasi G.: A semantics for Soft selection conditions in Fuzzy Databases based on a Fuzzy Inclusion, in Proc. of IFSA03 (2003).
5. Bosc P., and Prade H.: An Introduction to the Fuzzy Set and Possibility Theory-based Treatment of Flexible Queries and Uncertain and Imprecise Databases, in Uncertainty Management in Information Systems: from Needs to Solutions, A. Motro and P. Smets eds, Kluwer, (1997), 285-324.
6. Bosc P., Buckles B.B., Petry F.E., Pivert O.: Fuzzy Databases, in Fuzzy Sets in Approximate Reasoning and Information Systems, J. C. Bezdek, D. Dubois and H. Prade eds., Kluwer Ac. Pub., (1999), 403-468.
7. Bosc P., Pivert O.: A Formal Framework for Representation-Based Querying of Databases Containing Ill-known Data, in Proc. of the IIA'99 SOCO'99, (1999).
8. Cheng R., Kalashnikov D., Prabhakar S.: Evaluating probabilistic queries on imprecise data, in proc. of ACM SIGMOD, (2003).
9. Cheng R., Xia Y., Prabhakar S., Shah R., Vitter J.S.: Efficient indexing methods for probabilistic threshold queries over uncertain data. In Proc. of VLDB, (2004).
10. De Caluwe R.: Fuzzy and Uncertain Object-Oriented databases, Concepts and Models, World Scientific, (1997).
11. Dubois D., Prade H.: Tolerant Fuzzy Pattern matching: an Introduction, in Fuzziness in Database Management Systems, P. Bosc and J. Kacprzyk eds., Physica Verlag, (1994).
12. Gedik B., Liu L.: Mobiyes Distributed processing of continuously moving queries on moving objects in a mobile system, in Proc. of EDBT, (2004).
13. Gruteser M. , Grunwald D.: Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking. In Proc. of the 1st Int. Conf. on Mobile Systems, Applications and Services, (2003).
14. Hu H., Xu J., Wong W., Zheng B., Lee D., Lee L.: Proactive Caching for Spatial Queries in Mobile Environments, in Proc. of the 21st ICDE (2005).
15. Iliarri S., Mena E., Illarramendi A.: Location-dependent queries in mobile context: distributed processing using mobile agents. ACM TMC, 5(8) 2006.
16. Lee D., Xu J., and Zheng B.: Data management in location-dependent information services, IEEE Pervasive Computing, 1(3), (2002).
17. Petry F.E., Fuzzy Databases, Kluwer, (1996).
18. Pforser D., Jensen C.: Capturing the uncertainty of moving-objects representations. in Proc. of SSDBM, (1999).
19. Tao Y., Papadias D., Shen Q.: Continuous NN search. In Proc. of VLDB, (2002).
20. Vila M.A., Cubero J.C., Medina J.M., Pons O.: A Conceptual Approach for Dealing with Imprecision and Uncertainty in Object-Based data Models, Int. J. of Intel. Syst., 11(10), (1996), 791-806.
21. Warrior J., McHenry E., McGee K.: They know where you are, Spectrum, 40(7), (2003) 20-25.
22. Wolfson O., Sistla A., Chamberlain S., Yesha Y.: Updating and querying databases that track mobile units. Distributed and Parallel Databases, 7(3), (1999) 257-387.
23. Xu J., Zheng B., Lee W., Lee D., Energy Efficient Index for Querying Location_Dependent Data in Mobile Broadcast environment, in proc. of ICDE, (2003).
24. Xu J., Lee W., The D-Tree: An Index Structure for Planar Point Queries in Location-Based Wireless Services, IEEE Trans. on Knowledge and data Eng., 16(12), (2004), 1-16.
25. Zheng B., Xu J., Lee W., and Lee D., Energy-Conserving Air Indexes for NN Search, E. Bertino et al. Eds.: EDBT 2004, LNCS 2992, Springer-Verlag, (2004), 48–66.

Complex Quantified Statements Evaluated Using Gradual Numbers

Ludovic Liétard[1] and Daniel Rocacher[2]

[1] IRISA/IUT, rue Edouard Branly,
[2] IRISA/ENSSAT, rue de Kerampont, BP 30219 22302, Lannion cedex, France
BP 447, 22305, Lannion cedex France,
ludovic.lietard@univ-rennes1.fr, rocacher@enssat.fr

Abstract. This paper is devoted to the evaluation of quantified statements in the context of flexible querying of relational databases. Two types of quantified statements can be distinguished and the most sophisticated one (of type "*Q B* X are *A*") is considered in this paper. The contribution is to propose a new theoretical background for their evaluation (using an arithmetic on gradual numbers ($\mathbb{N}_f$, $\mathbb{Z}_f$, $\mathbb{Q}_f$)).

Keywords: flexible querying, linguistic quantifiers, gradual numbers.

1 Introduction

Flexible querying of relational databases aims at expressing preferences into queries instead of boolean requirements as it is the case for regular (or crisp) querying. As a consequence, a flexible query returns a set of discriminated answers to the user (from the best answers to the less preferred). Many approaches to define flexible queries have been proposed and it has been shown that the fuzzy set based approach is the more general [1]. An extension of the SQL language (namely SQLf [2]) has been proposed to define sophisticated flexible queries calling on fuzzy sets.

In this context, predicates are defined by fuzzy sets and are called fuzzy predicates. They can be combined using various operators such as generalized conjunctions and generalized disjunctions (respectively expressed by t-norms and t-conorms) or using more sophisticated operators such as averages. In addition, linguistic quantifiers [7] (which are quantifiers defined by linguistic expressions like "*most of*" or "*around 3*") allow to define a particular type of conditions called quantified statements.

Two types of quantified statements can be distinguished. A statement of the first type is denoted "*Q* X are *A*" where *Q* is a linguistic quantifier, X is a crisp set and *A* is a fuzzy predicate. Such a statement means that "*Q* elements belonging to X satisfy *A*". An example is provided by "*most of* employees are *young*" where *Q* is *most of*, X is a set of employees whereas *A* is the condition *to be young*. In this first type of quantified statements, the referential for the linguistic quantifier is a crisp set (denoted by X, the set of employees when considering the example). In the second type, the quantifier applies to a fuzzy set as in "*most of young* employees are *well-paid*"

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 46–53, 2007.
springerlink.com

where the referential for *most of* is a fuzzy set (of *young* employees). This second type of statement is written "*Q B* X are *A*" (when referring to the previous example, *Q* is *most of* while *B* is the predicate *to be young* and *A* is the predicate *well-paid*). Such a statement means that, among the elements from X which satisfy *B*, a quantity *Q* satisfies *A*.

Quantified statements can be used in flexible queries as in the query : "retrieve the firms where *most of young* employees are *well-paid*". After query evaluation, each firm is associated to a degree in [0,1] expressing its satisfaction with respect to the quantified statement "*most of young* employees are *well-paid*". The higher this degree, the better answer is the firm.

Two kinds of linguistic quantifiers can be distinguished: absolute quantifiers (which refer to an absolute number such as *about 3*, *at least 2*, ...) and relative quantifiers (which refer to a proportion such as *about the half*, *at least a quarter*, ...). To evaluate a quantified statement is to determine the extent to which it is true and this paper proposes a new theoretical framework to evaluate quantified statements of type "*Q B* X are *A*" (some elements devoted to the evaluation of quantified statement of type "*Q* X are *A*"- in the same theoretical framework - can be found in [4]). In addition, the case of an absolute quantifier is not dealt with since, in this case, the statement reverts to a quantified statement of type "*Q* X are *A*" ("*around 3 young* employees are *well-paid*" is equivalent to "*around 3* employees are *(young* and *well-paid)"*). Propositions are based on the handling of gradual integers ($\mathbb{N}_f$, $\mathbb{Z}_f$) [5] and gradual rational numbers ($\mathbb{Q}_f$) as defined in [6]. These specific numbers express well-known but gradual numbers and differ from usual fuzzy numbers which define imprecise (ill-known) numbers. In addition, since their definition is closely related to the concept of cardinality of a fuzzy set, their use to evaluate quantified statements appears to be natural.

Section 2 introduces the definition of linguistic quantifiers while section 3 states the principles advocated in this paper for the evaluation of quantified statements of type "*Q B* X are *A*" where *Q* is relative. Section 4 proposes a fuzzy truth value as result of the evaluation. Since a scalar value is mandatory in the context of SQLf, section 5 proposes two interpretations to obtain two different scalar evaluations of this fuzzy truth value.

2 Linguistic Quantifiers

This section proposes two different representations for linguistic quantifiers (which can be derived one from each other). The first definition [7] is the traditional one (based on a constraint on a cardinality or proportion), while the second one (based on a tolerance) is more suited to the framework of gradual numbers.

A first representation for an absolute quantifier (resp. relative quantifier) is a fuzzy subset Q of the real line (resp. of the unit interval [0,1]). This fuzzy subset is interpreted in terms of a matching between cardinalities (or proportions) and degrees of satisfaction. In both cases, $\mu_Q(j)$ represents the truth value of the statement "*Q* X are *A*" when j elements in X completely satisfy *A*, whereas *A* is fully unsatisfied by the others (j being a number or a proportion). The representation of an increasing linguistic quantifier satisfies:

i) $\mu_Q(0) = 0$, ii) $\exists$ k such as $\mu_Q(k) = 1$, iii) $\forall$ a, b if a > b then $\mu_Q(a) \geq \mu_Q(b)$. (1)

A decreasing linguistic quantifier is defined by:

ii) $\mu_Q(0) = 1$, ii) $\exists$ k such as $\mu_Q(k) = 0$, iii) $\forall$ a, b if a > b then $\mu_Q(a) \leq \mu_Q(b)$. (2)

A unimodal quantifier is a fuzzy subset *Q* such that:

i) $\mu_Q(0) = 0$, ii) $\exists!$ k such as $\mu_Q(k) = 1$,

iii) $\forall$ a < b < k then $\mu_Q(a) \leq \mu_Q(b)$ and $\forall$ a > b $\geq$ k then $\mu_Q(a) \leq \mu_Q(b)$. (3)

A second representation of a linguistic quantifier is possible when thinking in terms of a soft constraint (a tolerance) with respect to a precise value. In this case, the linguistic quantifier is decomposed into two parts, a value which can be either an integer or a proportion, and a tolerance function (T) with respect to this precise value. As an example, the linguistic quantifier *around 3* can be defined in term of a tolerance (T) with respect to the value 3 (k = 3).

These two representations are equivalent. When considering the second representation, the value or the proportion is the value k of the first representation (see (1-3)), while the tolerance function T is obtained using the translation $T(x\text{-}k) = \mu_Q(x)$. As a consequence, a quantifier is equivalently represented by a fuzzy set Q or a by a couple (k, T) made of the two items.

Example 1. The linguistic quantifier *about half* is represented by the fuzzy set given by Fig. 1 and by the couple (k=0.5, T) where the tolerance function is given by Fig.2.

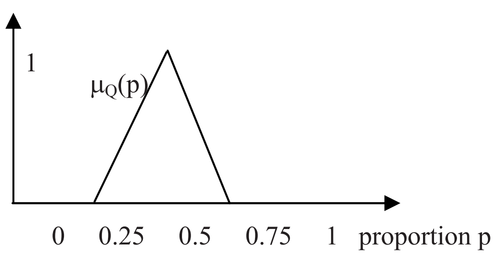


Fig. 1. The quantifier *about half*

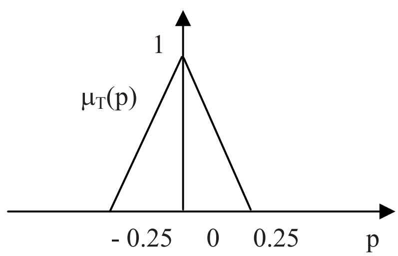


Fig. 2. The tolerance function associated to the linguistic quantifier *about half*

3 Principles for the Evaluation of Quantified Statements

First, we consider the evaluation of quantified statements in the particular case of crisp predicates. This situation is then adapted to the case of fuzzy predicates in order to propose principles for an interpretation in the general case.

3.1 The Case of Crisp Predicates

When the predicates A and B in the quantified statement "*Q* B X are A" are crisp (Q being relative), the evaluation is provided by $\mu_Q(p)$ where p is the proportion of elements which are A among the elements which are B. We get :

$$T(p) \text{ where } p = |\{x \text{ in } X \text{ such that } x \text{ is } A \text{ and } B\}| \,/\, |\{x \text{ in } X \text{ such that } x \text{ is } B\}|. \quad (4)$$

3.2 General Case

In the general case, the proportion p of (4) is a ratio between two cardinalities of fuzzy set.

We consider to define these cardinalities of fuzzy sets as gradual numbers (FGCount) and to carry out the computation of the tolerance (according to (4)) using the extended arithmetic defined on gradual numbers (as defined in [5, 6]). In this context, the result can be either a fuzzy truth value or a scalar truth value (i.e. a degree set in [0, 1]).

4 A Fuzzy Truth Value for the Evaluation

Section 4.1 introduces the cardinalities of fuzzy sets (gradual numbers), this definition being a key point for the evaluation of quantified statements. Section 4.2 considers the evaluation of quantified statements of type "*Q B* X are *A*", where *Q* is relative.

4.1 Gradual Numbers

When F is a fuzzy set, the cardinality c is described by a gradual integer given by:

$$\forall e \in \mathbb{N}, \mu_c(e) = \sup \{\alpha \,/\, |F_\alpha| \geq e\},$$

where F_α is the α-cuts of fuzzy set F ($F_\alpha = \{x \,/\, \mu_F(x) \geq \alpha\}$). This cardinality is proposed by Zadeh [7] and called the FGCount(F). The degree $\mu_c(e)$ expresses the extent to which there is at least *e* elements in fuzzy set F. The gradual integer c is a normalized convex conjunctive fuzzy set which is non decreasing. The set of gradual integers is denoted by $\mathbb{N}_f$, which can be equivalently defined as the set of FGCount()'s.

The α level cut of a gradual integer c is denoted by $c(\alpha)$ and is defined as the cardinality of the α level cut of any fuzzy sets having c as fuzzy cardinality. The value $c(\alpha)$ is also the highest integer value appearing in the description c associated with a degree at least equal to α.

4.2 Quantified Statements of Type "Q B X Are A" Where Q Is Relative

In this case, we need to compute T(p - k) where p = c/d such that :

- c is the cardinality of the fuzzy set A∩B(X) made of elements from X which satisfy fuzzy condition *A* and condition *B* ($\forall$x in X, $\mu_{A\cap B(X)} = \min(\mu_A(x)\,,\, \mu_B(x))$),

- d is the cardinality of the fuzzy set B(X) made of elements from X which satisfy fuzzy condition *B*.

The fuzzy rational number [6] c/d is defined by the couple (c, d). A canonical representation for c/d is [6] :

$$\forall \alpha \in [0, 1], p(\alpha) = c(\alpha)/d(\alpha).$$

This canonical definition is defined only when $d(\alpha) \neq 0$. The cardinality c (resp. d) being that of the fuzzy set A∩B(X) (resp. B(X)), we get :

$$\forall \alpha \in [0, 1], p(\alpha) = |A \bullet B(X)_\alpha| / |B(X)_\alpha|,$$

where $|B(X)_\alpha| \neq 0$. It means that $p(\alpha)$ is not defined for $\alpha > \max_{x \in X} \mu_B(x)$, and we can write :

$$\forall \alpha \in [0, \max_{x \in X} \mu_B(x)], p(\alpha) = |A \bullet B(X)_\alpha| / |B(X)_\alpha|.$$

A canonical representation for p-k is :

$$\forall \alpha \in [0, \max_{x \in X} \mu_B(x)], \Delta(\alpha) = p(\alpha) - k = |A \bullet B(X)_\alpha| / |B(X)_\alpha| - k.$$

The application of a predicate T on a fuzzy relative integer such as Δ gives a global satisfaction S defined by [6] :

$$\forall \alpha \in [0, \max_{x \in X} \mu_B(x)], \mu_S(\alpha) = T(\Delta(\alpha)) = T(|A \bullet B(X)_\alpha| / |B(X)_\alpha| - k)$$
$$= \mu_Q(|A \bullet B(X)_\alpha| / |B(X)_\alpha|).$$

The fuzzy truth value S expresses the satisfaction of each α-cut of A(X) and A∩B(X) with respect to the linguistic quantifier.

The value α is viewed as a quality threshold for the satisfactions with respect to *A* and *B*. When the minimum is chosen as a t-norm to define A∩B(X), the value of $\mu_S(\alpha)$ states that : "among the elements which satisfy *B* at least at level α, the proportion of elements x with $\mu_A(x) \geq \alpha$, is in agreement with *Q*" (since we have $A \cap B(X)_\alpha = A(X)_\alpha \cap B(X)_\alpha$). In other words, $\mu_S(\alpha)$ represents the truth value for "*Q B* X are *A*" when considering the two α-cuts $B(X)_\alpha$ and $A(X)_\alpha$ ($\mu_S(\alpha)$ is the truth value of "*Q* elements in $B(X)_\alpha$ are in $A(X)_\alpha$"). The value $\mu_S(\alpha)$ represents the truth value for the quantified statements when considering the interpretations at level α of the two fuzzy sets. The advantage of this representation is to provide the different results given by the different interpretations of the fuzzy sets. As a consequence, this result has a clear meaning and can be the base for further processing.

In addition, the fuzzy truth value S is not defined when $\alpha > \max_{x \in X} \mu_B(x)$ but a simple assumption is to assume that $\mu_S(\alpha) = 0$ in that case.

Example 2. We consider the statement "*about half B* X are *A*" where $X = \{x_1, x_2, x_3, x_4\}$ and the linguistic quantifier *about half* from example 1. The different satisfactions with respect to *B* and *A* are described by Fig. 3. We obtain the fuzzy truth value given by Fig. 4. As an example, we get $\mu_S(0.6) = 1/3$ because $|A \cap B(X)_{0.6}| / |B(X)_{0.6}| = 2/3$ and $\mu_Q(2/3) = 1/3$. The truth value of the statement "*about half* elements in {x such that $\mu_B(x) \geq 0.6$ } are in {x such that $\mu_A(x) \geq 0.6$}" is 1/3.

	x_1	x_2	x_3	x_4
$\mu_B(x_i)$	1	0.9	0.7	0.3
$\mu_A(x_i)$	0.8	0.3	1	1
$\mu_{A\cap B}(x_i)$	0.8	0.3	0.7	0.3

Fig. 3. The satisfaction to *B* and *A*

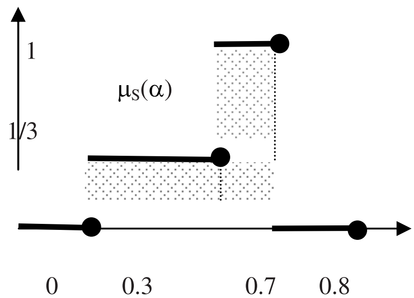


Fig. 4. The fuzzy truth value for "*about half B* X are *A*"

5 A Scalar Value for the Evaluation

The fuzzy truth value S computed in the previous section represents the different satisfactions of the different α-cuts with respect to the linguistic quantifier.

This fuzzy truth value can be defuzzified in order to obtain a scalar evaluation (set in [0,1]). Various interpretations can be associated to this defuzzification and we consider the following one (since it is the more natural) :

"the higher the scalar interpretation, the more α-cuts satisfies the constraint defined by the linguistic quantifier".

A scalar interpretation of 1 means that whatever is the chosen interpretation for A(x) and B(x) (i.e., whatever is the chosen quality threshold), the proportion of elements *A* among the *B* elements is **fully** in agreement with *Q*. Otherwise, the higher the scalar evaluation, the more there exists interpretations such that the proportion highly satisfies *Q*.

In section 5.1, we consider a quantitative defuzzification (since based on an additive measure – a surface) while in section 5.2 we consider a qualitative defuzzification (since based on a non additive process).

5.1 A Quantitative Approach

In this approach, the surface of the fuzzy truth value is delivered to the user. The scalar interpretation is then :

$$\delta = \int \mu_S(\alpha)\, d\alpha$$

Value δ is the area delimited by function μ_S. Since this function is a stepwise function, we get :

$$\delta = (\alpha_1 - 0) * \mu_S(\alpha_1) + (\alpha_2 - \alpha_1) * \mu_S(\alpha_2) + \ldots + (1 - \alpha_n) * \mu_S(1),$$

where the discontinuity points are $(\alpha_1, \mu_S(\alpha_1))$, $(\alpha_2, \mu_S(\alpha_2))$, ..., $(\alpha_n, \mu_S(\alpha_n))$ with $\alpha_1 < \alpha_2 < \ldots < \alpha_n$.

Example 3. We consider the statement "*about half B* X are *A*" of example 2 and the fuzzy truth value given by Fig. 4. We compute :

$$\delta = (0.7 - 0.3) * 1/3 + (0.8 - 0.7)*1 = 1/3.$$

This result is in accordance with our intuition since it seems that the proportion of elements which are *A* among the *B* elements is near to be 2/3 (with $\mu_Q(2/3) = 1/3$).

5.2 A Qualitative Approach

According to this approach, the scalar interpretation takes into consideration two aspects :

- a guaranteed (minimal) satisfaction value β associated to the α-cuts (β must be higher as possible),
- the repartition of β among the α-cuts (β should be attained by the most possible α-cuts).

Obviously, these two aspects are in opposition since, in general, the higher β, the smaller the repartition. The scalar interpretation δ reflects a compromise between these two aspects and we get :

$$\delta = \max_{\beta \text{ in } [0,1]} \min(\beta, \text{each}(\beta)),$$

where each(β) means "for each level α, $\mu_S(\alpha) \geq \beta$".

The truth value for each(β) can be a matter of degree and we propose to sum the lengths of intervals (of levels) where the threshold β is reached :

$$\text{each}(\beta) = \sum_{\substack{]\alpha_i,\alpha_j] \text{ such that} \\ \forall \alpha \in]\alpha_i,\alpha_j],\ \mu_S(\alpha) \geq \beta}} (\alpha_j - \alpha_i).$$

The higher each(β), the more numerous the levels α for which $\mu_S(\alpha) \geq \beta$. In particular, each(β) equals 1 means that for each level α, $\mu_S(\alpha)$ is larger than (or equal to) β.

In addition, from a computational point of view, the definition of δ needs to handle an infinity of values β. However, it is possible [3] to restrict computations to β values belonging to the set of "effective" $\mu_S(\alpha)$ values:

$$\delta = \max_{\{\beta \mid \exists\ \alpha \text{ such that } \beta = \mu_S(\alpha)\}} \min(\beta, \text{each}(\beta)),$$

Example 4. We consider the statement "*about half B* X are *A*" of example 2 and the fuzzy truth value given by Fig. 4. The values β to be considered are 1/3 and 1. Furthermore :

$$\text{each}(1/3) = 0.7, \qquad \text{each}(1) = 0.2.$$

We get $\delta = \max(\min(1/3, 0.7), \min(1, 0.2)) = 1/3$. A truth value of 1/3 for "*about half B* X are *A*" is coherent (see example 4).

6 Conclusion

This paper takes place at the crossword of flexible querying of relational databases using fuzzy sets and the fuzzy arithmetic introduced in [5, 6]. It shows that fuzzy arithmetic allows to evaluate quantified statements of type "*Q B* X are *A*". This evaluation is based on the different results provided by the different interpretations (α-cuts) of the two fuzzy sets (respectively of *A* and *B* elements).

The evaluation can be either a fuzzy truth value or a scalar value obtained by the defuzzification of that fuzzy value. Two types of scalar values can be distinguished : the first one corresponds to a quantitative view of the fuzzy value, the second one of a qualitative view. Our approach presents the advantage of providing a theoretical framework to evaluate "*Q B* X are *A*" statements. It is the first attempt to set this evaluation in the extended arithmetic described in [5, 6].

References

1. Bosc, P., Pivert, O.: Some Approaches for Relational Databases Flexible Querying, Journal of Intelligent Information Systems, 1, (1992) 323-354.
2. Bosc P., Pivert O.: SQLf: A relational database language for fuzzy querying. IEEE Transactions on Fuzzy Systems, vol. 3, (1995) 1-17.
3. Bosc, P., Liétard L.: A general technique to measure gradual properties of fuzzy sets, International Conference of the Fuzzy Sets Association (IFSA'05), (2005) 485-490.
4. Lietard L., Rocacher D.: A generalization of the OWA operator to evaluate non monotonic quantifiers", 4th Conference of the European Society for Fuzzy Logic and Technologie and 14ème Rencontres francophones sur la logique floues et ses applications, (EUSFLAT'05 and LFA'05), (2005) 183-188.
5. Rocacher D., Bosc P.: About Zf, the set of fuzzy relative integers, and the definition of fuzzy bags on Zf, Lecture Notes in Computer Science, LNCS 2715, Springer-Verlag, (2003) 95-102.
6. Rocacher D., Bosc P.: The set of fuzzy rational numbers and flexible querying, Fuzzy Sets and Systems, 155, 3, (2005) 317-339.
7. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages, Computer Mathematics with Applications, 9, (1983) 149-183.

Properties of Local Andness/Orness

Jozo J. Dujmović

Department of Computer Science, San Francisco State University, 1600 Holloway Ave, San Francisco, CA 94132, USA
jozo@sfsu.edu

Abstract. The goal of this paper is to investigate fundamental properties of local andness/orness. We analyze the distribution of local andness/orness in the unit hypercube from the standpoint of usability of these indicators in decision models.

1 Introduction

The local andness, orness, and their mean values were introduced in 1973 as indicators that are necessary as parameters of a generalized aggregator that enables continuous transition from conjunction to disjunction [5]. Andness indicates the degree of similarity of an aggregator and conjunction, and it was initially called the conjunction degree. Orness indicates the degree of similarity of an aggregator and disjunction, and it was initially called the disjunction degree. Andness and orness are parameters that are adjusted to attain desired properties of many decision models.

A generalized aggregator that combines conjunctive and disjunctive properties is called the *Generalized Conjunction/Disjunction* (GCD) [10] and symbolically denoted $y = x_1 \Diamond \cdots \Diamond x_n$, $x_i \in I = [0,1]$, $i = 1,...,n$, $y \in I$. The andness ($\alpha \in I$) indicates the similarity between GCD and conjunction, and the orness ($\omega \in I$) indicates the similarity between GCD and disjunction; α and ω are complementary: $\alpha + \omega = 1$. The parameters $\alpha = 1,\ \omega = 0$ denote the full conjunction and $\alpha = 0,\ \omega = 1$ denote the full disjunction. If $\alpha < 0.5 < \omega$ then $x_1 \Diamond \cdots \Diamond x_n$ is called the *partial disjunction* and symbolically denoted $x_1 \nabla \cdots \nabla x_n$. If $\alpha = 0.5 = \omega$ then $x_1 \Diamond \cdots \Diamond x_n$ is called the *neutrality function*, and implemented as the arithmetic mean. If $\alpha > 0.5 > \omega$ then $x_1 \Diamond \cdots \Diamond x_n$ is called the *partial conjunction* and symbolically denoted $x_1 \Delta \cdots \Delta x_n$ [10].

Andness and orness can be defined in several ways, and nine of them are identified and compared in [12,13]. The initial definition [5] introduced the local andness α_ℓ and the local orness ω_ℓ as follows:

$$\begin{aligned} x_1 \Diamond \cdots \Diamond x_n &= \omega_\ell (x_1 \vee \cdots \vee x_n) + (1 - \omega_\ell)(x_1 \wedge \cdots \wedge x_n) \\ &= (1 - \alpha_\ell)(x_1 \vee \cdots \vee x_n) + \alpha_\ell (x_1 \wedge \cdots \wedge x_n) \\ &= \omega_\ell (x_1 \vee \cdots \vee x_n) + \alpha_\ell (x_1 \wedge \cdots \wedge x_n) \end{aligned}$$

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 54–63, 2007.
springerlink.com

Consequently, the local andness and orness are indicators that can be computed in all points of $[0,1]^n$:

$$\alpha_\ell(x_1,...,x_n) = \frac{(x_1 \vee \cdots \vee x_n) - (x_1 \Diamond \cdots \Diamond x_n)}{(x_1 \vee \cdots \vee x_n) - (x_1 \wedge \cdots \wedge x_n)}$$

$$\omega_\ell(x_1,...,x_n) = \frac{(x_1 \Diamond \cdots \Diamond x_n) - (x_1 \wedge \cdots \wedge x_n)}{(x_1 \vee \cdots \vee x_n) - (x_1 \wedge \cdots \wedge x_n)}$$

$$\alpha_\ell(x_1,...,x_n) \in I, \quad \omega_\ell(x_1,...,x_n) \in I, \quad \alpha_\ell(x_1,...,x_n) + \omega_\ell(x_1,...,x_n) = 1$$

$$x_1 \vee \cdots \vee x_n \neq x_1 \wedge \cdots \wedge x_n$$

The local andness is not defined for $x_1 = ... = x_n$ and has a distribution that can substantially vary inside the input space $[0,1]^n$; an analysis of local andness was recently provided by Fernández Salido and Murakami [14]. The variations of local andness/orness are not desirable for decision analysts who have to select a given level of andness/orness for each aggregator they use in decision models. These problems are usually avoided by using either the mean local andness $\overline{\alpha}_\ell$ [5] or the global andness α_g [6,7]:

$$\overline{\alpha}_\ell = \int_0^1 dx_1 ... \int_0^1 \alpha_\ell(x_1,...,x_n)dx_n \ , \quad \overline{\omega}_\ell = \int_0^1 dx_1 ... \int_0^1 \omega_\ell(x_1,...,x_n)dx_n$$

$$\overline{\alpha}_\ell \in I, \quad \overline{\omega}_\ell \in I, \quad \overline{\alpha}_\ell + \overline{\omega}_\ell = 1 \ .$$

$$\alpha_g = \frac{\overline{x_1 \vee ... \vee x_n} - \overline{x_1 \Diamond ... \Diamond x_n}}{\overline{x_1 \vee ... \vee x_n} - \overline{x_1 \wedge ... \wedge x_n}} = \frac{n - (n+1)(\overline{x_1 \Diamond ... \Diamond x_n})}{n-1}$$

$$\omega_g = \frac{\overline{x_1 \Diamond ... \Diamond x_n} - \overline{x_1 \wedge ... \wedge x_n}}{\overline{x_1 \vee ... \vee x_n} - \overline{x_1 \wedge ... \wedge x_n}} = \frac{(n+1)(\overline{x_1 \Diamond ... \Diamond x_n}) - 1}{n-1}$$

$$\overline{x_1 \Diamond ... \Diamond x_n} = \int_0^1 dx_1 \cdots \int_0^1 (x_1 \Diamond ... \Diamond x_n) dx_n$$

$$\overline{x_1 \wedge ... \wedge x_n} = 1/(n+1), \quad \overline{x_1 \vee ... \vee x_n} = n/(n+1), \quad \alpha_g \in I, \quad \omega_g \in I, \quad \alpha_g + \omega_g = 1$$

Analytic computation of $\overline{\alpha}_\ell$ can be done using the Marichal method [18] and the computation of α_g can be based on techniques presented in [11]. However, analytic results are available only in special cases of GCD, and generally we must rely on numerical computation. Accurate numeric computation of $\overline{\alpha}_\ell$ and α_g is also not simple because of the following reasons:

- For some implementations of GCD there are difficulties in accurate computation of the value of GCD (e.g. see [20])
- $\alpha_\ell(x_1,...,x_n)$ has discontinuities inside $[0,1]^n$
- GCD in many cases has infinite derivatives $\partial(x_1\Diamond\cdots\Diamond x_n)/\partial x_i = +\infty$ at the edges of $[0,1]^n$.

Regardless of the selected values of $\overline{\alpha}_\ell$ or α_g, each aggregation occurs at some specific value of local andness/orness, and understanding the variations of local andness/orness is indispensable for building and using decision models. In particular, the discontinuities of local andness and the sensitivity of local andness to distance from idempotency are fundamental properties that can affect the way we use aggregators in decision models. Therefore, there is a clear interest to investigate and evaluate properties of local andness/orness, and this is our primary goal in this paper.

2 Implementations of GCD

GCD is a mean ($x_1 \wedge\cdots\wedge x_n \le x_1\Diamond\cdots\Diamond x_n \le x_1 \vee\cdots\vee x_n$) and all means can be interpreted as continuous logic functions [8,10]. If $x_1 = \cdots = x_n = x$ then $x \wedge\cdots\wedge x = x \vee\cdots\vee x = x$ and the consequence of $x \le x\Diamond\cdots\Diamond x \le x$ is the idempotency $x\Diamond\cdots\Diamond x = x$. Those means that have a continuously adjustable parameter enabling the transition from $x_1 \wedge\cdots\wedge x_n$ to $x_1 \vee\cdots\vee x_n$ are the best candidates for implementing the GCD aggregator. Such means are analyzed in [2,3,15,19]. A detailed presentation of mathematical aspects of the whole area of aggregation operators can be found in [17] and [4]. Mathematical properties of aggregation operators that are used for the synthesis of judgments (closely related to system evaluation) are analyzed in [1]. In this paper we will analyze the properties of local andness using the GCD implementation based on weighted power means (WPM):

$$M(x_1,...,x_n;r) = \begin{cases} (W_1x_1^r + ... + W_nx_n^r)^{1/r}, & 0 < |r| < +\infty \\ x_1^{W_1}\cdots x_n^{W_n}, & r = 0 \\ x_1 \wedge\cdots\wedge x_n, & r = -\infty \\ x_1 \vee\cdots\vee x_n, & r = +\infty \end{cases}$$

$$W_i > 0, \quad i = 1,...,n, \quad \sum_{i=1}^{n} W_i = 1$$

Other implementations of GCD include exponential, logarithmic, and counter-harmonic means [3], AIWA [16] and aggregators with positional weights, OWA [21,22] and ItOWA [9]. OWA aggregator has a remarkable feature $\overline{\omega}_\ell = \omega_g$. In [18] Marichal showed that this feature holds for any Choquet integral with n variables.

3 Distribution of Local Andness

For each form of GCD aggregator the local andness can be computed in any point of $[0,1]^n$. Let us define the local andness probability distribution function $P_\alpha(t)$ and its density $p_\alpha(t)$ as follows:

$$P_\alpha(t) = \Pr[\alpha_\ell(x_1,...,x_n) < t], \quad p_\alpha(t) = dP_\alpha(t)/dt, \quad (x_1,...,x_n) \in [0,1]^n$$

These functions characterize the distribution of global andness for any number of variables. In the case of power means the probability density function for n=3 is shown in Fig. 1. For convenience, we denote nine equidistant aggregators that have the global andness $\alpha_g = 0$, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1 using respectively symbols D, D+, DA, D-, A, C-, CA, C+, and C. In the case of arithmetic mean (A) the distribution is uniform in the range [1/3, 2/3]. If $r \le 0$, the local andness is distributed from 1/3 to 1. If $r > 1$, the maximum local andness is 2/3. Figure 1 shows that two inconvenient (and counterintuitive) properties: (1) the distributions significantly overlap, and (2) at some points the partial conjunction aggregators have $\alpha_\ell < 0.5$ and partial disjunction aggregators have $\alpha_\ell > 0.5$. In some regions of $[0,1]^n$ (in the vicinity of the idempotency line $x_1 = \cdots = x_n$) both the partial conjunction and the partial disjunction have the local andness $\alpha_\ell \approx 0.5$, i.e. their behavior is similar to the arithmetic mean (this property is discussed in Section 4).

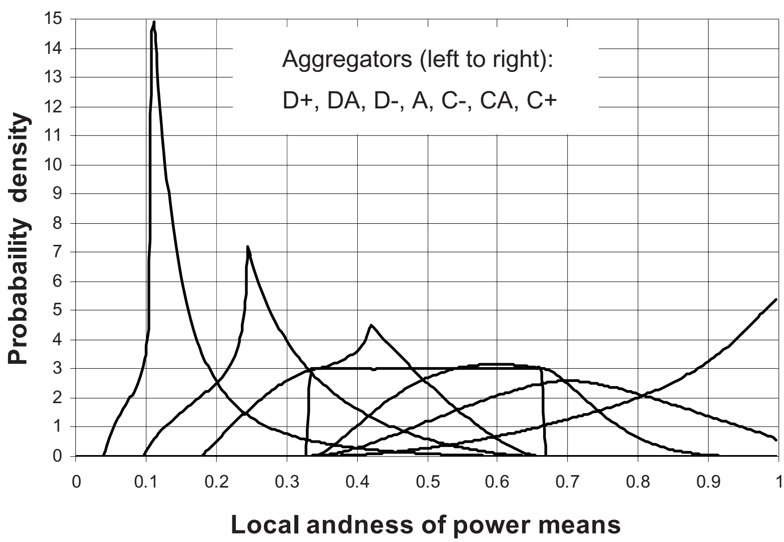


Fig. 1. Probability density of the local andness of power means of 3 variables

The local orness was recently investigated by Fernández Salido and Murakami [14]. Their *orness distribution function* is equivalent to the local orness and their *orness average value* is equivalent to the mean local orness. In order to provide graphical representability they focused on the case of two variables.

4 Discontinuities of Local Andness/Orness

Continuous transition from conjunction to disjunction generates a spectrum of aggregators with the arithmetic mean in the central position. Furthermore, in the vicinity of the idempotency line all aggregators behave similarly to the arithmetic mean. These are reasons why we must first investigate the behavior of local andness of the arithmetic mean.

In the simplest case of two variables the local andness of the arithmetic mean is constant, ½. If n=3, we have

$$\alpha_\ell(x,y,z) = \frac{(x \vee y \vee z) - (x+y+z)/3}{(x \vee y \vee z) - (x \wedge y \wedge z)}, \quad x \vee y \vee z \neq x \wedge y \wedge z$$

$$\alpha_\ell(0,0,z) = \frac{z - z/3}{z} = \frac{2}{3} = \alpha_\ell(0,y,0) = \alpha_\ell(x,0,0)$$

$$\alpha_\ell(1,1,z) = \frac{1-(2+z)/3}{1-z} = \frac{1}{3} = \alpha_\ell(1,y,1) = \alpha_\ell(x,1,1)$$

$$\alpha_\ell(x,x,z) = \frac{x \vee z - (2x+z)/3}{x \vee z - x \wedge z} = \begin{cases} \dfrac{z-(2x+z)/3}{z-x} = \dfrac{2}{3}, & z > x \\ \dfrac{x-(2x+z)/3}{x-z} = \dfrac{1}{3}, & z < x \end{cases}$$

$$\alpha_\ell(x,y,y) = \begin{cases} 2/3, & x > y \\ 1/3, & x < y \end{cases}, \quad \alpha_\ell(z,y,z) = \begin{cases} 2/3, & y > z \\ 1/3, & y < z \end{cases}$$

Therefore, along the idempotency line $x=y=z$ the local andness converges to different values at different directions, producing a discontinuity shown in Fig. 2.

In the general case of the arithmetic mean of n variables the local andness is

$$\alpha_\ell(x_1,\dots,x_n) = \frac{(x_1 \vee \cdots \vee x_n) - (x_1 + \cdots + x_n)/n}{(x_1 \vee \cdots \vee x_n) - (x_1 \wedge \cdots \wedge x_n)}, \quad x_1 \vee \cdots \vee x_n \neq x_1 \wedge \cdots \wedge x_n$$

The discontinuity of local andness of the arithmetic mean in the vicinity of idempotency increases if we increase the number of dimensions:

$$\alpha_\ell^+(x,\dots,x+\delta,\dots,x) = \frac{(x+\delta)-(nx+\delta)/n}{(x+\delta)-x} = 1 - \frac{1}{n}$$

$$\alpha_\ell^-(x,\ldots,x-\delta,\ldots,x) = \frac{x-(nx-\delta)/n}{x-(x-\delta)} = \frac{1}{n}, \quad \delta > 0$$

For n=2,3,4,5 the corresponding discontinuity gap (and the range of local andness) is $\rho=1\text{-}2/n$ = 0, 1/3, ½, 3/5, and is close to 1 for large n (the gap in Fig. 1 is 1/3)

There are $n(n\text{-}1)/2$ cases where two variables are different, and the remaining n-2 variables are their arithmetic mean. In such cases we have

$$x_i \neq x_j,\ \ i \in \{1,\ldots,n\},\ \ j \in \{1,\ldots,n\},\ \ n > 2,$$
$$x_k = (x_i + x_j)/2,\ \ k = 1,\ldots,n,\ \ k \neq i,\ \ k \neq j$$
$$\alpha_\ell(x_1,\ldots,x_n) = \frac{(x_1 \vee \cdots \vee x_n) - (x_1 + \cdots + x_n)/n}{(x_1 \vee \cdots \vee x_n) - (x_1 \wedge \cdots \wedge x_n)}$$
$$= \frac{(x_i \vee x_j) - \left[x_i + x_j + (n-2)(x_i + x_j)/2\right]/n}{(x_i \vee x_j) - (x_i \wedge x_j)} = \frac{(x_i \vee x_j) - (x_i + x_j)/2}{(x_i \vee x_j) - (x_i \wedge x_j)} = \frac{1}{2}$$

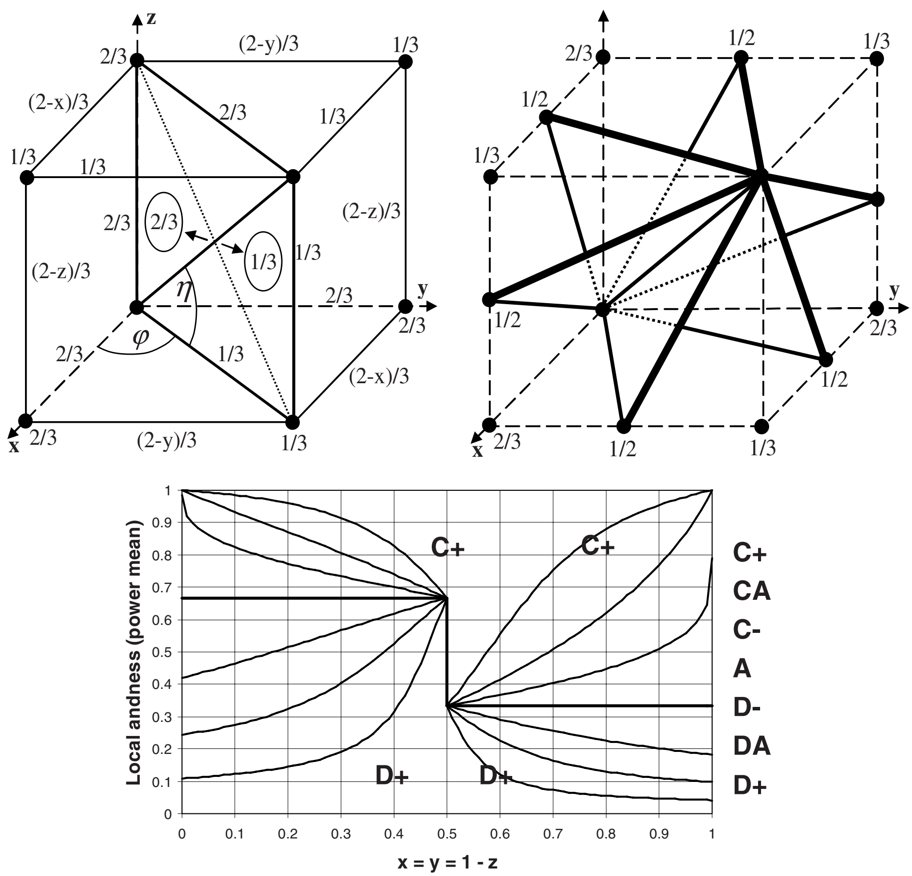


Fig. 2. Local andness of arithmetic mean for n=3 and α-discontinuity for GCD

Therefore, in the case of n variables ($n > 2$) the behavior of local andness of the arithmetic mean is similar to the behavior for n=3. Inside $[0,1]^n$ there are $n(n-1)/2$ hyperplanes where $\alpha_\ell = 1/2$, and between them the local andness is alternatingly attaining $n(n-1)/2$ extreme values of $1/n$, and $n(n-1)/2$ extreme values of $1-1/n$.

5 Local Andness in the Vicinity of Idempotency Line

In the case of three variables, in the vicinity of the idempotency line x=y=z the local andness has large variations going from the minimum 1/3 through the mean 1/2, to the maximum 2/3. These variations can be investigated if we move along a closed path around the central idempotency point x=y=z=1/2.

As expected from Figure 2, in any closed path around the idempotency line the local andness six times takes the mean value ½, three times the minimum value 1/3, and three times the maximum value 2/3. In the close vicinity of the idempotency line other characteristic power means (e.g. geometric, harmonic and quadratic), behave in the same way as the arithmetic mean (Fig. 3). All other GCD aggregators have similar behavior.

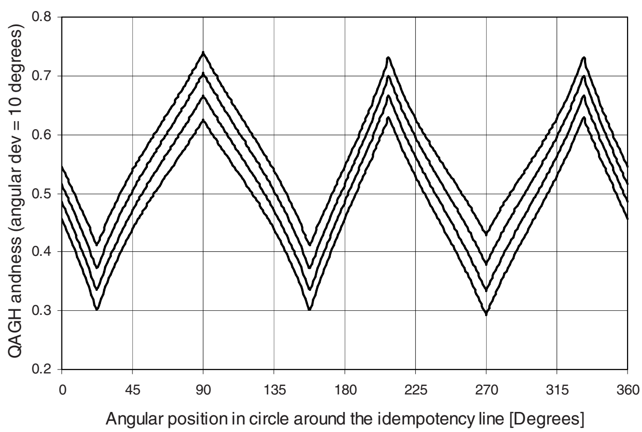


Fig. 3. Andness of harmonic, geometric, arithmetic, and quadratic mean (top curve = H > G > A > Q = bottom curve)

6 Similarity of GCD Aggregators and the Arithmetic Mean

In the vicinity of idempotency all averaging aggregators generate similar results and behave in a similar way, i.e. as the arithmetic mean. The mean distance from the idempotency line $x_1 = ... = x_n$ can be defined as follows:

$$D(n)=\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|x_i-x_j|,$$

$$D_{\max}(n)=\frac{\lfloor (n+1)/2\rfloor}{2\lfloor (n+1)/2\rfloor-1},$$

$$D_{norm}(n)=D(n)/D_{\max}(n),\quad 0\le D_{norm}(n)\le 1,$$

$$\bar{D}(n)=\frac{2}{n(n-1)}\int_0^1 dx_1\cdots\int_0^1\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|x_i-x_j|\,dx_n=\frac{1}{3},\quad n>1$$

The mean values of local andness for aggregators D+, DA, D-, A, C-, CA, C+ based on power means are shown in Figure 4. Regardless the level of global andness, all partial conjunctions and partial disjunctions based on adjustable means become similar to arithmetic mean (logical neutrality aggregator) in the vicinity of the idempotency line (i.e. for similar values of inputs). This property holds for any number of variables.

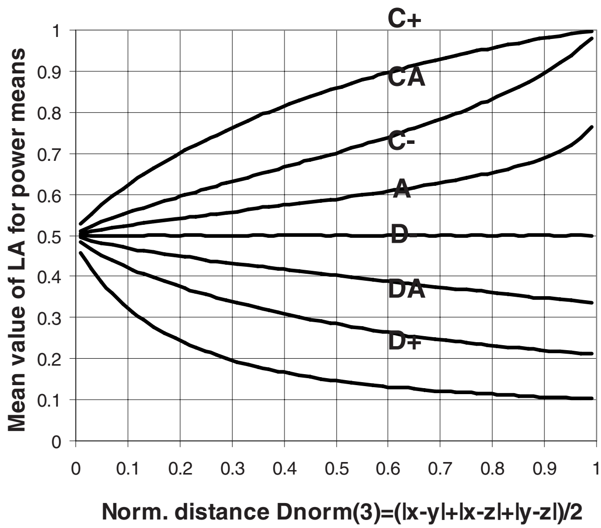


Fig. 4. Mean values of local andness for power means as functions of normalized distance from the arithmetic mean

This is not a property of aggregators with positional weights, such as OWA and ItOWA. In the case of OWA the mean value of local andness is constant for any value of the normalized distance D_{norm}. OWA has regular distribution of equidistant parallel lines, and ItOWA has parallel lines that are similar to OWA, but not equidistant.

The sensitivity of andness to the distance from idempotency does not seem to be a decisive disadvantage or advantage compared to aggregators that provide the same average andness in all regions of $[0,1]^n$. The selection of aggregators in applications is based on other factors, such as the ability to express variable relative importance (weights) and the ability to model mandatory requirements and building asymmetric aggregators [10]. The fact that for small differences of inputs some means are similar to the arithmetic mean is not a source of practical problems, since the resulting variations of aggregated values are very small and cannot significantly affect final results generated by decision models. On the other hand, regular and constant properties, such as those provided by OWA, contribute to controllability of decision models, and are certainly positive.

7 Conclusions

In this paper we identified and investigated several inconvenient properties of local andness and orness:

- The distribution of andness/orness inside $[0,1]^n$ is frequently nonuniform with large variations.
- There are points in $[0,1]^n$ where the local andness of conjunctive aggregators is less than ½ and the local orness of disjunctive aggregators is less than ½.
- The distribution of local andness/orness has a discontinuity at the idempotency line.
- For all aggregators based on adjustable (parameterized) means the mean local andness in the vicinity of idempotency is close to ½, which is the mean andness of the arithmetic mean. Consequently, in the vicinity of idempotency all averaging aggregators behave as the arithmetic mean.

The complexity of distribution of local andness/orness inside $[0,1]^n$ eliminates (or significantly reduces) the possibility of pointwise adjustment of andness/orness during the design of decision models. Designers of decision models must rely on average andness indicators, primarily on the global andness. The decision analysts must think globally, but the individual aggregators act locally, depending on the current position in the $[0,1]^n$ space. Therefore, the awareness of local properties of andness/orness is a necessary component in the process of designing and using decision models based on aggregators.

References

[1] Aczél, J., C. Alsina, On Synthesis of Judgements. Socio-Econ. Plann. Sci. Vol. 20, No. 6, pp. 333-339 (1986)
[2] Bullen, P.S., D.S. Mitrinović, and P.M. Vasić, Means and Their Inequalities. D. Reidel (1988)
[3] Bullen, P.S., Handbook of Means and Their Inequalities. Kluwer (2003)
[4] Calvo, T., A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators: Properties, classes and construction methods. In Studies in Fuzziness and Soft Computing, eds. T. Calvo, G. Mayor, and R. Mesiar, Vol 97, pp. 3-104, Springer (2002).

[5] Dujmović, J.J., A Generalization of Some Functions in Continuous Mathematical Logic – Evaluation Function and its Applications, (In Serbo-Croatian). Proceedings of the Informatica Conference, paper d27, Bled, Yugoslavia (1973).
[6] Dujmović, J.J., Two Integrals Related to Means. Journal of the University of Belgrade EE Dept., Series Mathematics and Physics, No. 412 - No. 460, pp. 231-232 (1973)
[7] Dujmović, J.J., Weighted Conjunctive and Disjunctive Means and their Application in System Evaluation. Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics, No. 483, 1974, pp. 147-158 (1974).
[8] Dujmović, J.J., Extended Continuous Logic and the Theory of Complex Criteria. Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics, No. 537, 1975, pp. 197-216.
[9] Dujmović, J.J., Preferential Neural Networks. Chapter 7 in Neural Networks - Concepts, Applications, and Implementations, Vol. II. Edited by P. Antognetti and V. Milutinović, Prentice-Hall Advanced Reference Series, Prentice-Hall, pp. 155-206 (1991)
[10] Dujmović, J.J., Continuous Preference Logic for System Evaluation. Proceedings of Eurofuse 2005, edited by B. De Baets, J. Fodor, and D. Radojević, pp. 56-80. ISBN 86-7172-022-5, Institute "MihajloPupin", Belgrade, 2005.
[11] Dujmović, J.J., Seven Flavors of Andness/Orness. Proceedings of Eurofuse 2005, edited by B. De Baets, J. Fodor, and D. Radojević, pp. 81-92. ISBN 86-7172-022-5, Institute "MihajloPupin", Belgrade, 2005.
[12] Dujmović, J.J., A Comparison of Andness/Orness Indicators. Proceedings of the 11th Information Processing and Management of Uncertainty international conference (IPMU 2006), pp. 691-698, Les Cordeliers, Edition E.D.K., ISBN 2-84254-112-X, Paris, 2006.
[13] Dujmović, J.J., Nine Forms of Andness/Orness. Proceedings of the IASTED International Conference on Computational Intelligence. ISBN Hardcopy: 0-88986-602-3 / CD: 0-88986-603-1, pp. 276-281, 2006.
[14] Fernández Salido, J.M. and S. Murakami, Extending Yager's orness concept for the OWA aggregators to other mean operators. Fuzzy Sets and Systems 139, pp. 512-542, 2003.
[15] Gini, C. et al., Means. Italian original, *Le Medie*, Milano 1958. Russian translation, *Средние Величины* published by Statistika, Moscow (1970).
[16] Larsen, H.L.: Efficient Andness-directed Importance Weighted Averaging Operators, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (Suppl.): 67–82 (2003).
[17] Marichal, J.-L., Aggregation Operators for Multicriteria Decision Aid. Ph.D. Dissertation, Universite de Liege, 1998.
[18] Marichal, J.-L., Multivariate integration of functions depending explicitly on the minimum and the maximum of the variables. Submitted for publication (2006).
[19] Mitrinović, D.S. and J.E. Pečarić, Mean Values in Mathematics (in Serbo-Croatian), Naučna Knjiga Publishers, Belgrade (1989).
[20] Slavić, D.V. and Dujmović, J.J., Numerical Computation of Weighted Power Means. Journal of the University of Belgrade, EE Dept., Series Mathematics and Physics, No. 485, pp. 167-171. (1974).
[21] Yager, R.R., On Ordered Weighted Averaging Aggregation Operators in Multi-Criteria Decision Making. IEEE Trans. SMC 18, pp. 183-190 (1988).
[22] Yager, R.R. and J. Kacprzyk (Editors), The Ordered Weighted Averaging Operators – Theory and Applications. Kluwer Academic Publishers (1997).

A New Approach for Boolean Query Processing in Text Information Retrieval

Leemon Baird and Donald H. Kraft

Department of Computer Science, U.S. Air Force Academy,
USAFA, CO 80840, USA
leemon.baird@usafa.af.mil,donald.kraft@usafa.af.mil

Abstract. The main objective of an information retrieval system is to be effective in providing a user with relevant information in response to a query. However, especially given the information explosion which has created an enormous volume of information, efficiency issues cannot be ignored. Thus, to be able to quickly process lists of documents that have the keywords stated in a given query assigned/indexed to them by merging via the Boolean logic of the query is essential in a Boolean query system. A new algorithm, based loosely on concurrent codes, is developed and discussed.

1 Introduction

There is a need for good information retrieval, especially in this, the information age. Such systems need to be effective in providing relevant information without burdening the user with too much data. Thus, performance measures such as recall and precision help one determine how effective an information retrieval system is.

Information retrieval systems also need to be efficient, especially in terms of time. Efficiency considerations are vital in the information explosion era where huge volumes of information make it imperative to find the right information in a timely manner. Boolean query mechanisms are quite common, even today, in modern retrieval systems. Search engines such as Google and digital libraries such as the one at Stanford University [Stanford2006] provide such mechanisms. Users tend to use simple queries, but some users can and do incorporate Boolean logic connectives (e.g., and, or, and not) into their queries [Spink2001]. Thus, it behooves us to find the most efficient mechanism possible to search large collections of text documents to find those documents that satisfy Boolean queries.

In addition, some text retrieval systems have incorporated weights to indicate the importance of terms in describing document contents (indexing) and in terms of describing queries [Meadow2000]. These weights have been interpreted as coordinates for documents and queries in a vector space and as probabilities, but can also be interpreted as fuzzy set membership functions. This, too, will have an impact on efficiency, so that an efficient mechanism to process weighted Boolean queries is needed.

In addition, data mining is a field of increasing importance lately, including security issues such as identifying terrorists and keeping information secure. Such data mining often involves complex searches of extremely large databases that combine

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 64–70, 2007.
springerlink.com

both structured data (e.g., relational databases) and unstructured data (e.g., text documents and emails). In unstructured document retrieval, the problem is often to find documents that contain certain keyword combinations, as described by a complex Boolean expression. In the structured case, the problem is generally to find records that contain certain attributes, as described by a complex Boolean expression. In both cases, the need for efficiency, in terms of both time and space (i.e., computer memory) is obvious.

2 Background

There has been but a little work of late in trying to improve the efficiency of query processing of Boolean queries using AND, OR, and NOT operators on indexed textual information retrieval systems. Consider the problem of retrieving documents in response to a query q = t1 AND t2, where t1 and t2 are given terms (e.g., t1 = "NLP" and t2 = "IR") that is posed to a collection of N documents. Suppose that an index has been set up as an inverted file [Meadow2000]. Suppose further that t1 has n1 documents with the tem t1 associated with them, while t2 has n2 documents with the term t2 associated with them, i.e., the lengths of the lists for terms t1 and t2 are n1 and n2, respectively. Suppose also that n3 the number of documents that have both terms t1 and t2 associated with them.

Zobel notes that to process query q, if the index is totally resident on the disk, a reasonable assumption, that one must fetch both lists which has O(n1+n2) time, while a straightforward merge would then cost an additional O(min(n1,n2)) [Zobel2006] . Zobel also notes that one could get O(√n1+√n2) if one has random access to the lists and if the sum of the square roots of the list lengths is greater than n3 by employing skiplists.

Baeza-Yates [Baeza-Yates2004] offers a fast set intersection algorithm that assumes that the two lists of documents for each term are sorted. His algorithm involves a hybrid of binary search, as his original problem involves matching a query that is a multiset of terms against a document collection that is a larger multiset of terms. He arrives at a good average case by not inspecting all elements. Still, he is dealing with O(min(n1,n2) log (max(n1,n2))).

Work has been done on using comparison-based insertion and interpolation search for intersecting large, ordered sets [Barbay2006, Demaine2003, Demaine2001, Demaine2000]. More specifically, the authors look at Boolean query operators (union and difference, as well as intersection) for large sets of text records, using adaptive algorithms based on binary search.

The bottom line is that most results lead to algorithms with algorithmic complexity O(min(n1,n2)), while some hierarchical methods can yield O(√n1+√n2), but with high overhead. We desire an algorithm that is linear, say O(n3).

3 The Algorithm

We employ a novel data structure and algorithm for speeding Boolean searches, optimizing for the case where the number of documents satisfying the query is small compared to the number of documents containing terms used in the query. The algorithm works for arbitrary Boolean queries, including arbitrarily-nested parentheses. The data structures are based on BBC codes, a recent development in concurrent code theory, a new branch of coding theory that has been developed recently for a very different application (wireless jam-resistant communication) [Baird2007].

Assume every document is assigned a random, *n*-bit binary number as its identifier. For each possible search term, there will be a list of documents IDs containing that term. These are stored in a BBC structure, which is something like a Bloom filter. However, in a Bloom filter it is difficult to read out all the items stored in it. In a BBC structure, it is easy to do so. This difference allows the structure to optimize queries where the final set of satisfying documents is very small compared to the number of documents containing each term.

There is a separate BBC structure for each search term. The structure is a 1D array of bits. Initially, all bits are set to zero. Then each document ID is added to the structure. An ID is added to the structure by setting a bit to 1 in a location determined by the hash of each prefix of the document ID. For example, if the document ID is 11010 then the prefixes are {1, 11, 110, 1101, 11010). Hashing them gives the set {H(1), H(11), H(1110), H(1101), H(11010)}. Any hash function can be used, as long as it returns values in the range of $0 \ldots 2^m$-1, where *m* is the number of bits in the BBC structure. A simple approach is to use the first *m* bits of a SHA-1 hash or MD5 hash. In this example, 5 bits in the vector are set to 1, those in positions H(1) through H(11010). Contrast this with a Bloom filter, where only the bit in position H(11010) would be set. Figure 1 shows the structure that would be built for term A, assuming there are only two documents containing A, and their IDs are 11010 and 01100. The location of each 1 bit is chosen by a hash function. The locations corresponding to 11010 are written above the rectangle, and those for 10100 are written below it.

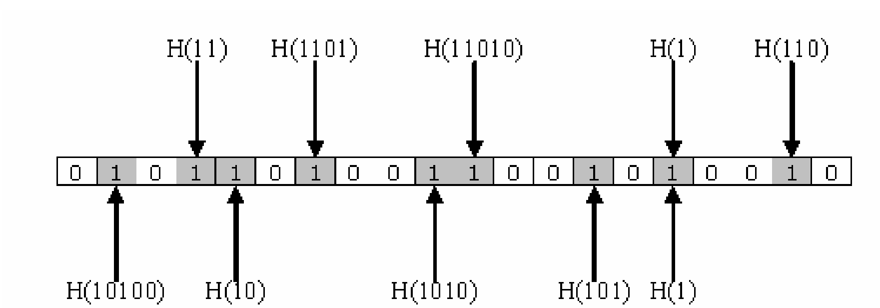


Fig. 1. A BBC structure for a given term, encoding the document IDs of all documents containing that term (in this example, 11010 and 10100)

Once all the document IDs have been inserted into the BBC structures for all the search terms, it is possible to perform queries efficiently. For example, consider the query "(A OR B) AND NOT C". We will first determine whether any of the document IDs satisfying this query begin with a 0. This is done by performing an OR of the bit in position H(0) in the BBC structures that were built for terms A and B. The result is then ANDed with the NOT of the bit in position H(0) in the BBC structure for term C.

If the result of this calculation is a 1, then we conclude that there is at least one satisfying document whose ID starts with 0. A similar check can be done for H(1). After performing these two checks, the *working set* of possible prefixes of document IDs satisfying the query is either {}, or {0}, or {1}, or {0,1}. The process can then be repeated to find the second bit of each satisfying document ID. If the first step yielded a set with 0 in it, then the second step will check positions H(00) and H(01). If the first step yielded a set with 1 in it, then the second step will check positions H(10) and H(11). This can continue until all the bits of all the document IDs have been found. In other words, a search is performed on a tree such as that in figure 2.

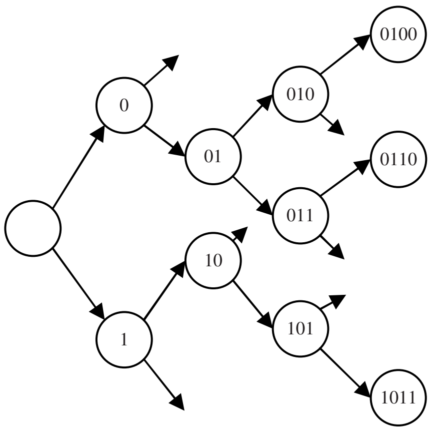


Fig. 2. A BBC decoding search that yields all documents that satisfy the query (in this case, documents 0110, 0100, and 1011).Upward branches add a 0, and downward add 1. Branches are pruned when the hash of a prefix doesn't satisfy the query.

The leaves that occur in the far right layer are the IDs of exactly those documents that satisfy the query. Initially, the set of potential prefixes is empty {}. After one iteration, searching the second layer from the left, the set is {0, 1}, because the bits at position H(0) for each term satisfied the Boolean expression of the query, and so did the bits at position H(1). At the next iteration, for layer 3, the set is only {01, 10} because the bits at positions H(01) and H(10) for each term satisfied the query, but the bits at H(00) and H(11) did not. In the third layer, the set grows to three elements, because the prefix 01 could be expanded to both 010 and 011, but the prefix 10 could only be expanded to 101. It couldn't be expanded to 100 because the bits in position H(100) for each term didn't satisfy the Boolean expression of the query.

If the final solution has very few satisfying documents, then the set of possibilities will shrink very quickly, and very little time will be expended on the entire search. Note that the search pruning has an interesting holographic property. In a traditional search for a query such as A AND B AND C AND D AND E, the search might proceed from left to right, processing each term in order to successively shrink the set of possible documents to return. In the approach using the BBC structure, the search reveals progressively more bits of the document ID. But at every step, all of the variables in the entire query are taken into account. This means that if the final answer is going to be a small set, then there will be substantial pruning at every stage, not just at the end. The entire Boolean expression will help prune at every step.

An example of how this would work is an optimization of the query (Attack OR Bomb) AND Car. A BBC packet (signature) could be created that contains the document IDs of every document containing the word “Attack”. A similar packet would be created for Bomb, and one for Car. These packets can then be combined bitwise. So the Attack and Bomb packets would be combined with a bitwise OR, then the result would be combined with the Car packet with a bitwise AND. The resulting packet encodes all of the documents that satisfy the query. Previous methods would allow such a packet to be constructed and queried for whether it contains a given document. But now with BBC coding, we can also extract all of the documents contained in the final packet. In other words, this is an extension of traditional signatures, which actually allows all of the documents satisfying the query to be found efficiently. In addition, by using a depth-first search of the BBC decoding tree, it is very efficient to extract just a single document (or just *n* documents) from the set of satisfying documents, with an amount of computation that is independent of the number of documents in the set.

4 Extending the Algorithm to Fuzzy Retrieval

In classical information retrieval models involving Boolean queries, one assumes that text records have been indexed with terms such that term t_j has or has not been assigned to document d_i. Thus, the “weight” of t_j on document d_i is either 1 (has been assigned) or 0 (has not been assigned). One can easily imagine an extension where these weights, rather than being in {0,1} can take on values in the interval [0,1]. One interpretation of these extended weights is to see them as fuzzy set membership functions, i.e., d_i is in the subset of documents “about” the concept(s) represented by t_j with a membership value if w_{ij}, the weight of t_j on d_i. These weights might be generated according to any of many forms of relevance feedback [Kraft1999].

The BBC-based approach can be adapted for fuzzy queries as well. If each (document, term) pair has a fuzzy weight, then some system can be chosen for combining fuzzy values to obtain an overall degree that any given document satisfies the given query. For example, using traditional fuzzy operators, the weights for each term can be combined with a maximum, minimum, or 1-x operator (for OR, AND, and NOT, respectively). The goal then is to find the best document (or top *n* documents) according to this function.

There are several ways to modify the algorithm to work in this case. The simplest method is a two-stage process. First treat every variable as crisp, noting only whether a document contains a term, not how many times it occurs. Use the algorithm from the previous section to find the set of documents that satisfy the crisp query. The fuzzy value for each document in the set can then be calculated, and the best document (or best *n* documents) can be returned. In the case where the solution set is small, this calculation will be very fast.

If the result of the crisp query is a large set, then this simple approach may be unacceptably slow. In that case, there is a more sophisticated algorithm that can be much more efficient. It will only generate one document (or *n* documents) rather than generating the entire set resulting from the crisp query.

For this fuzzy search algorithm, it is necessary to store an additional data structure. For each term T and each possible prefix P of a document ID, it will store the C_{min}(T,P) and C_{max}(T,P), which are the minimum and maximum weight for that term in all documents whose IDs start with that prefix. So, C_{max}(X, 011) for term X and the prefix 011 will store the count of how many times X appears in whichever document contains the most occurrences of X, from among all documents whose IDs start with the string 011. Similarly, C_{min}(X, 011) will store the minimum.

Given the C_{max} and C_{min} tables, it is possible to optimize the search process through the BBC decoding tree, by performing a uniform-cost, best-first search [Russell1995]. At any given point in the search, there will be a set of prefixes of IDs that might satisfy the query. The one with the most promising fuzzy value should be expanded next. For example, if the query is "(A OR B) AND NOT C", and a given prefix is 011, then the maximum value that any document could have starting with 011 would be min(max(C_{max}(A,011), C_{max}(B,011)), 1- C_{min}(C,011)). This was found by replacing OR with max, AND with min, and NOT with a subtraction from 1. The C_{max} structure was used for both the non-NOTed variables, and the C_{min} for the NOTed variable. This can be calculated for each prefix in the set, and the prefix with the highest value should be expanded next. For the crisp algorithm in the previous section, at any given time the working set always contains prefix strings of the same length. With this new approach, the working set will contain prefix strings of varying lengths, but which all have roughly the same maximum possible fuzzy weight.

Using this approach, the first time an entire ID is decoded, it's guaranteed to have the highest upper bound of all the IDs considered. As a heuristic, this document could simply be returned, and the search could end. But if it is desired to find the document with the highest possible fuzzy value, then additional work should be performed. First, the true fuzzy value for this document should be calculated. Then, any prefix in the working set that has an upper bound lower than this document should be deleted. Finally, the search should continue using all the remaining prefixes (if any). During the remainder of the search, a prefix can always be pruned if its maximum value is less than the best value already achieved. When no prefixes remain, the best answer discovered so far is guaranteed to be the best document possible for that query.

5 Future Work-Testing

In information retrieval research, one must go beyond algorithmic complexity and purely theoretical analysis by testing new algorithms against standard data. Such standard data must include textual documents, queries, and answer sets. The National Institute of Standards and Technology (NIST) has available such standard test data, known as the TREC or Tipster data collection, which would be employed to verify the improvements possible in the new algorithm. TREC collections vary in data type, query types, and data volumes. For example, some sample data sets include genomic, legal, and government oriented collections. Regardless of the data sets, they all provide document relevance judgments for each query listed. To demonstrate scalability, we plan in the future to employ the "Terabyte" collection. Although not actually a terabyte in terms of size (only 436 GB), efficiently and accurately processing this collection should demonstrate the potential of our approach.

References

[Baird2007] Baird, L. C., Bahn, W. L., and Collins, M. D., “Jam-resistant communication without shared secrets through the use of concurrent codes,” U.S. Air Force Academy Technical Report USAFA-TR-2007-01, 14 Feb 2007.

[Baeza-Yates2004] Baeza-Yates, R., “A fast set intersection algorithm for sorted sequences,” CPM 2004, Istanbul, Turkey, Springer Lecture Notes in Computer Science

[Baeza-Yates1999] Baeza-Yates, R. and Ribeiro-Neto, B. *Modern Information Retrieval*, England: ACM Press/Addison-Wesley

[Barbay2006] Barbay, J., Lopez-Ortiz, A., and Lu, T., “Faster adaptive set intersections for text searching,” WEA 2006

[Demaine2003] Demaine, E.D. and Lopez-Ortiz, A. “A Lower Bound on Index Size for Text Retrieval,” Journal of Algorithms, 48(1) pp. 2-15

[Demaine2001] Demaine, E.D., Lopez-Ortiz, A., and Munro, J.I., “Experiments on Adaptive Set Intersections for Text Retrieval Systems, *Proceedings of the 3rd Workshop on Algorithm Engineering and Experiments (ALENEX)*, Lecture Notes in Computer Science, Washington, DC, 2001.

[Demaine2000] Demaine, E.D. and Lopez-Ortiz, A. “Adaptive Set Intersections, Unions, and Differences,” *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*

[Kraft1992] Kraft, D.H. and Buell, D.A., "Fuzzy Sets and Generalized Boolean Retrieval Systems," In Dubois, D., Prade, H, and Yager, R. (eds.), *Readings in Fuzzy Sets for Intelligent Systems*, San Mateo, CA: Morgan Kaufmann Publishers

[Kraft1999] Kraft, D.H., Bordogna, G., and Pasi, G., "Fuzzy Set Techniques in Information Retrieval," in Bezdek, J.C., Didier, D. and Prade, H. (eds.), Fuzzy Sets in Approximate Reasoning and Information Systems, vol. 3, The Handbook of Fuzzy Sets Series, Norwell, MA: Kluwer Academic Publishers, 1999

[Meadow2000] Meadow, C.T., Boyce, B.R., and Kraft, D.H., *Text Information Retrieval Systems*, second edition, San Diego, CA: Academic Press

[Grossman2004] Grossman, D. and Frieder, O., *Information Retrieval: Algorithms and Heuristics,* Second Edition, Springer Publishers

[Russell1995] Russell, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.

[Sanchez2002] Sanchez, S.N., Triantaphyllou, E., and Kraft, D.H., "A Feature mining Based Approach for the Classification of Text Documents into Disjoint Classes," *Information Processing and Management*, 38(4), pp. 583-604

[Spink2001] Spink, A., Wolfram, D., Jansen, B.J., and Saracevic, T., "Searching the Web: The Public and Their Queries," *Journal of the American Society for Information Science,* 53(2), pp. 226-234

[Stanford2006] http://www-db.stanford.edu/~kevin/queryLanguage.html, October 16

[Zobel2006] Zobel, J. Personal communication. June

Multi-objective Evolutionary Algorithms in the Automatic Learning of Boolean Queries: A Comparative Study

A.G. Lopez-Herrera[1], E. Herrera-Viedma[2], F. Herrera[2], C. Porcel[3], and S. Alonso[4]

[1] Dept. Computer Sciences, University of Jaén, Spain
aglopez@ujaen.es
[2] Dept. Computer Sciences and A.I, University of Granada, Spain
{viedma,Herrera}@decsai.ugr.es
[3] Dept. Computer Sciences and N.A., University of Córdoba, Spain
carlos.porcel@uco.es
[4] Software Engineering Dept., University of Granada, Spain
salonso@decsai.ugr.es

Abstract. The performance of Information Retrieval Systems (IRSs) is usually measured using two different criteria, precision and recall. In such a way, the problem of tuning an IRS may be considered as a multi-objective optimization problem. In this contribution, we focus on the automatic learning of Boolean queries in IRSs by means of multi-objective evolutionary techniques. We present a comparative study of four multi-objective evolutionary optimization techniques of general-purpose (NSGA-II, SPEA2 and two MOGLS) to learn Boolean queries.

Keywords: Information Retrieval Systems, Genetic Programming, Inductive Query By Example, Multi-objective Evolutionary Algorithms, Query Learning.

1 Introduction

Information Retrieval (IR) may be defined as the problem of selecting documentary information from storage in response to searches provided by a user in form of queries [2], [24]. IRSs deal with documentary databases containing textual, pictorial or vocal information and they process user queries to allow the user to access relevant information in an appropriate time interval.

The Boolean IR model [26] is frequently used to build queries in the IRSs. However, it presents some limitations: a Boolean query is defined by a set of terms joined by the logical operators AND, OR and NOT, but to build Boolean queries is not usually easy neither very intuitive [2]. This problem becomes a more serious issue if the users do not have previous experience with the model. A possible solution to overcome this problem is to build automatic aid tools to assist users to express their information needs by means of Boolean queries. Inductive Query By Example (IQBE) [5], where a query describing the information contexts of a set of key documents provided by the user is automatically derived or learned, is an useful paradigm to assist users to express queries.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 71–80, 2007.
springerlink.com

Some approaches based on the IQBE paradigm to learn queries have been proposed in the specialized literature [8], [9], [10], [18], [20], [25]. They are based on a kind of Evolutionary Algorithms (EA) [1] as it is Genetic Programming (GP) [19], where queries are represented as expression syntax trees and adapted using selection, crossing and mutation methods. They are usually guided by a fitness function that combines precision and recall [7] (the traditional performance measures in IR) in an unique objective. One of the main characteristics of this approach is that it only provides a single solution query in each run (the one that maximizes the fitness function). However, there exist a kind of EAs specially designed for multi-objective problems, MOEAs, which are able to obtain different solutions to the problem in a single run [6]. As multi-objective problems are characterized by the fact that several objectives have to be simultaneously optimized, there is not usually a single best solution solving the problem, that is, being better than the remainder with respect to every objective, but there exist a set of solutions which are superior to the remainder when all the objectives are considered. This set is called the Pareto set. These solutions are known as non-dominated solutions [4], while the remainder are known as dominated solutions. Since none of the Pareto set solutions is absolutely better than the other non-dominated solutions, all of them are equally acceptable regarding the satisfaction of all the objectives.

Recently an IQBE MOEA in the context of automatic learning of Boolean queries in IRSs has been proposed [9]. This IQBE MOEA is based on the first version of Strength Pareto Evolutionary Algorithm (SPEA) [30] and uses GP concepts to extend, toward the multi-objective context, the single-objective Boolean IQBE EA proposal of Smith and Smith [25]. However, there exist other MOEAs which usually improve the performance of SPEA [12], [13], [15], [29].

In this work, a comparative study of the performance of four of the currently most successful MOEAs applied to the automatic learning of Boolean queries will be done. The studied MOEAs are: the second version of Non-dominated Sorting Genetic Algorithm (NSGA-II) [12], the second version of Strength Pareto Evolutionary Algorithm (SPEA2) [29], and two MOGLS (MOEAs using Local Search): MOGLS-I [13] and MOGLS-J [15]. All of them are adapted to use GP concepts and optimize both objectives (precision and recall) simultaneously, extending the Smith and Smith IQBE EA [25] proposal to the multi-objective context. Experimental results show that NSGA-II, adapted with GP concepts, obtain the best performance in the context of automatic learning of Boolean queries.

To do this, this paper is structured as follows. In Section 2 IRSs' foundations and the IQBE paradigm are drawn. Section 3 describes the four MOEAs with GP concepts which are used. In Section 4 the experimentation framework is described, and finally Section 5 presents some conclusions.

2 Preliminaries

In this section, we introduce the foundations of the IRSs, their components and procedure of evaluation and a brief description of the IQBE paradigm.

2.1 Information Retrieval Systems

An IRS is basically composed by three main components [3]:

The documentary database: This component stores the documents and the representation of their contents. Textual documents representation is typically based on index terms (that can be either single terms or sequences), which work as content identifiers for the documents. We assume that the database is built like in usual IRSs [2], [24]. Therefore, IRS-user interaction is unnecessary because it is built automatically. The database stores a finite set of documents $D=\{d_1,\cdots,d_m\}$, a finite set of index terms $T=\{t_1,\cdots,t_l\}$, and the representation R_{d_j} of each document d_j characterized by a numeric indexing function $F: D\times T \rightarrow [0,1]$ such that $R_{d_j}=\sum_{i=1}^{l} F(d_j,t_i)/t_i$ is the representation of d_j in fuzzy sets notation. Using numeric values in (0,1), F can weight index terms according to their significance in describing the content of a document in order to improve the retrieval of documents.

The query subsystem: It allows users to formulate their information needs (queries) and presents the relevant documents which are retrieved by the system. To do this, each query is expressed as a combination of index terms which are connected by the Boolean operators AND ($\wedge$), OR ($\vee$), and NOT ($\neg$).

The matching mechanism: It evaluates the degrees (the Retrieval Status Value (RSV)) to which the document representation satisfy the requirements expressed in the query, and it retrieves the documents that are judged to be relevant. To evaluate Boolean queries, the matching function uses a constructive bottom-up process based on the separability criterion [27]. This process includes two steps:

- Firstly, the documents are evaluated according to their relevance only to the terms of the query. In this step, a partial relevance degree is assigned to each document with respect to every term in the query.
- Secondly, the documents are evaluated according to their relevance to the Boolean combination of the terms (their partial relevance degree), and so on, working in a bottom-up fashion until the whole query is processed. In this step, a total relevance degree is assigned to each document that is used to rank the documents from the most relevant one to the less relevant.

2.2 Evaluation of Information Retrieval Systems

There are several ways to measure the quality of an IRS, such as the system efficiency and effectiveness, and several subjective aspects related to user satisfaction [2]. Traditionally, the retrieval effectiveness is based on the document relevance with respect to the users needs. There are different criteria to measure this aspect, but precision and recall are the most used. Precision is the ratio between the relevant documents retrieved by the IRS in response to a query and the total number of documents retrieved, whilst recall is the ratio between the number of relevant documents retrieved and the total number of relevant documents for the query that exist in the database [26]. The mathematical expression of each of them is:

$$P=\frac{D_{rr}}{D_{tr}},\ R=\frac{D_{rr}}{D_{rt}} \tag{1}$$

where D_{rr} is the number of relevant documents retrieved, D_{tr} is the total number of documents retrieved and D_{rt} is the total number of relevant documents for the query which exist in the database. P and R are defined in [0, 1], being 1 the optimal value.

We notice that the only way to know all the relevant documents existing for a query in the database (value used in the R measure) is to evaluate all documents. Due to this fact and tacking into account that relevance is subjective, there are some classic documentary databases (TREC, CACM, Cranfield) available, each one with a set of queries for which the relevance judgments are known, so that they can be used to verify the new proposals in the field of the IR [2], [22]. In this contribution, we use the Cranfield collection.

2.3 The IQBE Paradigm

The IQBE paradigm was proposed by Chen [5] as a process in which users provide documents (examples) and an algorithm induces (or it learns) the key concepts of the examples with the purpose of finding other and equally relevant documents. In this way, IQBE can be seen as a technique to assist users in the query building process by using automatic learning methods.

It works taking a set of relevant documents (and optionally non-relevant documents) provided by the user (they can be obtained from a preliminary query or from a browsing process through the documentary database) and applying an automatic learning process to generate a query that describes the user information needs (represented by the previous set of documents). The query that is obtained can be executed in other IRSs to obtain new relevant documents. In this way, it is not necessary for the user to interact with the IR process which is mandatory in other techniques for query refinement as the relevance feedback [22].

Several IQBE techniques for different IR models have been proposed [7]. The most used IQBE models are based on GP concepts, with queries being represented by expression syntax trees and the algorithms are articulated on the basis of the classic operators: cross, mutation and selection.

3 Structure of the MOEAs with GP Concepts

3.1 Components

The four studied MOEAs with GP (MOEAs-GP) concepts share the following components:

- *Codification scheme*: Boolean queries are encoded in expression syntax trees, whose terminal nodes are query terms and whose inner nodes are the Boolean operators AND, OR and NOT.
- *Crossover operator*: subtrees are randomly selected and crossover in two randomly selected queries.
- *Mutation operator*: changes a randomly selected term or operator in a randomly selected tree.
- *Initial population*: all individuals of the first generation are generated in a random way. The population is created including all the terms in the relevant documents

provided by the user. Those that appear in more relevant documents will have greater probability of being selected.

- *Objectives to optimize*: precision and recall.
- *Local search*: it is based on the Crossover Hill-Climbing (XHC) algorithm defined in [21] and adapted to expression syntax trees. This XHC operator uses hill-climbing as the move accepting criterion of the search and uses crossover as the move operator. XHC maintains a pair of predecessors and repeatedly performs crossover on this pair until some number of offspring, n_{off} , is reached. Then, the best offspring is selected and it replaces the current solution only if it is better. The process iterates n_{it} times and returns the final current solution. This XHC requieres values for n_{off} and n_{it}, and a starting pair of parents.

3.2 NSGA-II-GP

NSGA-II [12] is a very complete algorithm since, not only incorporates a strategy of preservation of an elite population, but in addition, it uses an explicit mechanism to preserve diversity.

NSGA-II works with a population of offsprings Q_t, which is created using a predecesor population P_t. Both populations (Q_t and P_t) are combined to form a unique population R_t, with a size $2 \cdot M$, that is examined in order to extract the front of the Pareto. Then, an arrangement on the non-dominated individuals is done to classify the R_t population. Although this implies a greater effort compared with the arrangement of the set Q_t, it allows a global verification of the non-dominated solutions that as much belong to the population of offsprings a the one of the predecesors.

Once the arrangement of the non-dominated individuals finishes, the new generation (population) is formed with solutions of the different non-dominated fronts, taking then alternatively from each of the fronts. It begins with the best front of non-dominated individuals and continues with the solutions of the second one, later with third one, etc.

Since the R_t size is $2 \cdot M$, it is possible that some of the front solutions have to be eliminated to form the new population.

In the last states of the execution, it is usual that the majority of the solutions are in the best front of not-domintad solutions. It is also probable the size of the best front of the combined population R_t be bigger than M. It is then, when the previous algorithm assures the selection a diverse set of solutions of this front by means of the method of niches. When the whole population converges to the Pareto-optimal frontier, the algorithm continues, so that the best distribution between the solutions is assured.

3.3 SPEA2-GP

SPEA2 [29] introduces elitism by explicitly maintaining an external population. This population stores a fixed number of the non-dominated solutions found from the beginning of the simulation.

In each generation, the new non-dominated solutions are compared with the existing external population and the resulting non-dominated solutions are preserved. In addition, SPEA2 uses these elite solutions in the genetic operations with the current population to guide the population towards good regions in the search space.

The algorithm begins with a randomly created population P_0 of size M and an external population $\underline{P}_0$ (initially empty) which has a maximum capacity $\underline{M}$. In each generation t, the best non-dominated solutions (belonging to the best non-dominated front) of the populations P_t and $\underline{P}_t$ are copied in the external population $\underline{P}_{t+1}$. If the size of $\underline{P}_{t+1}$ exceeds M, then $\underline{P}_{t+1}$ is reduced by means of a truncate operator; on the other hand, $\underline{P}_{t+1}$ is filled up with dominated solutions from P_t and $\underline{P}_t$. This truncate operator is used to maintain the diversity of the solutions.

From $\underline{P}_{t+1}$, a pool of individuals is obtained applying a binary tournament selection operator with replacement. These individuals are crossed and mutated to obtain the new generation P_{t+1}.

3.4 MOGLS-GP-I

MOGLS is an hybrid approach that combines concepts of MOEAs and Local Search for improving the current population. Several alternatives have been proponed in the specialized literature [13], [14], [15], [16]. In this paper we have chosen to study the performance of two of them, MOGLS-GP-I [13] and MOGLS-GP-J [15]. MOGLS-GP-I, is an adapted version of the first Ishibuchi proposal [13] which includes GP concepts. This approach has great part of its effort in obtaining the most extended Pareto front possible. To do so, it associates each individual of the population with weighting vector that points the direction, in the objective space, with which that individual was generated. These directions are later considered when generating new individuals.

This algorithm maintains elitism using two sets of solutions: the current population and a provisional population of non-dominated solutions. The algorithm begins generating an initial population of N_{pop} individuals, it evaluates them and updates the provisional population with the non-dominated individuals. Next, N_{pop} - N_{elite} predecessor solutions are obtained considering their direction vectors. After the crossing and mutation processes, the population will be completed with N_{elite} non-dominated individuals of the provisional population. Next, a local search (XHC) is applied to all N_{pop} individuals of the current population. In this stage, the direction vector of each individual will guide the local search process. Finally, the following generation is obtained with the N_{pop} improved individual of the actual population.

3.5 MOGLS-GP-J

In this subsection we describe an adapted version of the first Jaszkiewicz proposal [15] which use GP concepts. We call it MOGLS-GP-J. This MOGLS implements the idea of simultaneous optimization of all weighted Tchebycheff or all weighted linear utility functions by random choice of the utility function optimized in each iteration. The general idea of this MOGLS is similar to that used by Ishibuchi [13]. The main difference is in the the way that the solutions are selected for recombination. A Current Set of solutions (*CS*) is used. *CS* is initially filled up with *S* random solutions. In each iteration, the algorithm randomly draws an utility function *u*. From *CS*, *k* different solutions (with the best *u* evaluations) are selected to form a Temporary Population (*TP*). A crossover operador is applied to two randomly selected parents from the *TP*, and the new resulting solution *x* is locally optimized using the XHC local search operator. A Potentially Efficient (*PE*) set, with non-dominated solutions is updated with the new solution *x*. The random selection of utility functions may be seen as a mechanism that introduces some additional diversification.

3.6 Evaluation of MOEAs

In multi-objective optimization problems, the definition of the quality concept is substantially more complex than in single-objective ones, since the processes of optimization imply several different objectives. In the specialized literature several quantitative measures have been proposed [6], [11], [17], [28]. The most used is the *C* measure, whose expression:

$$C(A,B) = \frac{\left|\{a \in A; \exists b \in B : b \succ a\}\right|}{\left|A\right|} \tag{2}$$

measures the ratio of individuals of the Pareto *A* that are dominated by individuals of the Pareto *B*. A value of 1 indicates that all individuals of the Pareto *A* are dominated by individuals of the Pareto *B*; on the other hand, a value of 0 indicates that none of the individuals of *A* is dominated by individuals of *B*.

4 Experimental Study

The experimental study has been developed using the Cranfield collection, componed by 1398 documents about Aeronautics. The 1398 documents have been automatically indexed in the usual way, removing the stop-words, and obtaining 3857 different index terms in total. A *tf·idf* scheme[1] [22] has been used to represent the relevant index terms in the documents. Cranfield provides 225 queries, of which, those that have 20 o more relevant documents have been selected. The seven resulting queries (#1, #2, #23, #73, #157, #220 and #225) have 29, 25, 33, 21, 40, 20, 25 relevant documents associated respectively. Our MOEA-GP approaches based on the IQBE paradigm generate a set of queries from a relevant and a non-relevant documents sets. To do so, it is necessary to consider sufficiently representative number of positive examples (relevant documents), so queries with more relevant documents associated have been selected.

The studied MOEAs in this contribution have been run 30 times for each query (a total of 840 runs) with different initializations for each selected query during the same fixed number of fitness function evaluations (50.000) in a 1.5GHz Pentium Mobile computer with 2Gb of RAM. The common parameter values considered are a maximum of 19 nodes for trees, 0.8 arid 0.2 for crossover and mutation probabilities, respectively and population size of M = 800 queries. Additionally, MOGLS-GP-I and MOGLS-GP-J use an elite population size = 200, local search probability = 0.3, n_{off} = 3 and n_{it} = 10; MOGLS-GP-J use an initial population S = 1600; and SPEA2-GP use an elite population size = 200.

From each run a pareto set is obtained. The four pareto sets obtained by each run and query are compared with the performance *C* measure (in Table 1, average results

[1] To do so, we use the classical Salton's SMART IR [23].

of the *C* measure for each pair of MOEAs and query are showed). The experimental results show that NSGA-II, with GP concepts (NSGA-II-GP), is the IQBE MOEA technique that achieves a better performance achieves, i.e., it achieves better non-dominated solutions sets (view values in bold type-style in Table 1) in the process of learning Boolean queries in IRSs, than the other studied IQBE MOEAs.

Table 1. Average results of the *C* measure for each queries and pair of MOEAs-GP studied. W=NSGA-II, X=SPEA2, Y=MOGLS-GP-I and Z=MOGLS-GP-J.

Query	**C(W,X) / C(X,W)**	**C(W,Y) / C(Y,W)**	**C(W,Z) / C(Z,W)**
#1	**0.908** / 0.031	**0.803** / 0.129	**0.861** / 0.071
#2	**0.973** / 0.013	**0.902** / 0.037	**0.936** / 0.025
#23	**0.917** / 0.022	**0.895** / 0.030	**0.922** / 0.017
#73	**0.952** / 0.026	**0.943** / 0.044	**0.955** / 0.036
#157	**0.934** / 0.002	**0.925** / 0.002	**0.933** / 0.005
#220	**0.924** / 0.019	**0.917** / 0.036	**0.920** / 0.027
#225	**0.940** / 0.007	**0.909** / 0.007	**0.927** / 0.007
Query	**C(X,Y) / C(Y,X)**	**C(X,Z) / C(Z,X)**	**C(Y,Z) / C(Z,Y)**
#1	0.067 / 0.888	0.157 / 0.700	0.704 / 0.240
#2	0.068 / 0.867	0.176 / 0.681	0.675 / 0.204
#23	0.053 / 0.852	0.207 / 0.639	0.713 / 0.179
#73	0.033 / 0.869	0.156 / 0.715	0.641 / 0.207
#157	0.064 / 0.846	0.158 / 0.739	0.639 / 0.234
#220	0.054 / 0.877	0.196 / 0.693	0.698 / 0.229
#225	0.037 / 0.865	0.150 / 0.679	0.719 / 0.164

5 Conclusions

In this contribution he have presented a comparative study of performance, in the Boolean IR models context, of four of the most currently successful MOEAs in the specialized literature has been done. The studied MOEAs have been applied on the automatic learning of Boolean queries problem. The original proposals [12], [13], [15], [29] have been adapted to use GP concepts. All of them extend the Smith and Smith's IQBE EA propose [25] to work in a multi-objective context. The experimental results show that NSGA-II, with GP concepts (NSGA-II-GP), is the best IQBE MOEA technique, i.e., it achieves better non-dominated solutions sets in the process of learning Boolean queries in IRSs, than the other studied IQBE MOEAs.

In future works, we will perform a more exhaustive comparative study, using additional IQBE MOEAs and bigger database collections like TRECs.

Acknowledgments. This work has been supported by the projects: GLIRS-II, Ref. TIC-2003-07977 and Proyecto de Excelencia de la Junta de Andalucia SAINFOWEB, Cod. 00602.

References

1. T. Back, D. Fogel, and Z. Michalewicz, Handbook of evolutionary computation, IOP Publishing and Oxford University Press, 1997.
2. R. Baeza-Yates and B. Ribeiro-Neto, Modern information retrieval, Addison-Wesley, 1999.

3. G. Bordogna, P. Carrara, and G. Pasi, Fuzzy sets and possibility theory in database management systems, ch. Fuzzy Approaches to Extend Boolean Information Retrieval, pp. 231-274, Bosc, P. and Kacpryk, J., Eds. (Springer Verlag), 1995.
4. V. Chankong and Y.Y. Haimes, Multiobjective decision making theory and methodology, North-Holland, 1983.
5. H. Chen, G. Shankaranarayanan, L. She, and A. Iyer, A machine learning approach to inductive query by example: An experiment using relevance feedback, ID3, genetic algorithms, and simulated annealing, Journal of the American Society for Information Science 49 (1998), no. 8, 693-705.
6. C.A. Coello, D.A. Van Veldhuizen, and G.B. Lamant, Evolutionary algorithms for solving multi-objective problems, Kluwer Academic Publishers, 2002.
7. O. Cordón, E. Herrera-Viedma, C. López-Pujalte, C. Zarco and M. Luque, A review on the application of evolutionary computation to information retrieval, International Journal of Approximate Reasoning 34 (2003), 241-264.
8. O. Cordon, E. Herrera-Viedma, and M. Luque, Evolutionary learning of boolean queries by multiobjective genetic programming, Proceding of PPSN-VII, Granada, Spain, LNCS 2439, Springer, 2002, pp. 710-719.
9. O. Cordón, E. Herrera-Viedma, and M. Luque, Improving the learning of boolean queries by means a multiobjective IQBE evolutionary algorithm, Information Processing & Management 42 (2006), no. 3, 615-632.
10. O. Cordón, F. Moya, and C. Zarco, A GA-P algorithm to automatically formulate extended boolean queries for a fuzzy information retrieval system, Mathware & Soft Computing 7 (2000), no. 2-3, 309-322.
11. K. Deb, Multi-objective optimization using evolutionary algorithms, Wiley, 2001.
12. K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2002), 182-197.
13. H Ishibuchi and T. Murata, A multi-objective genetic local search algorithm and its application to flowshop scheduling, IEEE Transactions on Systems, Man and Cibernetics, Part C: Applications and Reviews 28 (1998), no. 3, 392-403.
14. H. Ishibuchi, T. Yoshida, and T. Murata, Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling, IEEE Transactions on Evolutionary Computation 7 (2003), no. 2, 204-223.
15. A. Jaszkiewicz, Genetic local search for mult-objective combinational optimization, European Journal of Operational Research 137 (2002), 50-71.
16. A. Jaszkiewicz, Do multiple-objective metaheuristics deliver on their promises? A computacional experiment on the set-covering problem, IEEE Transactions on Evolutionary Computation 7 (2003), no. 2, 133-143.
17. J. Knowles, L. Thiele, and E. Zitzler, A tutorial on the performance assessment of stochastic multiobjective optimizers, 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, feb 2006, revised version.
18. R. Korfhage and J. Yang, Query modifications using genetic algorithms in vector space models, International Journal of Expert Systems 7 (1994), no. 2, 165-191.
19. J. Koza, Genetic programming. on the programming of computers by means of natural selection, The MIT Press, 1992.
20. D.H. Kraft, F.E. Petry, B.P. Buckles, and T. Sadasivan, Genetic algorithms and fuzzy logic systems, ch. Genetic algorithms for query optimization in information retrieval: relevance feedback, pp. 155-173, Sanchez, E. and Shibata, T. and Zadeh, L.A., Eds., (World Scientific), 1997.

21. M. Lozano, F. Herrera, N. Krasnogor, and D. Molina, Real-coded memetic algorithms with crossover hill-climbing, Evolutionary Computation 12 (2004), no. 3, 272-302.
22. G. Salton, Automatic text processing: The transformation, analysis and retrieval of information by computer, Addison-Wesley, 1989.
23. G. Salton, The smart retrieval system. experiments in automatic document processing., Englewood Cliffs: Prentince-Hall, 1997.
24. G. Salton and M.J. McGill, An introduction to modern information retrieval, McGraw-Hill, 1983.
25. M. P. Smith and M. Smith, The use of genetic programming to build boolean queries for text retrieval through relevance feedback, Journal of Information Science 23 (1997), no. 6, 423-431.
26. C.J. van Rijsbergen, Information retrieval, Butterworth, 1979.
27. W.G. Waller and D.H. Kraft, A mathematical model of a weighted boolean retrieval system, Information Processing and Management 15 (1979), 235-245.
28. E. Zitzler, K. Deb, and L. Thiele, Comparison of multiobjective evolutionary algorithms: Empirical results, Evolutionary Computation 8 (2000), no. 2, 173-271.
29. E. Zitzler, M. Laumanns, and L. Thiele, SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization, Evolutionary Methods for Design, Optimization and Control with Application to Industrial Problems (EUROGEN 2001) (K. Giannakoglou et al. (Eds.), ed.), International Center for Numerical Methods in Engineering (CIMNE), 2002, pp. 95-100.
30. E. Zitzler and L. Thiele, Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach, IEEE Transactions on Evolutionary Computation 3 (1999), no. 4, 257-271.

Interactions Between Decision Goals Applied to the Calculation of Context Dependent Re-rankings of Results of Internet Search Engines

Rudolf Felix

F/L/S Fuzzy Logik Systeme GmbH, Joseph-von-Fraunhofer-Str. 20, 44227
Dortmund, Germany
felix@fuzzy.de

Abstract. Search keywords can be considered as decision criteria (goals), whereas the documents contained in the original result list of a search engine can be considered as decision alternatives in the sense of adequate search results. The paper shows how a decision making model based on interactions between decision goals can be used for re-ranking of the search results obtained by classical search engines. The decision making model was previously applied to real world problems in production optimization and business process management. In contrast to other approaches, the interactive structure of decision goals for each decision situation is calculated explicitly based on fuzzy types of interaction. In this paper, after a brief description of the model an application to the calculation of context depending re-rankings of search results of internet search engines is presented. It is shown that keywords of search queries can be understood as decision goals and that the decision making model can be successfully applied for context dependent re-ranking of the search results. Using the model, popularity based rankings can be re-ranked making use of context dependent information derived from the query keywords.

Keywords: Decision making, interaction of decision goals, context depending rankings, internet search, re-ranking of search results.

1 Introduction

Humans efficiently decide whether or not a listed document is interesting from the point of view of the search query they previously entered to the search engine. They rather perform a cross check of the documents instead of reading them in detail and rank them with respect to the initial intention of the search query expressed by the keywords of the query. The process of re-ranking the search results can be considered as a decision making process with the keywords used in the query being the decision goals. The decision alternatives are the selected documents themselves which are to be re-ranked with respect to the keywords. It is shown how the concept of interactions between decision goals (here: keywords) can be applied for ranking the documents (here: decision alternatives). In the subsequent sections first basic definitions used to formalize the concept of interactions between decision goals are presented. It is also discussed in which sense interactions between the goals are used to model the decision making. Then it is shown how the model is applied to the

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 81–88, 2007.
springerlink.com

problem of context dependent ranking of documents being search results of a classical internet search engine. Subsequently the results obtained from a system prototype are discussed and a summarizing conclusions are made.

2 Basic Definitions

Before we define interactions between goals as fuzzy relations, we introduce the notion of the positive impact set and the negative impact set of a goal. A more detailed discussion can be found in [6] and [7].

Def. 1)
Let A be a non-empty and finite set of potential alternatives, G a non-empty and finite set of goals, $A \cap G = \emptyset, a \in A, g \in G, \delta \in (0,1]$. For each goal g we define the two fuzzy sets S_g and D_g each from A into [0, 1] by:

1. Positive impact function of the goal g: $S_g(a) := \delta$, if a affects g positively with degree δ then $S_g(a) = \delta$, $S_g(a) := 0$ *else.*
2. Negative impact function of the goal g: $D_g(a) := \delta$, if a affects g negatively with degree δ then $D_g(a) = \delta$, $D_g(a) := 0$ *else.*

Def. 2)
Let S_g and D_g be defined as in Def. 1). S_g is called the *positive impact set of* g and D_g the *negative impact set of* g.
The set S_g contains alternatives with a positive impact on the goal g and δ is the degree of the positive impact. The set D_g contains alternatives with a negative impact on the goal g and δ is the degree of the negative impact.

Def. 3)
Let A be a finite non-empty set of alternatives. Let $\mathcal{P}(A)$ be the set of all fuzzy subsets of A. Let $X, Y \in \mathcal{P}(A)$, x and y the membership functions of X and Y respectively.

The *fuzzy inclusion* $\mathrm{I}: \mathcal{P}(A) \times \mathcal{P}(A) \to [0,1]$ is defined as follows:
$\mathrm{I}(X,Y) := \Sigma_{a \in A} (\min(x(a), y(a))) \,/\, \Sigma_{a \in A}$, for $X \neq \emptyset$ and $\mathrm{I}(X,Y) := 1$ for $X = \emptyset$, with $x(a) \in X$ and $y(a) \in Y$.
The *fuzzy non-inclusion* $\mathrm{N}: \mathcal{P}(A) \times \mathcal{P}(A) \to [0,1]$ is defined as:

$$\mathrm{N}(X,Y) := 1 - \mathrm{I}(X,Y)$$

The inclusions and non-inclusions indicate the existence of interaction between two goals. The higher the degree of inclusion between the positive impact sets of two goals, the more cooperative the interaction between them. The higher the degree of inclusion between the positive impact set of one goal and the negative impact set of the second, the more competitive the interaction. The non-inclusions are evaluated in a similar way. The higher the degree of non-inclusion between the positive impact sets of two goals, the less cooperative the interaction between them. The higher the degree of non-inclusion between the positive impact set of one goal and the negative impact set of the second, the less competitive the relationship.

The pair (S_g, D_g) represents the whole known impact of alternatives on the goal g. Then S_g is the fuzzy set of alternatives which satisfy the goal g. D_g is the fuzzy set of alternatives which are rather not recommendable from the point of view of satisfying the goal *g*.

3 Interactions Between Goals

Based on the inclusion and non-inclusion defined above, 8 basic fuzzy types of interaction between goals are defined. The different types of interaction describe the spectrum from a high confluence between goals (analogy) to a strict competition (trade-off).

Def. 4)
Let S_{g_1}, D_{g_1}, S_{g_2} and D_{g_2} be fuzzy sets given by the corresponding membership functions as defined in Def. 2). For simplicity we write S_1 instead of S_{g_1} etc.. Let $g_1, g_2 \in G$ where G is a set of goals.

The fuzzy types of interaction between two goals are defined as relations which are fuzzy subsets of $G \times G$ as follows:

1. g_1 is independent of g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{N}(S_1, D_2), \mathrm{N}(S_2, D_1), \mathrm{N}(D_1, D_2))$
2. g_1 assists g_2: <=> $\min(\mathrm{I}(S_1 S_2), \mathrm{N}(S_1, D_2))$
3. g_1 cooperates with g_2: <=> $\min(\mathrm{I}(S_1, S_2), \mathrm{N}(S_1, D_2), \mathrm{N}(S_2, D_1))$
4. g_1 is analogous to g_2: <=> $\min(\mathrm{I}(S_1, S_2), \mathrm{N}(S_1, D_2), \mathrm{N}(S_2, D_1), \mathrm{I}(D_1, D_2))$
5. g_1 hinders g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{I}(S_1, D_2))$
6. g_1 competes with g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{I}(S_1, D_2), \mathrm{I}(S_2, D_1))$
7. g_1 is in trade-off to g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{I}(S_1, D_2), \mathrm{I}(S_2, D_1), \mathrm{N}(D_1, D_2))$
8. g_1 is unspecified dependent from g_2: <=>
   $\min(\mathrm{I}(S_1, S_2), \mathrm{I}(S_1, D_2), \mathrm{I}(S_2, D_1), \mathrm{I}(D_1, D_2))$

The interactions between goals are crucial for an adequate orientation during the decision making process because they reflect the way the goals depend on each other and describe the pros and cons of the decision alternatives with respect to the goals. For example, for cooperative goals a conjunctive aggregation is appropriate. If the goals are rather competitive, then an aggregation based on an exclusive disjunction is appropriate. The interactions between the decision goals have been applied in a problem independent decision making model that has already been applied in many relevant real world solutions [9], [11]. For a more detailed discussion see for instance [7].

4 Approaches Related to the Decision Making Model Based on Interactions Between Goals

Since fuzzy set theory has been suggested as a suitable conceptual framework of decision making [2] several different approaches have been developed [3], [4], [5], [13], [14], [15], [17], [19],[20], [21]. The decision making approach based on interaction between goals significantly differs from these approaches. For a more detailed discussion see for instance [6], [8], [10], [12]. It also significantly differs from fuzzy rule based approaches applied for instance to the selection of web service compositions [1] as these approaches do not explicitly use negative selection information.

5 Application to Context Dependent Ranking of Internet Search Results

In the subsequent sections the application of the decision making model to the context dependent ranking of documents that are search results of a classical internet search engine is described. First the ranking problem is defined. Then it is indicated how the positive and negative impacts sets are used in order to translate the context of a search query to the decision making model. Subsequently examples of search queries are shown and the results of the rankings of the documents are discussed. The discussion of the resulting rankings is made by comparing the results with the rankings obtained by a well-known internet search engine.

5.1 Description of the Ranking Problem

Classical internet search engines generate rankings which are not based on content dependent information. The search keywords which represent the context of the query are used as index information. The rankings of the documents are then rather popularity based and concentrate on information like the number of input and output links. On the other hand a particular user evaluates the search result from his own point of view, that means from the point of view of the context of his own query. In consequence, the ranking generated by a classical search engine may not fit the intention of the user's search query. In such cases a document may be part of the set of documents found but may be ranked with a lower value and placed towards the end of the ranking generated by the search engine. The aim of the approach presented below is to re-rank the result of a classical search engine in a way that better fits the context of the query.

5.2 Positive and Negative Context Information

The search keywords can be considered as decision criteria (as goals in the sense of Def. 1) whereas the documents contained in the original result list of the search engine can be considered as decision alternatives. Using the decision making model presented in the previous sections the ranking problem can be solved in the following way.

For each search query the keywords used are considered as the set G of decision goals. The result list to this query generated by the conventional search engine is considered as the set of decision alternatives A. For every goal (keyword) the positive impact function's value for a particular document a $\in A$ is defined by counting the number of occurrences of the keyword (goal) normalized by the total number of words contained in the document (alternative). The more occurrences of the search keyword in the document the higher the value of the positive impact function of the keyword (goal) for the document (alternative) and the lower the value of the negative impact function of the keyword (goal) for the document (alternative). In an analog way we model: The less occurrences of the positive keyword (goal) in the document the lower the value of the positive impact function and the higher the value of the negative impact function.

The user also has the possibility to define the so-called negative keywords which are key-words that should not appear in the document. Using the concept of impact functions for each negative keyword again the documents are evaluated in an analog way as the positive keywords but with opposite increasing and decreasing values.

In the sense of the search query defined by the user the positive keywords describe what the user is tentatively looking for, the so-called positive context of the search query. The negative keywords describe what the user tentatively is not looking for, the so-called negative context of the search query.

Every search query is evaluated in the way that every document which is contained in the ranking generated by the conventional search engine is considered as a decision alternative $a \in A$. For every keyword the values of both the positive and the negative impact functions are calculated as described above. If for every keyword both the positive and the negative impact functions are calculated and the impact sets are passed to the decision model based on interactions between decision goals (here search keywords). The decision model calculates a new ranking of the documents which better correspond then to the context of the search query defined by the user.

5.3 Examples of Search Queries and Resulting Rankings

Let us explain some experimental results of the search concept obtained by using some common sense examples. The examples presented subsequently are based on the internet search results of Google and Google's index information. As already explained each search query consists of two sets of keywords. The first one is the positive keyword set (PKWS). The PKWS consists of keywords which describe what the user is searching for, that means the positive context of the search query. The second set of keywords is the negative keyword set (NKWS). The NKWS consists of keywords that describe the context the user is rather not interested in, that means the negative context of the search query. A search query is always a pair (PKWS:={positive keywords},NKWS:={negative keywords}). The search is organized as follows: First the search engine Google is used to generate a list of documents based on a search query with the PKWS as keyword list and the result is the initial search result list (ISRL). The order of the links in the ISRL is the initial ranking (IR) of the links or documents. The ISRL is used as input for the re-ranking procedure based on the interactions between the keywords (goals) as presented in the previous sections. The result of the re-ranking is the final ranking (FIR). For simplicity, in the subsequent examples the ISRL consisted of at most 50 links. Therefore both IR and FIR have at maximum 50 ranking positions.

Let us consider a first search query Q1=(PKWS:={Klose},NKWS=:{}). The positive keyword set consists of the keyword “Klose”, the negative keyword set is empty. The IR of the query consists of 50 positions of which 24 links refer to documents about the German soccer player Miroslav Klose. If we modify the search query by adding the negative keyword “Fussball” which is the German word for “soccer” we obtain the search query Q2=(PKWS:={Klose}, NKWS:={Fussball}). The first 10 ranking positions of FIR of Q2 compared to the first 10 ranking positions FIR of Q1 and the links are given in Fig. 1. It can be seen that links with relation to the soccer player Miroslav Klose dominate the FIR of Q1 and that the dominance in the result of Q2 is reduced to 0 because in the FIR of Q2 the positions 1 to 10 do not contain any link related to the soccer player Miroslav Klose.

FIR of Q1			FIR of Q2		
Links	Resulting Ranking	Google Ranking	Links	Resulting Ranking	Google Ranking
Miroslave Klose – die offfizielle Webseite *http://www.miroslavklose.de*	1	1	Friedrich Klose – Wikipedia http://de.wikipedia.org/wiki/F...	1	3
Miroslav Klose – Wikipedia http://de.wikipedia.org/wiki/M...	2	2	Die Klose Kollektion GmbH http://www.klosekollektion.de/	2	4
Friedrich Klose – Wikipedia http://de.wikipedia.org/wiki/F...	3	3	Staudengaertner-Klose, Paeonien, Staudenversand, Garten *http://www.staudengaertner-klo...*	3	6
Die Klose Kollektion GmbH http://www.klosekollektion.de/	4	4	Galerie Klose http://www.galerie-klose.de/	4	11
Staudengaertner-Klose, Paeonien, Staudenversand, Garten *http://www.staudengaertner-klo...*	5	6	BBQ Pits by Klose – Houston, TX *http://www.bbqpits.com/*	5	12
2006 FIFA World Cup Germany – Player Profile Page – KLOSE Miroslav – Germany *http://fifaworldcup.yahoo.com/...*	6	7	Rainer Klose Werkzeuge für's Handwerk, Münster *http://www.klose-ms.de/*	6	13
2006 FIFA World Cup Germany – Player Profile Page – KLOSE Miroslav – Germany *http://fifaworldcup.yahoo.com/...*	7	8	Fehlermeldung http://www.bundestag.de/mdb15/...	7	14
ZEIT online - Fussball --- WM 2006: Es traf Klose *http://www.zeit.de/2006/24/Fus...*	8	9	Klose, Hans-Ulrich http://www.bundestag.de/mdb/bi...	8	15
ZEIT online – Fussball --- Länderspiel: Überragender Klose *http://www.zeit.de/online/2006...*	9	10	BW – Hilchenbach.de http://www.klose-antriebstechn...	9	18
Galerie Klose http://www.galerie-klose.de/	10	11	Familie Klose aus Heisingen *http://www.klose-family.de/*	10	19

Fig. 1. Comparison of the results of the search queries Q1 and Q2

Let us now consider an another search query Q3=(PKWS:={Ronaldo}, NKWS=:{}). Again for simplicity the IR was reduced to at most 50 links. Without going into details it was observed that the ranking of Q3 contained in total 34 links which refer to the Brazilian soccer player Ronaldo and 7 links referring to the Portuguese soccer player Christiano Ronaldo. Obviously there is an ambiguity because of the nickname of the Brazilian player and the last name of the Portuguese player. However, among the first 20 links in the ranking there were 17 links to Ronaldo and only 2 links to Christiano Ronaldo. This overrepresentation of the Brazilian Ronaldo was due to the fact that he is more famous and therefore there were more links to internet sites referring to him. Let us compare the results of Q3 with the results obtained for Q4=(PKWS:={Ronaldo}, NKWS=:{Brasil}). We used the radical “Brasil” of the German word Brasilien for Brazil and defined it as element of the NKWS of the search query Q4. With this query we tried to lower the ranking positions of the Brazilian Ronaldo. And in fact we observed that this was the case: Among the first twenty positions of the ranking for Q4 there were 15 links to the Brazilian Ronaldo and 5 links to the Portuguese Christiano Ronaldo. Please, note that we did not obtain the same result by using the Google search with the excluding keyword "Brasil". The Google search result with "Ronaldo –Brasil" provided a ranking which still contained 17 links to the Brazilian among the first 20 positions in the ranking and 2 to the Portuguese. We considered the first 10 ranking positions of both Q3 and Q4 observed that in the ranking of Q3 we had 8 links to the Brazilian Ronaldo and only 2 to Christiano Ronaldo. Compared to this Q4 provided a ranking with 5 links to the Brazilian Ronaldo and also 5 to the Portuguese Cristiano Ronaldo.

6 Usability of Positive and Negative Keywords

When using search engines, users normally make their queries with two or three words. One might assume that it will be difficult motivating the user to provide negative keywords. However, if the search queries are not trivial, the number of keywords may increase. It can be expected that if the classical popularity ranking doesn't match the user's interest (like in the Ronaldo-Example) the documents on the top of the classical ranking will be negative examples of search results that do not fit what the user was looking for. If so, it will be easy for the user to identify negative keywords for instance by observing the titles and /or some keywords of these documents. Already after a few clicks there will be enough information available for writing down two or three negative keywords and starting a new (extended) search query again. This kind of user behavior may also be part of the user model and the user's interaction may be evaluated in the sense of an automated acquisition of both positive and negative query information [16].

Please note that using negative keywords as considered in the paper is different to the notion of negative selectivity as discussed in [18], where the selectivity is understood as the cardinality of the result of the query and is called negative if the cardinality is equal to 0.

7 Conclusions

The approach presented in this paper shows that the decision making model based on interactions between decision goals may be successfully used to re-rank search results of internet search engines with classical popularity based rankings. The process of cross-checking documents can be modeled as a decision making process within which the keywords are the decision goals and the decision alternatives are the documents to be re-ranked. Using both positive and negative keywords helps to better describe the search context. The interaction between the keywords considered as interaction between goals during the execution of the search queries is in opinion of the author a promising idea of extending popularity based search by keyword oriented context depending re-ranking of search results.

References

1. Agarwal, S. and Lamparter, S., "User Preference Based Automated Selection of Web Service Compositions", IBM Research Report, Proceedings of the International Workshop on Dynamic Web Processes (DWP 2005), Amsterdam, Netherlands, 2005
2. Bellman, R. E. and Zadeh, L. A., "Decision Making in a Fuzzy Environment", Management Sciences, 17, 141-164, 1970.
3. Biswal, M. P., "Fuzzy programming technique to solve multi-objective geometric programming problems", Fuzzy Sets and Systems, Vol.51, 67-71, 1992.
4. Dubois, D. and Prade, H., "Criteria Aggregation and Ranking of Alternatives in the Framework of Fuzzy Set Theory", Studies in Management Sciences, Volume 20, 209-240, 1984.
5. Dubois, D. and Prade, H., "Fuzzy Sets and Possibility Theory: Some Applications to Inference and Decision Processes", Fuzzy Logic - Theorie und Praxis, ed B Reusch: Springer-Verlag, 66-83, 1992.
6. Felix, R., "Entscheidungen bei qualitativen Zielen", PhD Dissertation, University of Dortmund, Department of Computer Sciences, 1991.
7. Felix, R., "Relationships between goals in multiple attribute decision making", Fuzzy Sets and Systems, Vol.67, 47-52, 1994.
8. Felix, R., "Fuzzy decision making based on relationships between goals compared with the analytic hierarchy process", Proceedings of the Sixth International Fuzzy Systems Association World Congress, Vol.II, Sao Paulo, Brazil, 253-256, 1995.
9. Felix, R., Kühlen, J. and Albersmann, R., "The Optimization of Capacity Utilization with a Fuzzy Decision Support Model", Proceedings of the fifth IEEE International Conference on Fuzzy Systems, New Orleans, USA, 997-1001, 1996.
10. Felix, R., "On fuzzy relationships between goals in decision analysis", Proceedings of the 3rd International Workshop on Preferences and Decisions, Trento, Italy, 37-41, 2000.
11. Felix, R., "Fuzzy decision making with interacting goals applied to cross-selling decisions in the field of private customer banking", Proceedings of the 10th IEEE International Conference on Fuzzy Systems, Melbourne, Australia, 2001.
12. Felix, R., "Efficient Decision Making with Interactions Between Goals", Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, Honolulu, Hawaii, USA, 2007.
13. Freeling, A. N. S., "Possibilities Versus Fuzzy Probabilities - Two Alternative Decision Aids", Studies in Management Sciences, Volume 20, Fuzzy Sets and Decision Analysis, 67-82, 1984.
14. Grabisch, M., “Fuzzy integral in multicriteria decision making”, Fuzzy Sets and Systems, 69, 279-298, 1995.
15. Modave, F., Dubois, D., Grabisch, M. and Prade, H., “A Choquet integral presentation in multicriteria decision making”, AAAI Fall Symposium, Boston, Ma, November 1997.
16. Parcus de Koch, N., "Software Engineering for Adaptive Hypermedia Systems", PhD Dissertation, Ludwig-Maximilians-Universität, Munich, Department of Mathematics and Computer Sciences, 2000.
17. Saaty, T. L., "The Analytic Hierarchy Process", Mc Graw-Hill, 1980.
18. Spiegel, J., Polyzotis, N., "Graph-Based Synopses for Relational Selectivity Estimation", SIGMOD 2006, June 27-29, Chicago, Ill., 2006.
19. Torra, V., "Weighted OWA operators for synthesis of information“, Proceedings of the fifth IEEE International Conference on Fuzzy Systems, New Orleans, USA, 966-971, 1996.
20. Yager, R.R., “Families of OWA operators”, Fuzzy Sets and Systems, 59, 125-148, 1993.
21. Zimmermann, H. J., "Fuzzy Set Theory and its Applications", Kluver/Nijhoff, 1991.

Using Fuzzy Logic to Handle the Users' Semantic Descriptions in a Music Retrieval System

Micheline Lesaffre and Marc Leman

IPEM, Institute for Psychoacoustics and Electronic Music, Department of Musicology,
Ghent University, Blandijnberg 2, 9000 Ghent, Belgium
Micheline.Lesaffre@UGent.be, Marc.Leman@UGent.be

Abstract. This paper provides an investigation of the potential application of fuzzy logic to semantic music recommendation. We show that a set of affective/emotive, structural and kinaesthetic descriptors can be used to formulate a query which allows the retrieval of intended music. A semantic music recommendation system was built, based on an elaborate study of potential users of music information retrieval systems. In this study analysis was made of the descriptors that best characterize the user's understanding of music. Significant relationships between expressive and structural descriptions of music were found. A straightforward fuzzy logic methodology was then applied to handle the quality ratings associated with the descriptions. Rigorous real-world testing of the semantic music recommendation system revealed high user satisfaction.

Keywords: semantic description, music information retrieval, user profile, music recommendation, query by emotion, fuzzy systems.

1 Introduction

Research on content-based music information retrieval aims at defining the search and retrieval of music in terms of semantic descriptors. Rather than having to specify the name of the composer or the title of the song, semantic description would allow one to specify musical content using descriptors such as `happy', 'sad', 'dynamic' and 'harmonious'. Such descriptions focus on high-level properties, whose meaning ranges from structural to kinaesthetic to affective/emotive qualities of the music [1, 2]. However, one of the weaknesses of this approach is that most often, there is a lack of knowledge about the user's background, such as education, gender, familiarity with music. Semantic descriptions are meant to function in a social context and the meaning of semantic descriptors is often determined by tacit knowledge about the user's intentions, the user's background and the common cultural context in which the communication is taking place. As a result, there is a semantic gap between user and system. Semantic descriptors of music are meant to mediate between the user's verbally described search intention and the audio contained in a music library, yet the system lacks the tacit knowledge about intentions, background and common cultural context.

Up to now, most solutions to the problem of the semantic gap are based on systems that correlate extracted audio features with semantic descriptors, using techniques

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 89–98, 2007.
springerlink.com

based on probabilistic learning methods (e.g. [3]). However, such mappings often assume the homogeneity of the users involved. However, in practice, users may group into categories, or users may use semantic descriptors in a particular way, depending on subjective factors such as education and gender. Therefore, content-based music search and retrieval cannot be fully accomplished when the particularities of users are not taken into account. What is needed is (1) a better definition of the users of such systems, (2) better and more elaborate databases with semantic annotations of music, (3) better tools for handling flexible processing of semantic descriptions and (4) better tools for system evaluation.

This paper consists of four parts. In the first part a brief overview is given of related work on semantic description of music. The second part addresses a user study that preceded the development of the semantic music recommendation system. In the third part the use of fuzzy logic to flexible querying is explained. Finally, in the fourth part, the testing of the semantic music recommendation system in the real-world is discussed.

2 Background

During the last decade, the fuzzy logics field has witnessed a tremendous growth in the number of applications across a wide variety of domains that focus on humanlike behavior. It is possible that in the near future, the Semantic Web will be a major field of applications of fuzzy logic [4]. However, to the best of our knowledge, there are no music recommendation systems available that use fuzzy logics to handle semantic descriptions of affective/emotive, structural and kinaesthetic features of music provided by users of music information retrieval systems.

Usability is a topic of interest in the musical digital library community. Although the importance of interface and system usability is acknowledged [5], it has only recently been suggested that users themselves should be consulted. Previous studies rather focus on trying to find out what people do and what people would like to do with music. These studies involve, for example, analyzing music queries posted to Usenet News [6] and to the Google ask-an-expert service [7] or watching people's behavior in CD stores [8]. The usability of existing systems and various methodologies, however, has not been tested with real music information retrieval users. Indeed, the most common method used for studying usability is laboratory-based testing [9].

So far, the use of semantic descriptors for music is based on two approaches. Linking approaches aim at collecting the users' descriptions of music in application contexts. Kalbach [10] praises the innovative character of these linking approaches, because they are based on a large population of users dedicated to search and retrieval of music. Yet the semantic description often relies on an ad hoc taxonomy (e.g. MoodLogic, http://www.moodlogic.com/). In contrast, annotation approaches collect the user's description of music in pursuit of system evaluations and algorithm testing (e.g. [11, 12, 13]). Unfortunately, most studies provide scarce reference material in terms of how these ratings were obtained, and how representative the population of users was, despite requests for more input from psychologists and musicologists [14].

A specific field of interest concerns the relationship between different categories of semantic description. In this context, a number of studies have explored the relationship between descriptions of musical structure and descriptions of emotional appraisal (e.g. [15, 16, 17]). The latter form an important sub-category of the category of semantic descriptions. Most studies reveal that semantic/emotive descriptors rely on a number of subjective factors. Yet, these studies are often not related to music information retrieval and therefore they suffer from a lack of representative population and musical excerpts.

The present research expands on earlier studies carried out by Leman et al. [1, 18]. In these studies, descriptions of emotional and affect appraisal of music were collected from a group of university students, while descriptions of musical structural were collected from a group of musicologists. These studies have been expanded by recruiting and involving a large set of users that are potentially interested in content-based music information retrieval.

3 Foregoing Users Study

A large-scale study has been set up, which consisted of two parts. In the first part, a survey of the demographic and the musical background of potential users of music information retrieval systems was carried out. In the second part, a representative set of these users was asked to annotate music by means of semantic descriptions. The study provided a large database that was then used to build a semantic music recommendation system.

3.1 Global Setup

The survey resulted in a dataset with information about personal, demographic and musical background of 774 participants. From this group, a sample of subjects was recruited for the annotation experiment. This provided an annotation dataset with semantic descriptions (i.e. quality ratings) of 160 music excerpts. The latter were selected from 3021 titles of the favorite music of the participants in the survey. The music stimuli (i.e. excerpts of 30 seconds) thus reflected the musical taste of the targeted population. 79 subjects rated the whole set of 160 musical excerpts. In the annotation experiment, a representative population of users, described music using a set of semantic adjectives. Our model distinguished between affective/emotive, structural and kinaesthetic descriptors. Apart from this, for each of the 160 rated musical excerpts, subjects were also asked to give additional information on how familiar they were with the music they heard and what was their personal judgment. (See Lesaffre [2] for a detailed description).

3.2 Summary of Results

Survey

With 774 participants, a representative sample of the targeted population was reached and a global profile of the envisaged users of content-based music information retrieval systems could be defined. The average music information retrieval system users: are younger than 35 (74%); use the Internet regularly (93%); spend 1/3 of

Internet time on music related activities; do not earn their living with music (91%); are actively involved with music; have the broadest musical taste between 12 and 35; have pop, rock and classical as preferred genres; are good at genre description; have difficulties assigning qualities to classical music and assign most variability to classical music.

Multiple relationships between the categorical variables gender, age, musical background, and musical taste were found. For example, it was found that: of users who cannot sing 74% are men; of users who can dance very well 93% are women; of classical music listeners 70% are music experts; of musically educated users 86% play an instrument; of users older than 35 years 74% listen to classical music.

Annotation experiment

There was a significant influence of demographic and musical background such as gender, age, musical expertise, broadness of taste, familiarity with classical music and active musicianship on the use of semantic descriptors. For example, men rated the musical excerpts more restrained, more harmonious and more static, whereas women judged the music more beautiful and more difficult. Subjects older than 35 found the music more passionate and less static than younger listeners did. Lay listeners judged the music as being more cheerful, passionate and dull than experts did. Equal results were found for the influence of musicianship. People with a broad musical taste judged the music to be more pleasing and more beautiful than those with a narrow taste. Familiarity with the music is highly significant for all affective/emotive descriptors. The above results led to a categorization of users in four different groups, based on education (musical and non-musical) and gender (male and female).

Factor analysis revealed that several affective/emotive descriptors were correlated and that three dimensions may account for a large proportion of the variance, namely *high intense experience*, *diffuse affective state* and *physical involvement*. These factors are closely related to the dimensions *Interest*, *Valence* and *Activity* uncovered in previous research [18]. In a similar way, the structural descriptors also revealed three dimensions. With regard to unanimity among the descriptors subjects agreed most on loudness and tempo, whilst less on timbre and articulation.

Interesting relationships were found between affective/emotive and structural descriptors. There is a strong correlation between the appraisal descriptor (tender-aggressive) and the structural descriptor loudness (soft-hard). This result is suggestive of the possibility to decompose semantic descriptors in terms of structural descriptors, which mediate the connection with acoustical descriptors.

4 Semantic Music Recommendation System

A semantic music recommendation system was built based on the results of the foregoing user study. The system incorporates the annotations, that is, the ratings of semantic descriptors made by the participants in the experiment. In the present study we had to deal with vagueness that arose from the quality ratings which used concepts like 'rather', 'moderate' and 'very'. In the case of qualitative adjectives there are semantic ambiguities between levels and there exists no definite threshold for which an emotion becomes too 'passionate' or 'carefree'. Rather we have to differentiate between descriptors which are perfectly acceptable for the user. Obviously, the meaning of vague expressions like 'rather passionate' is user dependent. In this context, the multi-valued logic of fuzzy logic was considered as a possible option to account for the vague descriptors. The interest of using fuzzy logic for a user is a better representation of the user's preferences.

4.1 Design and Procedure

An interface of the semantic music recommender tool was designed for use at exhibitions and other testing environments that address different user populations. The tool basically consists of four parts: (1) definition of the user profile (gender and musical interest); (2) specification of the search query using semantic descriptors; (3) recommendation of music, using the music database and (4) evaluation tasks.

The search screen presents four categories of semantic descriptors, allowing any combination of choices between (1) five genre categories (classical, pop/rock, folk/country, jazz and world/ethnic), (2) eight emotion labels (cheerful, sad, tender, passionate, anxious, aggressive, restrained and carefree), (3) four adjective pairs referring to sonic properties of music (soft-hard, clear-dull, rough-harmonious and void-compact) and (4) three adjective pairs reflecting movement (slow-quick, flowing-stuttering and dynamic-static).

The output is a hierarchically ordered list with music titles. The user can browse the list and listen to the music. Each time a user listens to a recommended piece of music a popup window provides the user with individual scores for each descriptor in the query. These scores reflect the agreement among the participants in the experiment.

In addition to the recommendation of music, two assessment tasks are included (see below, Real-world testing). First, the user is requested to assign a degree of satisfaction in using the system for the particular search task, after having listened to a recommended piece of music. Secondly, the user is asked to evaluate the general usability of content-based querying and of the distinct descriptor sets.

4.2 Fuzzy Logic Functions

To deal with the problem of interfacing linguistic categories, such as adjectives, with numerical data and for expressing user's preference in a gradual and qualitative way, ideas of fuzzy logics are used. To the best of our knowledge, fuzzy set methods have not been applied yet to the representation of annotations provided by real users of music information retrieval systems. In what follows, a description is given of the fuzzy logic functions of the semantic music recommender system. This component was built using Visual Basic .NET for Microsoft Access Databases. Fuzzy logic was applied in three steps. Firstly, fuzzy functions, which account for the vagueness of the semantic descriptors, were calculated per semantic descriptor and per user profile. Secondly, scores were calculated per music excerpt, semantic descriptor and user profile. Thirdly, combined scores were calculated.

Fuzzy functions per semantic descriptor and user profile

The semantic music recommendation takes into account four different types of users, based on gender (male, female) and musical expertise (expert, novice). As a consequence, for each adjective, four fuzzy functions were calculated. Each function is characterized by three numbers, namely, the 25th, 50th and 75th percentile values of the cumulative rating value. To obtain that function the rating values attributed by all the subjects who fit a specific profile (i.e. female novice, female expert, male novice and male expert) were sorted in ascending order. After that, the 3 values according to the cumulative percentages of 25%, 50% and 75% were calculated. These 3 values each define a fuzzy function score. Then, the cumulative distribution function is built using the number of ratings given by a user group for a semantic descriptor for five data points (i.e. not, little, moderate, rather, very). From this discrete set of data points a new fuzzy function is built on a set of three data points (i.e. the three fuzzy function scores).

The 3 values v1, v2 and v3 define the following fuzzy function score:

IF x ≤ v1 THEN score(x) = 0
IF v1 < x ≤ v2 THEN score(x)=0,5[(x – v1)/(v2 – v1)] (a number between 0 and 0,5)*
IF v2 < x ≤ v3
THEN score(x)=0,5+0,5[(x – v2)/(v3 –v2)] (a number between 0,5 and 1)*

v1, v2 and v3 are calculated as follows:

IF “rating x” = 0 : x / frequencyRatings(0)
ELSE “rating x”+(x - frequencyRatings (0.. ”rating value x” - 1))/frequencyRatings (”rating value x”) to which frequencyRatings (0..y) = the number of the rating values ≤ y

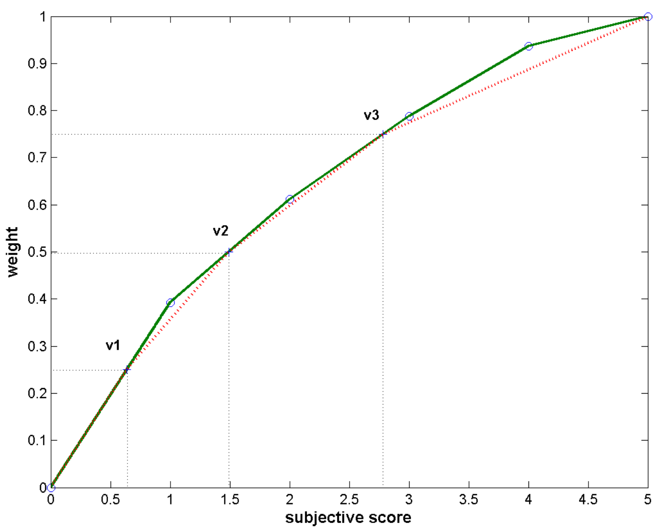


Fig. 1. Cumulative function and fuzzy function

In what follows an example of the calculation of v1, v2 and v3 for the profile 'male expert' and the descriptor 'cheerful' is given. Figure 1 shows a plot of the fuzzy function for this example.

The cumulative distribution function (see Fig. 1 full line) is built on the number of ratings given by male experts for 'cheerful'.

not *[frequencyRatings(0)] = 1252*;
little *[frequencyRatings(1)] = 698*;
moderate *[frequencyRatings(2)] = 564*;
rather *[frequencyRatings(3)] = 474*;
very *[frequencyRatings(4)] = 202*.

The total number of evaluations = 3190. From this discrete set of five known data points (1252, 698, 564, 474 and 202) a new, fuzzy function is built (see Fig. 1 dotted line) on a set of three data points (v1, v2, and v3).

Calculation of recommendation scores per music excerpt, adjective and profile

In order to determine the recommendation scores per music excerpt, the rating values attributed by all subjects who fit a specific user profile were sorted in ascending order as to semantic descriptor and excerpt number respectively. After that, the cumulative median value was calculated. The score for each adjective, profile and excerpt resulted in the following function value: *score(median)* with score being the fuzzy function corresponding to the adjective and profile concerned.

In what follows an example of the calculation of the score for music excerpt one (J.S. Bach, Kommt, ihr Töchter, helft mir klagen, from Matthäus-Passion, BWV 244), the profile 'male expert' and the descriptor 'cheerful' is given.

not *[frequencyRatings(0)] = 12;*
little *[frequencyRatings(1)] = 6;*
moderate *[frequencyRatings(2)] = 0;*
rather *[frequencyRatings(3)] = 2;*
very *[frequencyRatings(4)] = 0.*

The total number of evaluations =20.
Calculation for 1/2(x = 20 / 2 = 10):
"rating value 10" = 0 (< 12), thus :
median = 10 / frequencyRatings (0) = 10 / 12 = 0,83
*score(median) = score(0,83) = 0.5 * (0,83 – v1) / (v2 – v1),*
then v1 = 0,64 < 0,83 < v2 = 1,491...
*=> score(0,83) = 0.5 * (0,83– 0,64) / (1,49– 0,64) = 0,11*

Calculation of combined recommendation scores per music excerpt

If no adjectives are selected, the combined recommendation score is 1. If one adjective is selected then the combined recommendation score equals the score for the semantic descriptor concerned. If multiple (n) adjectives are selected then the combined recommendation score equals the nth power of the product of the adjective scores.

In what follows an example of the calculation of combined scores for music excerpt one, the profile ‘male expert’ and the descriptors ‘cheerful’, ‘sad’ and ‘passionate’ is given. Cheerful = 0,11; sad = 1; passionate = 0,97

Case 1: ‘cheerful’ and ‘sad’ are selected:
*score = square root(0,115 * 1) = 0,34*

Case 2: ‘cheerful’, ‘sad’ and ‘passionate’ are selected:
*score = 3th power root(0,11 * 1 * 0,97) = 0,48*

5 Real-World Testing

The semantic music recommendation system was tested in three different real-world environments, addressing four different types of users. The system was first tested by 626 trade fair visitors (i.e. ACCENTA 2005), then by 20 intellectuals (i.e. ALUMNI 2006), and finally by 34 school children and 119 science fair visitors (i.e. Wetenschapsfeest 2006). The tests aimed at evaluating the effectiveness of the fuzzy logic approach to semantic descriptors. In other words, we investigated whether users would agree with the judgments made by participants in our experimental study. We were interested in their assessment of the usability of the system and the descriptor sets.

Quantitative analysis of the users’ satisfaction ratings showed that on the average three quarter of all users were satisfied with the recommendation system (‘well’ to ‘very well’). There are some minor differences between four user groups. However, children reported that 31% of the recommendations did not or little matches their expectations. This can be explained by the fact that the music in the database is too much ‘middle of the road’ for this population. In their drive to hear the music they like, they might not have taken the satisfaction rating task very seriously.

From the query behavior of the four groups of users we learned that their first preference is for affective/emotive descriptors (50% selected by trade fair visitors up to 57% by school children). Second preferred are sonic descriptors (24% selected by intellectuals to 28% by trade fair visitors). Third preferred are movement descriptors (18% selected by school children to 23% by trade fair visitors).

Although several group dependent differences were found, for all groups the most selected emotion descriptor is ‘cheerful’. School children for example tend to search for music that is ‘aggressive’, ‘restrained’ and ‘anxious’ whereas trade fair visitors search for music that is ‘passionate’, ‘tender’ and ‘carefree’. As similar observation was made for sonic descriptors in that there is an overall agreed on interest in ‘clear’ music. Instead, school children search for ‘hard’ and ’rough’ music whereas trade fair visitors search for ‘harmonious’ and ‘soft’ music. For movement descriptors most agreement is on the descriptor ‘dynamic’. School children however like to find ‘quick’ music whereas trade fair visitors prefer music that is ‘flowing.

Over 90% of the trade fair visitors, intellectuals and science fair visitors responded positively to the overall usability of the system, except from the school children (82%). With regard to the usability of the semantic descriptor sets, affect/emotive descriptors are found most useful, followed by movement descriptors and after everything else sonic descriptors.

6 Conclusion

In this paper we described the development and real-world testing of a music information retrieval system that uses fuzzy logic for handling the vagueness of semantic descriptors used for music annotation. Our results show that a fuzzy logic methodology, combined with a user-oriented approach to music information retrieval, may be effective for the development of a content-based music recommendation system. The study reveals that the framework of affective/emotive, structural and kinaesthetic descriptors has an inter-subjective basis whose vagueness can be handled with fuzzy logics.

Users who tested the system in a real-world environment confirmed the usability of semantic-based music information retrieval systems. Even if the tests that were carried out with different user groups showed user dependencies, in mainly 75% of the cases users were satisfied with the music that was recommended.

Our study is suggestive of applying fuzzy logic to a predefined semantic descriptor set. It can be assumed that this methodology may provide a stable basis for further development of content-based access to music.

Acknowledgements

This research has been conducted in the framework of the MAMI (Musical Audio Mining) project (2002-2005), and the DEMCO (New Methods for the description of Musical Content) project (2004-2007). The authors wish to thank Prof. Dr. B. De Baets, Prof Dr. H. De Meyer and MA K. Vermeulen for their assistance with the development of the MAMI research tool.

References

1. Leman, M., Vermeulen, V., De Voogdt, L., Taelman, J., Moelants, D., & Lesaffre, M. (2004). Correlation of gestural musical audio cues and perceived expressive qualities. In A. Camurri & G. Volpe (Eds.), Gesture-based communication in human-computer interaction (40-54). Berlin Heidelberg: Springer-Verlag.
2. Lesaffre, M., (2005). « Music Information Retrieval. Conceptual framework, Annotation and User Behaviour". PhD. dissertation, Ghent University.
3. Whitman, B., (2005) "Learning the meaning of music", Ph.D. dissertation, Department of Architecture, Massachusetts Institute of Technology.
4. Straccia, U., (2005) *A fuzzy description logic for the semantic web*. In Sanchez, E., ed.: Capturing Intelligence: Fuzzy Logic and the Semantic Web. Elsevier
5. Arms, W. Y. (2000). Digital Libraries. Cambridge, MIT Press
6. Downie, J. S., and Cunningham, S.J. (2002) Toward a theory of music information retrieval queries: system design implications. Proceedings of the International Symposium on Music Information Retrieval (ISMIR02), Paris (France), pp 299 - 300.

7. Bainbridge, D., Cunningham, S. J., & Downie, J. S. (2003). How people describe their music information needs: a grounded theory analysis of music queries. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR03)*, Baltimore, pp 221-222.
8. Cunningham, S. J. (2002). User studies: a first step in designing a MIR testbed. In *The MIR/MDL Evaluation Project White Paper Collection,* 2nd ed., pp 17-19.
9. Covey, D. T. (2002). *Usage and usability assessment: library practices and concerns.* Washington, DC: Digital Library Federation and Council on Library and Information Resources. (CLIR Report105).
10. Kalbach, J. (2002). Classifying emotion for information retrieval: three websites. *Notes, 59*(2), 408-411.
11. Tzanetakis, G., & Cook, P. (2000). *Experiments in Computer-Assisted Annotation of Audio.* Paper presented at the International Conference on Auditory Display (ICAD), Atlanta, Georgia, USA.
12. Yang, D., & Lee, W. (2004). Disambiguating Music Emotion Using Software Agents. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR04*), Barcelona, 52-58
13. Lesaffre, M., Leman, M., De Baets, B., & Martens, J.-P. (2004). Methodological Considerations Concerning Manual Annotation of Musical Audio in Function of Algorithm Development. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR04),* Barcelona,pp 64-71.
14. Futrelle, J., & Downie, J. S. (2002). Interdisciplinary communities and research issues in Music Information Retrieval. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR02),* Paris, pp 215-221.
15. Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson & H. H. Goldsmith & K. H. Scherer (Eds.), *Handbook of affective sciences*. New York: Oxford University press, pp 503-534.
16. Gabrielsson, A., & Lindström, S. (2001). The Influence of Musical Expression on Emotion in Music. In P. Juslin & J. Sloboda (Eds.), *Music and Emotion: Theory and Research*. New York: Oxford University Press, pp 223-248.
17. Juslin and Sloboda Juslin, P. N., & Sloboda, J. A. (2001). *Music and Emotion: Theory and Research*. New York, Oxford: Oxford University Press.
18. Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesaffre, M. (2005). Prediction of Musical Affect Attribution Using a Combination of Structural Cues Extracted from Musical Audio. *Journal of New Music Research, 34*(1), pp 39-67.

Part III

Philosophical and Human-Scientific Aspects of Soft Computing

Between Empiricism and Rationalism: A Layer of Perception Modeling Fuzzy Sets as Intermediary in Philosophy of Science

Rudolf Seising

Medical University of Vienna, Core Unit for Medical Statistics and Informatics, Spitalgasse 23, A-1090 Vienna, Austria
Rudolf.seising@meduniwien.ac.at

Abstract. In philosophy of science we find two epistemological traditions: rationalism and empiricism. Rationalists believe that the criterion of knowledge is not sensory but intellectual and deductive whereas from the empiricist point of view the source of our knowledge is sense experience. Bridging this gap between these theories of knowledge has been a problem in philosophical approaches, both past and present. This philosophical paper focuses on using fuzzy sets and systems (FSS), computing with words (CW), and the computational theory of perceptions (CTP) as methodologies to help bridge the gap between systems and phenomena in the real world and scientific theories. It presents a proposal in which fuzzy methods are used to extend the so-called structuralist view of scientific theories in order to represent the relation of empiricism and theoretical structures in science.

Keywords: Philosophy of science, epistemology, rationalism, empiricism, fuzzy sets, computing with words, computational theory of perceptions, theory, reality, perceptions.

1 Introduction

In science we have a traditional division of work: on the one hand we have fundamental, logical and theoretical investigations and on the other hand we have experimental and application side examinations. The theoretical work in science is using logics and mathematics to formulate axioms and laws. It is linked with the philosophical view of rationalism whereas the other aspects of science using experiments to find or prove or refute natural laws have their roots in the philosophical empiricism.

In both directions – from experimental results to theoretical laws or from theoretical laws to experimental proves or refutations – scientists have to bridge the gap that separates theory and practice in science.

Beginning as early as the 17^{th} century, a primary quality factor in scientific work has been a maximal level of exactness. Galileo and Descartes started the process of giving modern science its exactness through the use of the tools of logic and mathematics.

The language of mathematics has served as a basis for the definition of theorems, axioms, definitions, and proofs. The works of Newton, Leibniz, Laplace and many others led to the ascendancy of modern science, fostering the impression that scientists were able to represent all the facts and processes that people observe in the

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 101–108, 2007.
springerlink.com

world, completely and exactly. But this optimism has gradually begun to seem somewhat naïve in view of the discrepancies between the exactness of theories and what scientists observe in the real world.

From the empiricist point of view the source of our knowledge is sense experience. John Locke used the analogy of the mind of a newborn as a "tabula rasa" that will be written by the sensual perceptions the baby has later. In Locke's opinion this perceptions provide information about the physical world. Locke's view is called "material empiricism" whereas the so called idealistic empiricism was hold by Berkeley and Hume: there exists no material world, only the perceptions are real.

This epistemological dispute is of great interest for historians of science but it is ongoing till this day and therefore it is of great interest for today's philosophers of science, too. Searching a bridge over the gap between rationalism and empiricism is a slow-burning stove in the history of philosophy of science. In this paper, Lotfi Zadeh's hierarchy stack of methodologies, fuzzy sets and systems (FSS), computing with words (CW) and the computational theory of perception (CTP), is recommended to build a bridge over this gap.

In my original research work on the history of the theory of fuzzy sets and systems (FSS) I could show that Lotfi A. Zadeh established this new mathematical theory in 1964/65 to bridge the gap that reflects the fundamental inadequacy of conventional mathematics to cope with the analysis of complex systems [1, 2, 3].

In the last decade of the 20th century Zadeh set up computing with words (CW) [4] and the computational theory of perceptions (CTP) [5, 6] and he erected the methodologies of CTP and CW on the basic methodology of FSS.

In this non-historical but philosophical paper this methodology stack for bridging the gap between real and theoretical systems will be examined from a philosophical point of view. To this end, the so-called structuralist approach of scientific theories in the philosophy of science will first be reviewed in section 2 and then this approach will be modified in section 3 – i.e. it will be "fuzzified" – by extending the structuralist framework with fuzzy sets and fuzzy relations to model perceptions of observers. This approach provides a new view of the "fuzzy" relationship between empiricism and theory.

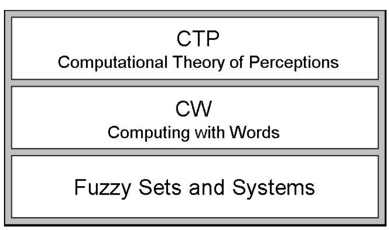


Fig. 1. Lotfi Zadeh's hierarchical stack of methodologies: FSS, CW, CTP

2 The Structuralist View of Theories

Two trends in obtaining systematic rational reconstructions of empirical theories can be found in the philosophy of science in the latter half of the 20th century: the *Carnap* *approach*[1] and the *Suppes approach*[2]. In both, the first step consists of an axiomatization that seeks to determine the mathematical structure of the theory in question. However, whereas in the Carnap approach the theory is axiomatized in a formal language, the Suppes approach uses informal set theory. Thus, in the Suppes approach, one is able to axiomatize real physical theories in a precise way without recourse to formal languages. This approach can be traced back to Patrick Suppes' proposal in the 1950s to include the axiomatization of empirical theories of science in the metamathematical programme of the French group "Bourbaki" [7].

Later, in the 1970s, Joseph D. Sneed[3] developed informal semantics meant to include not only mathematical aspects, but also application subjects of scientific theories in this framework, based on this method. In his book [8], Sneed presented the view that all empirical claims of physical theories have the form "x is an S", where "is an S" is a set-theoretical predicate (e.g., "x is a classical particle mechanics"). Every physical system that fulfils this predicate is called a model of the theory. For example, the class M of a theory's models is characterized by empirical laws that consist of conditions governing the connection of the components of physical systems. Therefore, we have models of a scientific theory, and by removing their empirical laws, we get the class M_p of so-called potential models of the theory. Potential models of an empirical theory consist of *theoretical terms*, i.e. observables with values that can be measured in accordance with the theory. This connection between theory and empiricism is the basis of the philosophical "problem of theoretical terms".

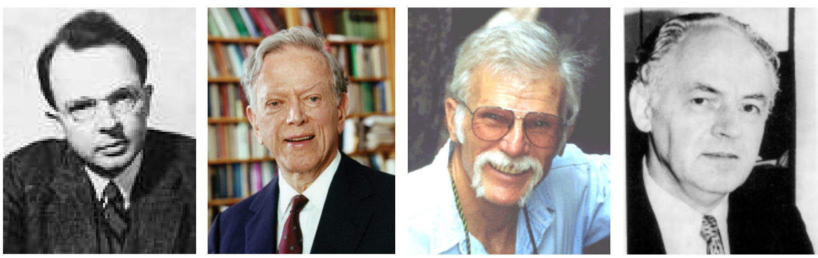


Fig. 2. From left to right: Rudolf Carnap, Patrick Suppes, Joseph Sneed, Wolfgang Stegmüller

If we remove the theoretical terms of a theory in its potential models, we get structures that are to be treated on a purely empirical layer; we call the class M_{pp} of these structures of a scientific theory its "partial potential models". Finally, every physical theory has a class I of intended systems (or applications) and, of course, different intended systems of a theory may partially overlap. This means that there is a class C of

[1] The German philosopher Rudolf Carnap (1891-1970) was a professor in Vienna (1926-1931) and a member of the Vienna Circle. He was a professor in Prague (1931-1935), Chicago (1936-1952), at the Institute for Advanced Study in Princeton (1952-1954), and at the University of California in Los Angeles (1954-1970).

[2] The American mathematician and philosopher Patrick Suppes (born in 1922) was and is a professor at Stanford University in the USA.

[3] The American physicist and philosopher Joseph D. Sneed is a professor at the Colorado School of Mines in the USA.

constraints that produces cross connections between the overlapping intended systems. In brief, this structuralist view of scientific theories regards the core K of a theory as a quadruple $K = \langle M_p, M_{pp}, M, C\rangle$. This core can be supplemented by the class I of intended applications of the theory $T = \langle K, I\rangle$.[4] To make it clear that this concept reflects both sides of scientific theories, these classes of K and I are shown in Fig. 2. Thus we notice that M_{pp} and I are entities of an empirical layer, whereas M_p and M_{pp} are structures in a theoretical layer of the schema.

Now this approach of the structuralist view of theories will be extended by using fuzzy sets and fuzzy relations to represent perceptions as important components in the interpretation of scientific theories. This will be very suitable in future invetstigations in the philosophy of science, because in new theories of the 20th century, such as relativity theory and quantum mechanics in physics, the observer and his/her perceptions play a central and important role [9].

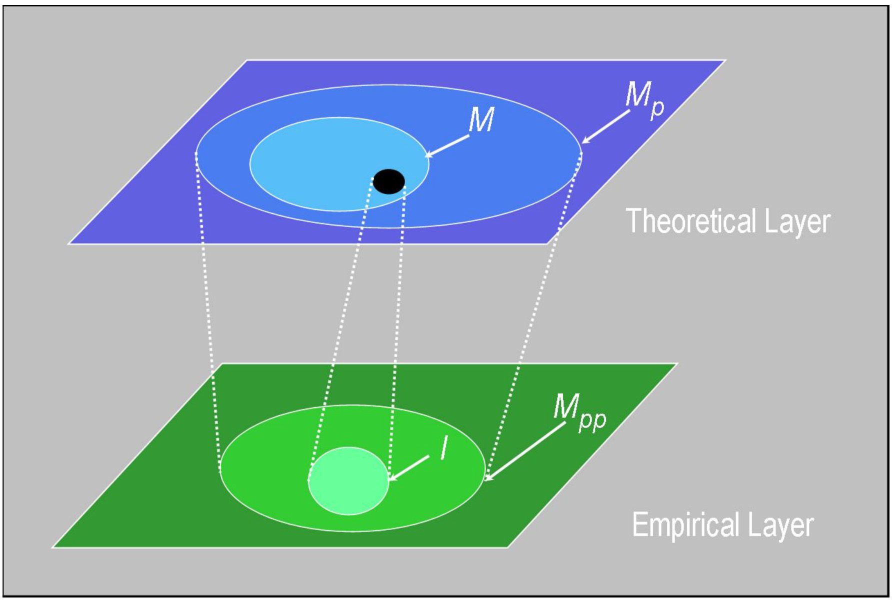


Fig. 3. Empirical and theoretical structural layers in the analysis of scientific theories

3 A Fuzzy Layer as Intermediary in Philosophy of Science

The proposed modification of the structuralist approach in philosophy of science pertains to the empirical layer in Fig. 3. A distinction can be made between real systems and phenomena, on the one hand, and perceptions of these entities, on the other. Thus a lower layer – the real layer – is introduced and the former empirical layer is renamed the "fuzzy layer", as the partial potential models and intended systems are not

[4] Sneed, Wolfgang Stegmüller, C. Ulises Moulines, and Wolfgang Balzer, developed this view into a framework intended to analyze networks of theories and the evolution of theories [10].

real systems because a minimal structure is imposed by the scientist's observations. These are perception-based systems and thus must be distinguished from real systems and phenomena that have no structure before someone imposes one upon them.

Now there is a layer of perceptions between the layer of real systems and phenomena and the layer of theoretical structures. In accordance with Lotfi A. Zadeh's computational theory of perceptions (CTP), perceptions in this intermediate layer can be represented as fuzzy sets. Whereas measurements are crisp, perceptions are fuzzy, and because of the resolutions achieved by our sense organs (e.g. aligning discrimination of the eye), perceptions are also granular – in 2001 Zadeh wrote in the *AI Magazine*: "perceptions, in general, are both fuzzy and granular or, for short *f*-granular [6]. Fig. 4 shows Zadeh's depiction of *crisp* (*C*) and *fuzzy* (*F*) granulation of a linguistic variable.

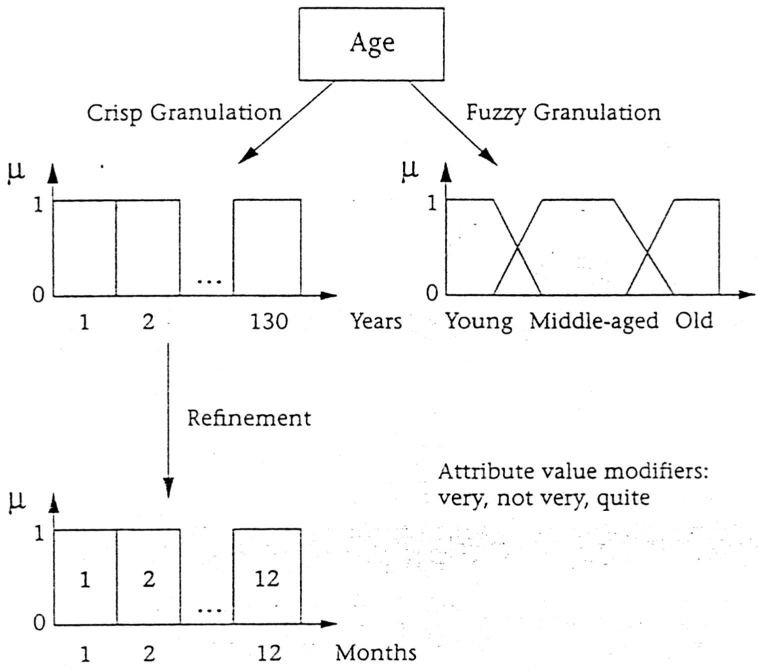


Fig. 4. Empirical and theoretical structural layers in the analysis of scientific theories [6]

When Zadeh established CTP on the basis of computing with words (CW), which in turn is based on his theory of fuzzy sets and systems [5], he earnestly believed that these methodologies would attain a certain importance in science: "In coming years, computing with words and perceptions is likely to emerge as an important direction in science and technology." [15]. Taking Zadeh at his word, his methodologies of fuzzy sets and computing with words and perceptions are here incorporated into the structuralist approach in the philosophy of science. As discussed above, a fuzzy layer of perceptions is inserted between the empirical layer of real systems and phenomena, and the theoretical layer, where there are structures of models and potential models. Thus the relationship of real systems and theoretical structures has two dimensions: fuzzification and defuzzification.

3.1 Fuzzification

Measurements are crisp and perceptions are fuzzy and granular. To represent perceptions we use fuzzy sets, e.g. A^F, B^F, C^F, ... It is also possible that a scientist observes not just a single phenomenon, but interlinked phenomena, e.g. two entities move similarly or inversely, or something is faster or slower than a second entity, or is brighter or darker, or has an analogous smell, etc. Such relationships can be characterized by *fuzzy-relations* f^F, g^F, h^F, ….

3.2 Defuzzification

"Measure what is measurable and make measurable what is not so" is a sentence attributed to Galileo. In modern scientific theories this is the way to get from perceptions to measurements or quantities to be measured. Here this transfer is interpreted as a defuzzification from perceptions represented by fuzzy sets A^F, B^F, C^F, … and relations between perceptions represented by fuzzy relations f^F, g^F, h^F, … to ordinary (crisp) sets A^C, B^C, C^C, … and relations f^C, g^C, h^C, … These sets and relations are basic entities for the construction of (potential) models of a scientific theory in the theoretical layer.

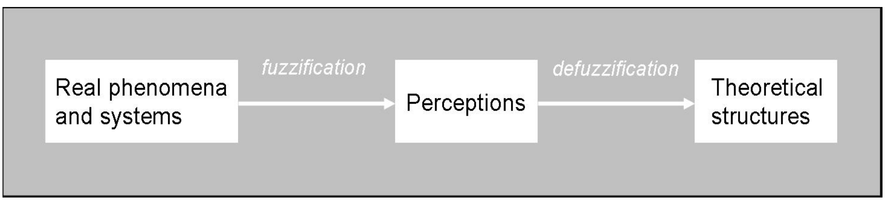


Fig. 5. Fuzzification and defuzzification in the fuzzy structuralist view

3.3 Theoretization Results from Fuzzification and Defuzzification

The serial operation of fuzzification and defuzzification (see Fig. 5) yields the operation of a relationship T that can be called "initial theoretization", because it transfers phenomena and systems in the real (or empirical) layer into structures in the theoretical layer (see Fig. 6).

In the structuralist view of theories, the general concept of theoretization is defined as an intertheoretic relation, i.e. a set theoretical relation between two theories T and T'. This theoretization relation exists if T' results from T when new theoretical terms are added and new laws connecting the former theoretical terms of theory T with these new theoretical terms of theory T' are introduced.

Successive addition of new theoretical terms establishes a hierarchy of theories and a comparative concept of theoriticity. In this manner the space-time theory arose from Euclidean geometry when the term "time" was added to the term "length", and classical kinematics developed from classical space-time theory when the term "velocity"

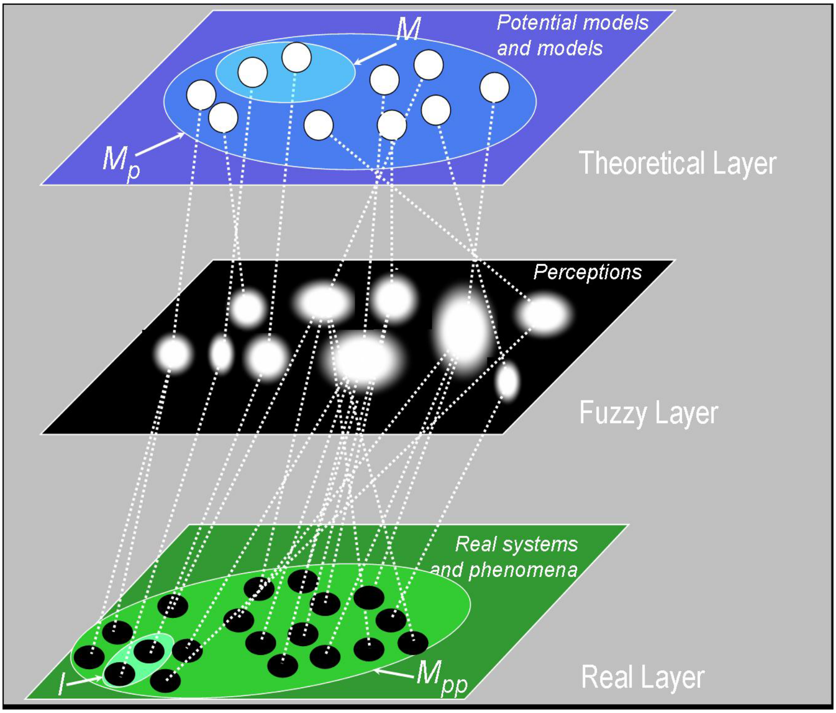


Fig. 5. Empirical, fuzzy, and theoretical layers of (fuzzy) structures in scientific research

was added. Classical kinematics turned into classical (Newtonian) mechanics when the terms "force" and "mass" were introduced.

The new theory *T'* adds a new theoretical layer to the old theory *T*. *T*-theoretical terms are not *T'*-theoretical but *T'*-non-theoretical terms, and reciprocally they may not be any of the T-non-theoretical terms. The old theory must not be changed in any way by the new theory. In this approach the higher terms are in the hierarchy, the more theoretical they are. The lower layers contain the non-theoretical base of the theory.

What is the situation in the lowest layer of this hierarchy? A theory *T* with theoretical terms and relations exists there, but it is not a theoretization of another theory. This theory *T* covers phenomena and intended systems with initial theoretical terms. This is an initial theoretization, because the *T*-theoretical terms are the only theoretical terms at this level. They have been derived directly as measurements of observed phenomena. This derivation was designated as "initial theoretization" above and it is a serial connection of fuzzification and defuzzification.

4 Conclusion

The computational theory of perceptions is an appropriate methodology to represent efforts of scientific research to bridge the gap between empirical observations and the abstract construction of theoretical structures.

In the classical, i.e. non-fuzzy, structuralist view of theories there is an empirical layer of real phenomena and systems that have some minimal structure and a theoretical layer of potential models and models that are fully structured entities. But there is no representation of the observer's role and his/her perceptions. The modified view of the structuralist approach presented in this paper as a proposal that will be worked out in detail in the near future comprises a layer of fuzzy sets and fuzzy relations as a means of dealing with the difference between real phenomena and systems on the one hand and the observer's perceptions of these real entities on the other. This extended structuralist view – which can be called the "fuzzy structuralist view" – of scientific theories may open up a new and fruitful way to understand scientific research.

Acknowledgments. I would like to thank Wolfgang Balzer (Munich), my former supervisor in the philosophy of science for his introduction to the structuralist view of theories and Dieter Straub (Munich) for important and enjoyable discussions. Many thanks go to Witold Pedrycz (Edmonton) for his encouragement to make a shift towards this interdisciplinary and science-oriented research project.

This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

1. Seising, R.: Die Fuzzifizierung der Systeme. Die Entstehung der Fuzzy Set Theorie und ihrer ersten Anwendungen – Ihre Entwicklung bis in die 70er Jahre des 20. Jahrhunderts. Stuttgart: Franz Steiner Verlag (Boethius: Texte und Abhandlungen zur Geschichte der Mathematik und der Naturwissenschaften Band 54) 2005. English edition: Seising R.: The Fuzzification of Systems. The Genesis of Fuzzy Set Theory and Its Initial Applications – Its Development to the 1970s. Berlin [u.a.]: Springer (2007, in print).
2. Seising, R.: Pioneers of Vagueness, Haziness, and Fuzziness in the 20th Century. In: Masoud Nikraves, Janusz Kacprzyk, Lotfi A. Zadeh (eds.): Forging the Frontiers I. Studies in Fuzziness and Soft Computing. New York [u.a.]: Springer (Series in Fuzziness and Soft Computing) (2007, in print).
3. Seising, R.: On the Absence of Strict Boundaries – Vagueness, Haziness, and Fuzziness in Philosophy, Medicine, and Science, Applied Soft Computing, Special Issue: Forging the Frontiers (2006, in print).
4. Zadeh, L. A.: Fuzzy Logic = Computing with Words, IEEE Transactions on Fuzzy Systems, 4, 2 (1996) 103-111.
5. Zadeh, L. A.: From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions, IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications, 45, 1 (1999) 105-119.
6. Zadeh, L. A.: A New Direction in AI. Toward a Computational Theory of Perceptions. AI-Magazine, 22, 1 (2001) 73-84.
7. Bourbaki N. (pseudo.): Elements of Mathematics: Theory of Sets, Addison-Wesley, Reading, Mass. 1968.
8. Sneed, J. D.: The Logical Structure of Mathematical Physics, Dordrecht: Reidel (1971).
9. Seising, R.: Can Fuzzy Sets be Useful to (Re)Interpret Uncertainty in Quantum Mechanics? K. Demirli and A. Akgunduz (eds.): Proceedings of the 2006 Conference of the NAFIPS, June 3-6 (2006), Montréal, Canada (CD, ISBN: 0-7803-9188-8).
10. Balzer, W., Moulines, C. U., Sneed, J. D.: An Architectonic for Science. The Structuralist Program. Dordrecht: Reidel (1987).

Ontological and Epistemological Grounding of Fuzzy Theory

I. Burhan Türkşen

Head Department of Industrial Engineering, TOBB-Economy and Technology University, Sögütözü Cad. No:43, Sögütözü 06560 Ankara/Turkey,
Tel.:+90 312 292 4068
bturksen@etu.edu.tr

Abstract. An Ontological and Epistemological foundation of Fuzzy Set and Logic Theory is reviewed in comparison to Classical Set and Logic Theory. It is shown that basic equivalences of classical theory breakdown but are re-established as weak equivalences as a containment relation in fuzzy theory. It is also stressed that the law of conservation of information is still upheld within fuzzy theory.

1 Introduction

About 40+ years ago (1965), L. A. Zadeh published "Fuzzy Sets"[13]. Later, he has further, exposed the fact that there is the theory of "Possibility" in contrast to the theory of "Probability"[9,12]. After 1978, "Fuzzy System Models" began to influence "Decision-Support System Models" in a substantial manner with the introduction of "Possibility" as opposed to "Probability" and as a new dimension of uncertainty in system modelling.

On the bases of philosophical foundations, fuzzy sets, logics and systems in comparison to classical sets, logics and systems may be contrasted in analogy to a comparison of the philosophical underpinnings of "modernism" ver sus "post-modernism" (Dan Simon, [4]).It is important, however, to point out that such a comparison is limited. While "modernism", in stressing crisp, black- white occurences, may be anologous to Classical theory, "post-modernism", in stressing only uncertainty could be partly analogous to fuzzy theory. In reality, fuzzy theory contains both the uncertainty interval of]0,1[in anology to "post-modernism" and the certainty boundaries of {0,1} in anology to "modernism", i.e., in fuzzy theory , a membership function, μ, maps occurences, X, to [0,1] interval; $\mu : X \rightarrow [0,1]$.

In another perspective, Zadeh`s "Computing With Words" [8] approach may be considered analogous to Turing`s [3] philosophy of man and machine, and Popper`s [2] repudiation of the classical observationalist-inductivist account of scientific method. It is clear that both of these two fine philosophers of science expressed their ideas within the classical perspective. Hence, once again the analogy would be limited. Zadeh`s approach is essentially stressing the fact that there is a "meaning" of words

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 109–118, 2007.
springerlink.com

that can at best be represented by fuzzy sets as "a matter of degree" and that it is context dependent. For these reasons and others, Hodge [1] states that Zadeh`s approach generates deeper roots and new understanding. In order to really comprehend these deeper insights, one needs to properly investigate the ontological and epistemological underpinnings of fuzzy theory Türkşen [5].

Ontology: 1) A branch of metaphysics concerned with the nature and relations of being. 2) A particular theory about the nature of being or the kinds of existents.

Ontology lays the ground for the structural equivalences in classical theory; whereas it lays the ground for the structural breakdown of classical equivalences in fuzzy theory. On the other hand it reveals essential Laws of Conservation based on assumptions of existences in a different manner in both theories.

At this ground level of inquiry, one ought to ask:

"What linguistic expressions capture our positions to realty?".

"What PNL, Precisiated Natural Language, (Zadeh, [8]) expressions capture our positions to reality?"

"What are the basic equivalences or the lack of them and the Laws of Conservation that capture our position to reality?"

Epistemology: The study or theory of the nature and grounds of knowledge, esp., with reference to its limits and validity.

Epistemology lays the ground work for the assessment of consistency and believability of a set of propositions by either a priory or evidentiary basis. Evidentiary basis could be subjective and/or objective.

At this second level, one ought to state general epistemological questions:

"What linguistic encoding allows us to access truth or knowledge?
"What linguistic expressions cause the assessment of truth and knowledge?".

As well, one must to ask:

What accounts as good, strong, supportive evidence for Belief? What is the degree of Belief?

This requires that we have to come up with a "good" in terms of a "matter of degree", i.,e., "Explication" of criteria of evidence or its justification.

What is the connection between a belief being well-supported by good evidence, and the likelihood that it is true? What is the degree of likelihood and its degree of truth?

This inquires into a new definition of "Validation" criteria and their assessment to a degree. In particular, we should investigate "Belief", "Plausibility", and "Probability", related assessments for "Validation".

In Table 1, ontological and epistemological levels of theoretical inquiry are shown.

Table 1. Ontological and Epistemological Levels of a Theoretical Inquiry

EPISTEMOLOGICAL LEVEL	4. How do we validate our knowledge? How do we know it is true? What criteria do we use to assess its truth-value? "What PNL expressions cause the assessment of truth and knowledge?" 3. What is our access to truth and knowledge in general? Where is knowledge and its truth to be found? How or from what are they constituted? "What PNL encoding allows us to assess truth or knowledge?"
ONTOLOGICAL LEVEL	2. What is our position or relation to that Reality (if we do assume that it exists on level 1 below)? "What PNL expressions capture our positions to reality?" 1. Is there any reality independent or partially independent of us? Does any absolute truth exist? Does fuzziness exists? "What PNL explicates reality?" "Are they crisp or fuzzy representations of linguistic variables and their linguistic connectives?" "What are the basic equivalences generated by PNL expressions for crisp or fuzzy representations?"

In Table 2, the same two levels of inquiry are shown for classical theory which is based on classical axioms that are shown in Table 3.

Table 2. Levels of Inquiry on Classical Set and Logic Theory

EPISTEMOLOGICAL LEVEL	4. Correspondence theory of Validity only Objective 3. Objectivist, empiricists, certain
ONTOLOGICAL LEVEL	2. sRo Cartesian dualism 1. Realism, crisp meaning representation of linguistic Variables and connectives are defined with two-valued sets and logic theory. Equivalences in "normal forms" together with classical laws of conservation, as well as formulae for Belief, Plausibility, Probability, etc.

Table 3. Axioms of Classical & Set & Logic Theory, where A, B are crisp, two valued, sets and c(.) is the complement, X is the universal set and ϕ is the empty set

Involution	$c(c(A)) = A$
Commutativity	$A \cup B = B \cup A$
	$A \cap B = B \cap A$
Associativity	$(A \cup B) \cup C = A \cup (B \cup C)$
	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributivity	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Idempotency	$A \cup A = A$
	$A \cap A = A$
Absorption	$A \cup (A \cap B) = A$
	$A \cap (A \cup B) = A$
Absorption by X and ϕ	$A \cup X = X$
	$A \cap \phi = \phi$
Identity	$A \cup \phi = A$
	$A \cap X = A$
Law of contradiction	$A \cap c(A) = \phi$
Law of excluded middle	$A \cup c(A) = X$
De Morgan Laws	$c(A \cap B) = c(A) \cup c(B)$
	$c(A \cup B) = c(A) \cap c(B)$

The positions taken by some of the fuzzy sets and logic theorists on these two levels of theoretical inquiry are shown in Table 4 which is based on the Meta-Linguistic interpretations of the Classical axioms which are shown in Table 5.

Table 4. Position Taken by some of Fuzzy Set and Logic Theorists on the Hierarchy of Levels of Theoretical Inquiry

GENERAL EPISTEMOLOGICAL LEVEL	4.Correspondence theory of Validity both objective and subjective.Approximate Reasoning 3.Subjective-objective, experimental and empiricist, e.g., expert and fuzzy data mining based.
ONTOLOGICAL LEVEL	2. $\mathbf{S} \overset{\mathbf{R}}{\longleftrightarrow} \mathbf{O}$ schema gives credence both on the Level of subject and the object interaction 1. Realism - fuzzy and uncertain Generation of "Fuzzy Canonical Forms" that are not equivalent to each other in contrast to "Classical Normal Forms". Generation of new Laws of Conservation for t-norms, co-norms, Belief, Plausibility, Probability, etc.

At this point, we ought to ask :“what are the basic axioms of CWW?”, and introduce “Meta-Linguistic Axioms” as a foundation for CWW which are shown in Table 5 below.

Table 5. Meta-Linguistic Expression of the Axioms for CWW

Involution:	NOT(NOT(A)) vs A
Commutativity:	A AND B vs B AND A
	A OR B vs B OR A
Associativity:	(A AND B) AND C vs A AND (B AND C)
	(A OR B) OR C vs A OR (B OR C)
Distributivity:	A AND (B OR C)vs(A AND B) OR (A AND C)
	A OR (B AND C)vs(A OR B) AND (A OR C)
Idempotency:	A AND A vs A
	A OR A vs A
Absorption :	A OR (A AND B) VS A
	A AND (A OR B) VS A
Absorption by X and ∅:	A OR X vs X
	A AND ∅ vs∅
Identity:	A OR ∅ vs A
	A AND X vs A
Law of contradiction:	∅ vs A AND NOT(A)
Law of excluded middle:	A OR NOT(A) vs X
De Morgan’s Laws:	NOT(A AND B) vs NOT(A) AND NOT(B)
	NOT(A OR B) vs NOT(A) OR NOT(B)

2 Classical Theory

On the bases of Tables 2 and 3, we hava the well known Classical equivalences between the Disjunctive and Conjunctive Normal Form, DNF(.) and CNF(.). For example, on the ontological levels, we hava the Classical equivalences such as:

DNF(A OR c(A)) ≡ CNF(A OR c(A)),
DNF(A AND c(A)) ≡ CNF(A AND c(A)),

and the associated law of conservation:
μ[DNF(A OR c(A)) ≡ CNF(A OR c(A))]
+ μ[DNF(A AND c(A)) ≡ CNF(A AND c(A))] = 1

We have their equivalences such as:
DNF(A AND B) ≡ CNF(A AND B), and
DNF(A OR B) ≡ CNF(A OR B),

Furthermore for T-norms and co-norms, Belief, Plausibility, Probability, etc., we receive the well known formulae:

T(a,b) = 1 – S(n(a), n(b)),
Bel(A) + Pl(c(A)) = 1
Pl(A) + Bel(c(A)) = 1
Pr(A) +Pr(c(A)) = 1

On the Epistemological Level, we find various developments of system models with application technologies known as statistical methods, such as multi-variate regression equations, programming methods such as linear and non-linear optimization algorithms or optimal control schemas developed on objective data that are obtained by measurement devises and depend on description, D, and validation, V, frameworks on the two valued theory, {D{0,1}, V{0,1}}.

As well the validation of the models are assessed with domain-specific test data that are assumed to be standing on descriptive and verified framework of {D{0,1}, V{0,1}}. The validations of the domain specific models are executed with the classical inference schemas such as Modus Ponens. This may entail a re-computation of, say, regression, or programming or control models with test data. Results obtained from such models are assumed to be on {D{0,1}, V{0,1}} framework based on some level of statistical risk.

Furthermore, "validation" of such system models is assumed in terms of certain criteria that are developed by classical statistical methods. For example, some of these are known as

RMSE – Root Mean Square Error
R2 – How successful the fit is in explaining the variation in the data
Accuracy of Prediction
Power of Prediction

These are defined as:

$$RMSE = \sqrt{MSE} \quad MSE = \frac{SSE}{n} \quad R^2 = 1 - \frac{SSE}{SST} \quad SST = \sum_{i=1}^{n} (y_i - \overline{y}_i)^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

MSE: Mean Square Error
SSE: Sum of Square Errors
SST: Total Sum of Squares

Accuracy t (%) = X / P
Power t (%) = X / A

X: Total number of predicted values that are hit correctly in the interval t
P: Total number of predicted values at interval t
A: Total number of actual values at interval t

It should be noted that these are crisp definitions. We need to develop their fuzzy versions for the fuzzy theory.

3 Fuzzy Theory

However, in most theoretical and applied investigations of fuzzy theory, only a part of the Classical axioms are considered as shown in Table 6. Since, they are crisp axioms, we suggest that this is a myopic adaptation.

Recall that in fuzzy theory, briefly, every element belongs to a concept class, say A, to a partial degree, i.e., $\mu A: X \to [0,1]$, $\mu A(x)=a \in [0,1]$, $x \in X$, where $\mu A(x)$ is the membership assignment of an element $x \in X$ to a concept class A in a proposition. Thus most of all concepts are definable to be true to a degree.

"Fuzzy Truth Tables" and in turn the derivation of the combination of concepts for any two fuzzy sets A and B, when they are represented by a Type 1 fuzzy sets, turn to reveal two canonical terms as Fuzzy Disjunctive and Conjunctive Canonical Forms. For example for "AND", "OR" and "IMP" linguistic expressions, we receive the following two forms for each combination of concepts:

$$\text{"A AND B"} = \begin{cases} \text{FDCF(A AND B)} = A \cap B \\ \text{FCCF(A AND B)} = (A \cup B) \cap (c(A) \cup B) \cap (A \cup c(B)), \end{cases} \quad \text{and}$$

$$\text{"A OR B"} = \begin{cases} \text{FDCF(A OR B)} = (A \cap B) \cup (c(A) \cap B) \cup (A \cap c(B)) \\ \text{FCCF(A AND B)} = A \cup B, \end{cases} \quad \text{and}$$

$$\text{"A IMP B"} = \begin{cases} \text{FDCF(A IMP B)} = (A \cap B) \cup (c(A) \cap B) \cup (c(A) \cap c(B)) \\ \text{FCCF(A IMP B)} = c(A) \cup B, \end{cases}$$

etc., in analogy to the two-valued set and logic theory where FDCF(.)=DNF(.) and FCCF(.)=CNF(.) in form only.

Furthermore, as it is shown in Türksen [5,6,7], the equivalence, DNF(.)≡CNF(.), breaks down, i.e., we have FDCF(.)not=FCCF(.) and in particular we get FDCF(.)⊆FCCF(.) for certain classes of t-norms and t-conorms that are strict and nilpotent Archimedean.

For example, particular consequences that we receive are:

(1) FDCF(A OR NOT A)⊆FCCF(A OR NOT A)

which is the realization of the law of "Fuzzy Middle" as opposed to the Law of Excluded Middle and

(2) the Law of "Fuzzy Contradiction, FDCF(A AND NOT A) ⊆FCCF(A AND NOT A)

as opposed to the Law of Crisp Contradiction.

As a consequence of these, we obtain new Laws of Conservation in fuzzy theory as:

$$m[\text{FDCF}(A \text{ AND } c(A))] + m[\text{FCCF}(A \text{ OR } c(A))] = 1;$$

which is well known but re-interpreted for fuzzy sets and

$$m[\text{FDCF}(A \text{ OR } c(A))] + m[\text{FCCF}(A \text{ AND } c(A))] = 1.$$

which now exists for fuzzy sets only!

Hence we once again observe that the "Principle of Invariance" is re-established in Interval-Valued Type 2 fuzzy set theory, but as two distinct Laws of Conservation.

This means that linguistic connectives "AND", "OR", "IMP", etc., are not interpreted in a one-to-one correspondence, i.e., non-isomorphic, to be equal to "$\cap$", "$\cup$", "c(.), $\cup$", etc. That is the imprecise and varying meanings of linguistic connectives are not precisiated in an absolute manner and there is no absolute precisiation of the meaning of words nor is there an absolute precisiation of the meaning of connectives. This provides a framework for the representation of uncertainty in the combination of words and hence in reasoning with them as a foundation for CWW.

The break down of the equivalences in Fuzzy theory, i.e., FDCF(.)not=FCCF(.), in turn generates new additional formulae for t-norm-conorms, Belief, Plausibility and Probability. That is, we now obtain:

Two T-Norm-Conorms Formulae for fuzzy sets in fuzzy theory:

(1) T(a,b) = 1 – S(n(a),n(b))

which is well known but re-interpreted in fuzzy theory; and

(2) T[T(S(a,b), S(n(a),b)), S(a,n(b))]
= 1 – S[S(T(n(a), n(b)), T(a,n(b))], T(n(a),b)]

which is new in fuzzy theory!

Two Belief and Plausibility measures over fuzzy sets at particular α-cuts:

(1) Pl[FDCF(A AND B)] + [Bel[FCCF(c(A) OR c(B)] = 1
Pl[(A$\cap$B) + Bel(c(A) $\cup$ c(B)) = 1

which is well known but re-interpreted for fuzzy theory; and

(2) Pl[FCCF(A AND B)] + Bel[FDCF(c(A) OR c(B))] = 1
Pl[(A$\cup$B) $\cap$ (c(A) $\cup$ B) $\cap$ (A $\cup$ c(B))] + Bel[(c(A) $\cap$ c(B)) $\cup$ (A $\cap$ c(B)) $\cup$ (c(A) $\cap$ B)] = 1

which is new in fuzzy theory!

Two Probability measures over fuzzy sets at particular α-cuts:

Pr(A AND B) + Pr(c(A) OR c(B) = 1
Pr[FDCF(A AND B] + Pr[FCCF(c(A) OR c(B)] = 1

which is well known but re-interpreted for fuzzy theory; and

(3) Pr[(A$\cap$B) + Pr(c(A) $\cup$ c(B)) = 1
Pr[FCCF(A AND B)] + Pr[FDCF(c(A) OR c(B))] = 1

Pr[(A$\cup$B) $\cap$ (c(A) $\cup$ B) $\cap$ (A $\cup$ c(B))]
+ Pr[(c(A) $\cap$ c(B)) $\cup$ (A $\cap$ c(B)) $\cup$ (c(A) $\cap$ B)] = 1

which is new in fuzzy theory!

This particular interpretation of measures at all a-cut levels of fuzzy sets together with knowledge representation and reasoning formulae form a unique foundation for Type 2 fuzzy set theory in general and in particular for Interval-Valued Type 2 fuzzy set theory generated by the combination of linguistic concepts with linguistic connectives even if the initial meaning representation of words are to be reduced to Type 1 membership representation. More general representations start with Type 2 representation schema and then form Type 2 reasoning schemas to capture both imprecision and uncertainty.

Therefore on the Epistemological level, we first have approximate reasoning models expressed in particular as Interval-Valued Type 2 fuzzy set as:

$$\text{A IMP B} = \begin{cases} \text{FDCF(A IMP B)} \\ \text{FCCF(A IMP B)} \end{cases}$$

as a descriptive model, i.e., an Interval-Valued Type 2 rule, a premise. That is {{D[0,1] V{0,1}} IMP {D[0,1] V{0,1}}}={D[0,1] V{0,1}} which is within the framework of a fuzzy inference schema such as Generalized Modus Ponens, GMP, originally proposed by Zadeh as Compositional Rule of Inference, CRI [10]. In this framework, the first premise {D[0,1] V{0,1}} for “A IMP B” combined with a second premise {D[0,1] V{0,1}} for “A′” result in a consequence {D[0,1] V{0,1}} for B*. The validation is based on a fuzzy comparison of the actual output for a given test input data and model output for the same test input data. The error is usually accepted to be a true, V{0,1}, verification but based on a risk statistically but fuzzily evaluated assessment dependent on a fuzzy test of hypothesis. It should be noted that all of the proceeding exposition which is made for the Descriptive fuzzy set paradigm. A similar exposition is applicable to the Veristic fuzzy set paradigm.

It is to be noted that we are yet to develop “Fuzzy RMSE”, “Fuzzy R2”, “Fuzzy Accuracy of Prediction”, and “Fuzzy power of Prediction” in analogy to their classical versions.

It is in these respects that many of the familiar revisions and alternatives to classical thinking, suggested by Black, Lukasiewicz, Kleene, etc., were preliminary break away strategies from the classical paradigm. With the grand paradigm shift caused by Zadeh’s seminal work and continuous stream of visionary proposals, it is now clear that most of them reflect very different stances adopted at the more fundamental levels of our proposed ontology and epistemology. Those changes, it appears, have sometimes been made only in a more tacit and implicit manner. In our studies, it became obvious that the most radical revisions are likely to be the ones that stem from modification to be made at the ontological and epistemological levels where the basic grounding of a theory takes place.

4 Conclusions

In this exposition, we have stated that there are essential properties of the fuzzy theory that comes to light at the ontological and epistemological levels of theoretical inquiry which are quite different from the properties of the classical theory. We have stated these both in linguistic expressions as well as in terms of classical axiomatic assumptions for the classical theory but in terms of Meta-Linguistic interpretations of classical axioms for the grounding of the fuzzy theory in a comparative manner.

In this comparative approach, we have also stated that there are two distinct fuzzy canonical forms in fuzzy theory that correspond to all classical normal forms which form equivalences in classical theory. This causes the generation of "Interval-Valued Type 2 Fuzzy Sets". The law of conservation of information is still upheld within fuzzy theory but in two distinct forms.

References

1. Hodge, R. (2001), "Key Terms in Fuzzy Logic Deep Roots and New Understanding", University of Western Sydney, Australia.
2. Popper, K.(1959) Logic of Scientific Discovery, Hutchison, London
3. Turing, A. (1950) "Computing Machinery and Intelligence" Mınd,Vol. 59, pp. 433-460.
4. Simon, D. 2007) "Fuzzy Logic and Western Culture" (to appear).
5. Türkşen I.B. (2006) An Ontological and Epistemological Perspective of Fuzzy Theory, Elsevier, New York.
6. Türkşen I.B. (1986), "Interval-Valued Fuzzy Sets based on Normal Forms", Fuzzy Sets and Systems, 191-210.
7. Türkşen I.B. (2001), "Computing with Descriptive and Veristic Words: Knowledge Representation and Reasoning", in: Computing With Words, P.P. Wang(ed.), Chapter 10, Wiley, New York, 297-328.
8. Zadeh, L.A. (2001), "From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions", in: P.P. Wang(ed.) Computing With Words, Wiley Series on Intelligent Systems, Wiley and Sons, New York, 35-68.
9. Zadeh, L.A. (1978), "Fuzzy Sets as a Basis for a Theory of Possibility", Fuzzy Sets and Systems, 3-28.
10. Zadeh, L.A. (1973), "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", IEEE Trans. Systems Man Cybernet, 3, 28-44.
11. Zadeh, L.A. (1971), "Similarity Relations and Fuzzy Ordering", Information Sciences, 3, 177-200.
12. Zadeh, L.A. (1968), "Probability Measures of Fuzzy Events", J.Math. Analysis and Appl., 10, 421-427.
13. Zadeh, L.A. (1965), "Fuzzy Sets", Information and Control Systems, Vol.8, 338-353.

Application of Fuzzy Cognitive Maps to Business Planning Models

Vesa A. Niskanen

Dept. of Economics & Management, University of Helsinki, PO Box 27,
00014 Helsinki, Finland
vesa.a.niskanen@helsinki.fi

Abstract. Soft computing modeling of business planning is considered. Our ultimate aim is to provide new proactive and innovative resolutions for designing a soft computing simulation tool for learning business planning. In our Learning Business Plan Project we have established a theory on business planning which assumes that we should replace the traditional business planning with a more creative approach and thus we should focus more on invention and development of business ideas. We have designed a model of five stages according to our theory, and we apply soft computing and cognitive maps for our model simulations. In simulation we particularly apply linguistic cognitive maps which seem more versatile than the corresponding numerical maps. Two modeling examples are also provided.

Keywords: business planning, entrepreneurship, soft computing, cognitive maps.

1 Introduction

This paper considers a novel approach to business planning from the soft computing modeling standpoint, in particular, we focus on the processes in which invention and innovation are involved when new business ideas are created. Our ultimate aim is to provide new proactive and innovative resolutions for designing a soft computing simulation tool for learning business planning. Our economical and behavioral-scientific theories stem from the ideas suggested by [2-4], and according to them, instead of conducting the traditional business planning, we should be more creative and flexible as well as we should focus more on invention and development of business ideas.

Our approach, which is studied in our *Learning Business Plan Project* (LBP Project), models the invention of business ideas by combining theories on creativity in order to provide a new proactive and innovative resolution for learning business planning. Our approach also attempts to combine certain relevant philosophical aspects and results of the behavioral sciences (more details are provided e.g. in [5]).

We thus assume that human intelligence is creative and capable of interacting with reality, and these assumptions lead us to certain methodological and theoretical choices in our model construction. Our models, in turn, are simulated in a computer environment by using soft computing methods. In practice we make both concept map and cognitive map configurations on the business planning problems of the real world, and then we construct the corresponding soft computing models in order to simulate our phenomena [1,6,8] (Fig. 1).

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 119–127, 2007.
springerlink.com

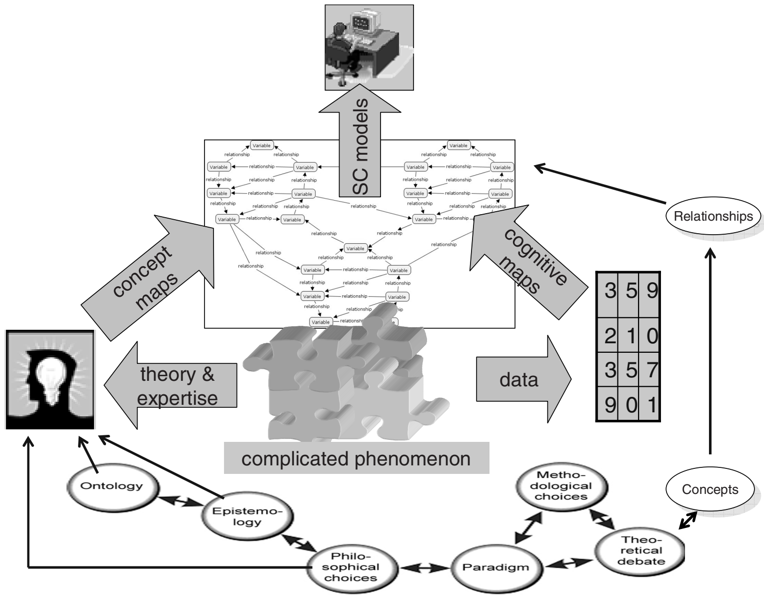


Fig. 1. The Framework for Our LBP Modeling

Below we introduce our approach which combines our theories and computer simulations. We also provide simplified examples which hopefully illuminate how we will proceed in practice. Since more concrete results will be published in the forthcoming papers, we at present mainly expect comments and feedback on our approach from the scientific community.

Section 2 briefly introduces the theoretical basis of our project. In Section 3 we consider our linguistic approach to cognitive map modeling. Section 4 provides two examples, and Section 5 concludes our paper.

2 Theoretical Basis of the LBP Project

In [5] a model including five stages for business planning is constructed (Fig. 2). First, the business idea is created; second, we develop our idea; third, we evaluate the idea by performing certain transformations, financial calculations and feasibility analysis; fourth, we have the implementation stage for our idea, and finally, we apply various re-evaluation and follow-up procedures. At present we focus on stages one to three.

Hence, we have designed a model which adds action and creativity as essential constituents to business planning whereas the previous models have focused on stage three. These five stages constitute a complicated network the nodes and interrelationships of which we consider in a computer environment.

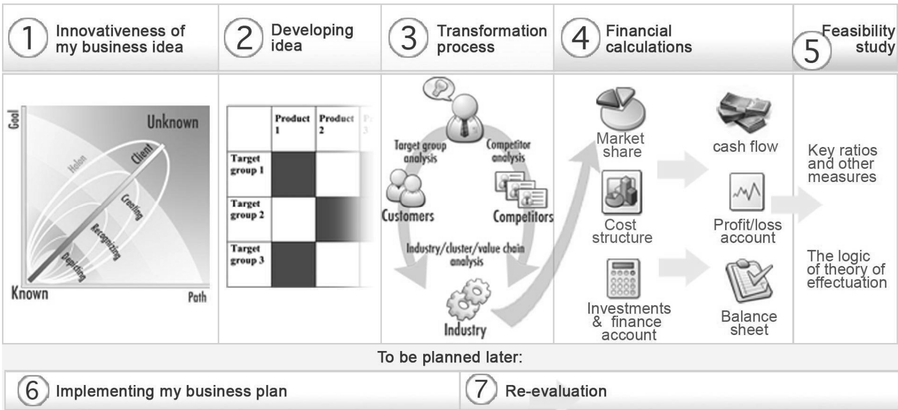


Fig. 2. Five Principal Stages in Business Planning ((3)-(5) comprise Stage three)

However, we still lack sufficient knowledge on the dynamics of the five stages above, and the available knowledge seems contradictory and confusing. At present we focus on the interrelationships between the stages one to three, and hence this problem area is only studied in this paper.

Another problem is that all business ideas are not novel, but rather they can be imitations of the existing ideas. One reason for this is that the concepts of invention, creativity and imitation are quite confusing in entrepreneurship research and their analyses has lead to various debates. However, there seems to be a general acceptance for the assumption that creativity and ability to innovations are connected to opportunity recognition in entrepreneurship research. For example, one definition is that entrepreneurial creativity means the entrepreneur's recognition of opportunities for development and the exploitation of resources. Hence the concepts of creativity and recognition of opportunity seem to have almost similar meanings.

To overcome this confusion we have applied Eijnatten's ideas on novelty [3]. He draws a distinction between individual and collective entrepreneurial novelty with improvement. Eijnatten suggests that, first, if we know both the goal and path we can only make improvements, second, if either of them is unknown, we renew our practices, and finally, if both of these aspects are unknown, we can reach the real novelty.

Eijnatten's approach enables us to understand how to recognize, create and mimic or depict opportunities. We can thus assume that if we know both the path and the goal, we can depict our business idea. If we recognize something, it is either the goal or path. If we create something, it refers to entities which are still non-existent to us, and we encounter unknown goals and paths. We thus have three basic alternatives and their different variations, viz. depicting, recognizing and creating opportunities to be exploited, as well as various degrees of known or unknown (Fig. 3).

By connecting the foregoing assumption to our model of learning business planning means that the more creative is our business idea, the more unknown it is to us.

Since the role of a client is also crucial to business, we can add this dimension to our model. Thus the degree of knowing leads to three zones of innovativeness, viz. depiction, recognition and creation in a three-dimensional space which constitutes the goal, path and client. In each zone these three dimensions have to be in balance, and

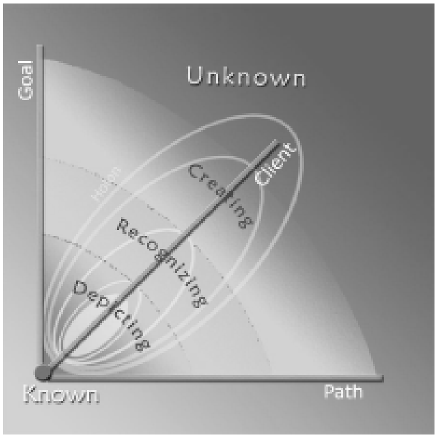


Fig. 3. Dynamics of Innovativeness in Business Planning

thus the creativity of the business idea is dependent upon person's ability to integrate these dimensions. If we provide answers to this process in business planning, we can model the dynamics between the five foregoing stages. In Section 4 we provide a simulation model of this idea.

3 Modeling with Soft Computing

The foregoing theoretical frameworks also presuppose simulation models which are good in practice. To date it has been problematic to model these phenomena in a computer environment because the conventional quantitative models have been fairly complicated whereas qualitative computer modeling as such arouses difficulties. By virtue of soft computing, on the other hand, we can construct both usable quantitative and qualitative models which also correspond well with human reasoning [6-7,10-11].

At general level, we apply the modeling depicted in Figure 1. Hence, we first design configurations by using concept and cognitive maps and these maps include our variables and their interrelationships [1,8]. Then we construct computer models which base on the idea of cognitive map modeling.

If (empirical) data on the behavior of a cognitive map is unavailable, we only operate with *a priori* cognitive maps, and thus our constructions are only based on our theories and expertise, otherwise we can also construct *a posteriori* maps, and then we can apply such methods as statistics (e.g., regression and path analysis), fuzzy systems, neural networks, Bayesian networks or evolutionary computing [9]. According to [9], most cognitive maps are still a priori maps and thus manual work seems to play an essential role in this context. However, more automatic or semi-automatic procedures are expected in this area.

To date numeric cognitive maps have been usual and they provide us a fairly good modeling basis as well as they are quite simple systems from the mathematical standpoint. They also allow us to use feedback or loops in our models, and this feature is problematic at least in the Bayesian networks.

However, the numeric cognitive maps also arouse some well-known problems. First, most of them have been a priori maps. Second, they can only establish monotonic causal interrelationships between the variables. Third, only numerical values and interrelationships can be used, and thus they are less user-friendly than the linguistic maps. Fourth, time delays are problematic because some phenomena can take place in a short term whereas others can occur in a long term [6].

The author has suggested that we use appropriate linguistic cognitive maps instead in order to resolve most of the foregoing problems [6]. In linguistic maps we use fuzzy linguistic variables and we establish the interrelationships between these variables by using fuzzy linguistic rule sets [6-7,10-11]. Hence, this approach, allows us to employ both quantitative and qualitative variables, use more versatile variable values and interrelationships, apply non-linear and non-monotonic modeling and construct more user-friendly systems. In addition, we can construct a posteriori linguistic cognitive maps in an automatic or a semi-automatic manner if we apply statistics and neuro-fuzzy systems [6].

4 Simple Modeling Examples

Consider first Eijnatten's model above (Fig. 3). In this context our modeling can base on Figure 4 and we thus use three input variables, Path, Goal and Client. These variables represent the degrees of knowing, and we can assign such values to them as *known*, *fairly known*, *medium*, *fairly unknown* and *unknown*. The output variable represents the degree of creativity, and we assign to it the values *depicting*, *recognizing* and *creating*. The relationships, which are established with fuzzy rules, are analogous with positive correlation (either linear or non-linear), i.e., the more unknown the input situation, the more creative is our business idea and vice versa (if we use the scales known - unknown and depicting - recognizing - creative).

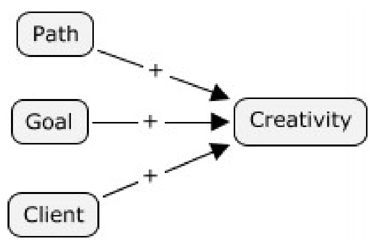


Fig. 4. The Modified Eijnatten's Fuzzy Model

Hence, we operate with such fuzzy rule sets as

1. If path is unknown and goal is unknown and client is unknown, then our business idea is creative.
2. If path is medium and goal is medium and client is medium, then our business idea is recognizing.
3. If path is known and goal is known and client is known, then our business idea is depicting.

Our inference engine operates with the foregoing rules and appropriate reasoning algorithms, for example, with the Mamdani or Takagi-Sugeno algorithms.

A more concrete example is the zero-order rule set in Figure 5 which was generated by using grid partition method in Matlab's Fuzzy Logic Toolbox. In this case the input values of 1 mean fully unknown situation and correspondingly the output value of 1 means fully creative business idea. In the real world, however, our variables constitute networks of variables, and this standpoint is taken into account in our LBP modeling.

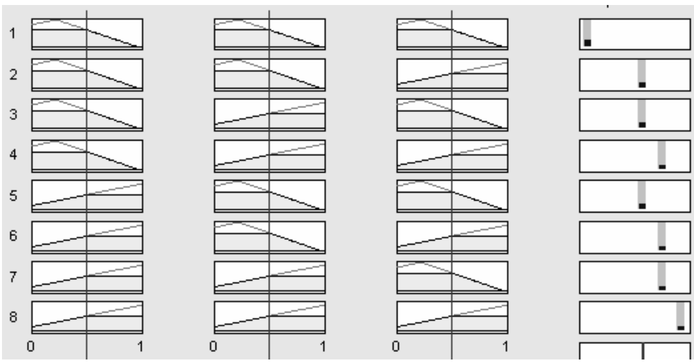


Fig. 5. Eight Fuzzy Rules for Eijnatten's Model. Inputs (which can be in any order): 0 = known, 1 = unknown. Output: 0 = depicting, 1 = creative.

Our second example considers our application of Eijnatten's chaotic growth of an high-tech enteprise [3]. He maintains that the founding of a high-tech start-up may be seen as an entrepreneurial process of successfully escaping the known valley ("business as usual"), climbing the mountain ("exploring the radical innovation"), and gliding into a new unknown valley ("the new business process"). For example, the product of service prototype may change or the business focus may change. Hence, coming from a relative stable state the enterprise is entering a relative unstable one. It will experience all kinds of dilemmas and contradictions, and literally "feels" the turbulence.

One version of Eijnatten's idea is a cognitive map which models the variation in sales of a given enterprise when the need for innovations varies more or less randomly or in a "chaotic" manner in a given period of time. We can thus assume that whenever our sales is increasing sufficiently, our enterprise meets new challenges (e.g. more competition in the markets) and we have to contribute more to innovations, otherwise our sales can decrease.

If we apply the simplified cognitive map in Fig. 6, our sales is dependent upon the variation in the needs for innovations (which is determined by various factors in practice) and our contributions to meet and fill these needs. If we focus on sales, we can generate such rules as

1. If the previous sales was fairly small and our contribution to innovations is slightly less than the true needs for innovations, then our sales is small.
2. If the previous sales was fairly large and our contribution to innovations is slightly more than the true needs for innovations, then our sales is large.

Fig. 7 depicts a corresponding tentative zero-order fuzzy rule set for the output variable Sales which was generated by using the grid technique.

If we construct fuzzy reasoning systems with linguistic rule bases for each variable as well as a meta-level system which applies these systems in simulation, we obtain a cognitive map system analogous to the prevailing numerical cognitive maps but in our case we can use more versatile interrelationships between the variables. The historical curves in Fig. 8 are based on this modeling approach and they show a simulation in which our sales is at first fairly small and our contribution to innovations is slightly less than the true need for innovations. In addition, our contribution is kept at a constant level. We notice that in this case our policy leads us to a zero-level sales.

The foregoing simplified examples hopefully illustrate our simulation approach with our novel linguistic modeling method. At present, we are constructing cognitive maps for the stages one to three, and these results will be presented in the further papers.

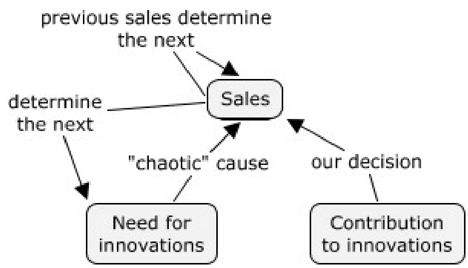


Fig. 6. An Example of Eijnatten's Chaotic Model

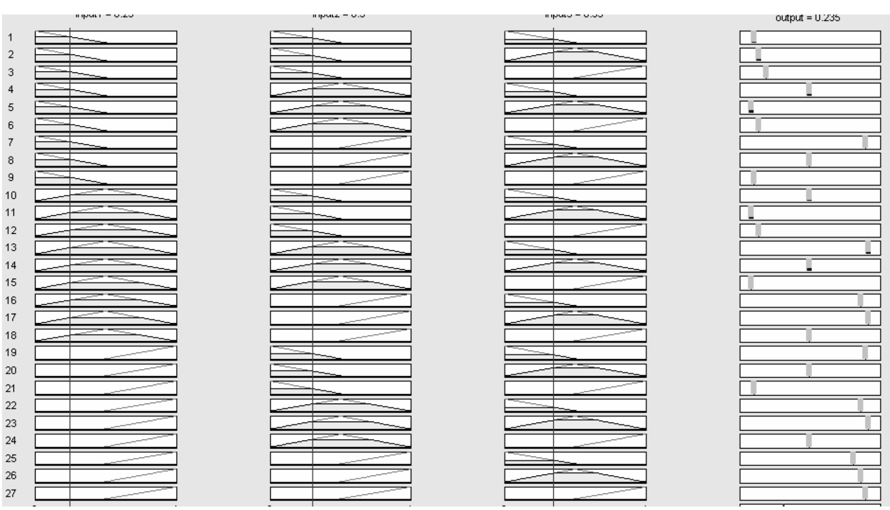


Fig. 7. Tentative Fuzzy Rules for Sales (Inputs: previous sales, contribution to innovations, need for innovations. Output: sales)

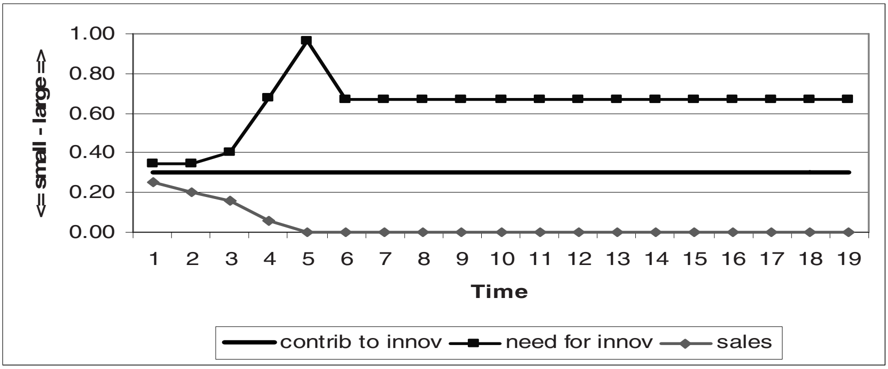


Fig. 8. Historical Curves in the Chaotic Model

5 Conclusions

Soft computing aspects of business plan modeling was considered. Our aim is to provide new proactive and innovative resolutions for designing a soft computing simulation tool for learning business planning. We have suggested two main ideas for this problem area. First, in our LBP project we have combined creativity to business planning and thus succeeded in establishing more usable logical and theoretical grounds for further modeling.

Second, in a computer environment we have applied soft computing and concept and cognitive maps to our simulations and tool construction. In this context we are applying novel linguistic maps and thus we can construct better models and tools. Two modeling examples was also sketched.

We still have unresolved problems concerning the nature of human behavior in business planning as well as in the corresponging simulation designs, and these issues will be considered in the future studies of the Project.

Acknowledgements

We would like to express our gratitude to the Finnish Ministry of Trade and Industry as well as to the European Union Social Fund for financing our Project. I also express my thanks to the referees.

References

1. Axelrod, R.: Structure of Decision, The Cognitive Maps of Political Elites, Princeton UniversityPress Princeton (1976).
2. Carrier, C.: Pedagogical challenges in entrepreneurship education, in P. Kyrö and C. Carrier (2005): The dynamics of learning entrepreneurship in a crosscultural university context, Entrepreneurship Education Series 2/2005, Hämeenlinna: University of Tampere, Research Centre for Vocational and Professional Education.

3. Eijnatten, Frans M. van: A chaordic lens for understanding entrepreneurship and intrapreneurship, In Fayolle, A.. Kyrö, P. & Uljin, J. (Eds.) (2005): Entrepreneurship Research in Europe: Perspectives and Outcomes, Cheltenham.
4. Hindle, Kevin: An Enhanced Paradigm of Entrepreneurial Business Planning. Swinburne University of Technology, Swinburne (1997).
5. Kyrö P. and Niskanen V.A.: Modelling innovativeness and creativity in business planning, Proceedings of the KES '06 conference, (2006).
6. Niskanen, V. A., Fuzzy Linguistic Cognitive-Map Modeling in Human Sciences, submitted to consideration to Fuzzy Sets and Systems, (2006).
7. Niskanen V. A., Soft Computing Methods in Human Sciences. Springer Verlag Heidelberg (2003).
8. Novak J.: Learning, Creating and Using Knowledge, Concept Maps as Faciliative Tools in Schools and Corporations. Lawrence Erlbaum Associates Mahwah New Jersey (1998).
9. W. Stach, and al.: Genetic Learning of Fuzzy Cognitive Maps, Fuzzy Sets and Systems 153 (3) (2005) 371-401.
10. Zadeh, L.: Fuzzy Logic = Computing with Words, IEEE Transactions on Fuzzy Systems, vol. 2, 103-111 (1996).
11. Zadeh, L.: Toward a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic, Fuzzy Sets and Systems vol. 90, no. 2, 111-127 (1997).

Perceptions for Making Sense, Description Language for Meaning Articulation

Tero Joronen

University of California at Berkeley, Berkeley Initiative in Soft Computing (BISC),
199 MF Cory Hall, Berkeley, CA, 94720, USA
tero.joronen@metso.com

Abstract. The aim of this article is to introduce a novel way to illustrate human-like reasoning by using a pictorial language and fuzzy sets. We follow the newest developments on soft computing, but approach the perceptions from a different point of view. We introduce a pictorial language to define human-like multi-domain reasoning. We start to further develop an existing simple pictorial language Bliss and increase its expressive power by fuzzy sets and by presenting several sentences in a graph. These extra sentences are used to show the right use (meaning) of a concept in the reasoning. Several examples of the approach are provided.

Keywords: pictorial language, meaning, fuzzy sets, computational theory.

1 Introduction

The aim of this article is to introduce a novel way of illustrating human-like reasoning by using a pictorial language and fuzzy sets. The language is called Description Language for Meaning Articulation (DLMA). The newest development in the field of soft computing is Prof. Zadeh's computational theory of perceptions [21] and Percisiated Natural Language [23]. The development continues towards the Generalized Theory of Uncertainty [24]. We follow the approach of the perceptional computation and continue our studies on semantics of fuzzy sets [8] but we approach the problem from a different point of view. A pictorial language is introduced to define human-like multi-domain reasoning. An existing simple pictorial language *Bliss* [2] is further developed, and fuzzy sets and human-like reasoning in graph form increase its expressive power. The graph contains several Bliss sentences that are used to show the right way of reasoning. We call these extra sentences *moves*.

The paper is divided into four parts. The first part explores the motivation of our approach. The second part introduces the DLMA. Part two is further divided into five subsections: 1) background, 2) structure 3) Bliss 4) fuzzy sets in DLMA, and 5) moves. The third part is dedicated to some illustrative examples of DLMA. The last part includes the conclusion and future work.

2 Motivation

In this part we argue that there is a need to find new ways to define and solve scientific and everyday problems. Our pictorial language is a good way to illustrate human-like

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 128–138, 2007.
springerlink.com

reasoning. Later in this paper, we will show that by including pictorial elements to reasoning we will solve some problematic definitions.

Humans are unique in applying languages such as natural language or mathematics. Yet, there is still a lot to accomplish in the mathematics of natural language. Zadeh [23, p. 75] argues: ‘It is a deep-seated tradition in science to view the use of natural languages in scientific theories as a manifestation of mathematical immaturity. We may argue also that the mathematics is immature not to be capable to use natural language.’ Zadeh continues [23, p. 76] ‘Due to their imprecision, perceptions do not lend themselves to meaning representation and inference through the use of methods based on bivalent logic.’ Several scientific theories are based on bivalent logic. For instance, the classical information theory stands on the mathematical theory of communication [14]. For Shannon communication is a probabilistic phenomenon. Shannon emphasizes that the semantics are irrelevant to the engineering problems of the information transfer. Shannon's narrow engineering approach is still prevailing in many of the technical and scientific approaches of information theory today, and indeed, there is a lack of ability to handle the meaning of natural language.

The foundations of the probability theory are confused. There is an ongoing debate between the Bayesian and the purely frequential interpretations of the probability. Jaynes [7, p. 21-22], a well-known Bayesian, notes that the English verb *is* has two entirely different meanings. He gives examples: ‘The room is noisy’ and ‘There is noise in the room’. The latter is ontological, the physical existence and the former is epistemological, expressing someone’s personal perception. He calls this problem of recognizing this ‘mind projection fallacy’. According to Zadeh [22] ‘In mathematics, though not logic, definability of mathematical concepts is taken for granted. But as we move further into the age of machine intelligence and automated reasoning, the issue of definability is certain to grow in importance and visibility, raising basic questions that are not easy to resolve.’ This is also true in the definition of the fuzzy membership functions. Our approach tries to figure out the right use (meaning) of the fuzzy membership function. This is explored later in this paper.

A rising approach in cognitive science is the paradigm of two minds [5]. In the paradigm two distinct systems underlie in the reasoning. System 1 is old in evolutionary terms and shared with other animals: it consists of autonomous domain-specific subsystems. System 2 is more recent and distinctively human: it permits abstract reasoning and is relatively slow and limited by intellectual capacity. Both the logic of the argument (system 2) and the believability of it (system 1) substantially influence the reasoning. Evans [5, p. 455-456] reports scientific evidence that confirms this. In an experiment, the portion of wrong answers for an invalid but unbelievable argument was 10 times smaller compared to an invalid but believable argument, although the tested persons were college students. In another logical test, the degree of success for a realistic and linguistic problem was about 5 times better compared to an abstract and symbolic problem.

Based on the arguments of two minds, we argue that by realistic pictorial elements the reasoning can be made easier. A multi-domain definition dramatically clears the way out from the problem of the ‘mind projection fallacy’, as the pictorial language clearly depicts the difference between the existence and the evaluation and is easy to understand.

Scientific evidence proves that a combined representation of pictures and language sentences is easier to handle compared to plain written text. Firstly, information presented verbally and pictorially is easier to remember. Already in 1973, Paivio and Csapo [12] argued, based on careful tests, that when information is processed verbally and pictorially, recall is better than if information is processed only verbally or pictorially. Secondly, pictorial languages are easier to learn. Muter and Johns [10] have shown by scientific experiments that extracting the meaning from logographs (Blissymbols or Chinese characters) was substantially and reliably better compared to English words written in an unfamiliar alphabetic code. They indicate that under a reasonably wide range of conditions, logographic writing systems may be substantially easier to learn to read than alphabetic systems. Thirdly, according to Najjar [11, p. 9], in many cases the superiority effect of pictures in multimedia boosts learning.

3 Description Language of Meaning Articulation

The description language is utilized in a larger domain called the Computational Theory of Meaning Articulation (CTMA). The CTMA is under development and is used in the design of artificial intelligence systems. The first task in the CTMA is to make the meaning of the domain information as clear as possible. This is made by Description Language of Meaning Articulation (DLMA). In it, a pictorial language illustrates the contextual information of the case. Pictorial symbols are used to show the right use of information, a way of proper inference. In the next part we introduce our approach, but first the philosophical background is explored.

3.1 Background

Before we can describe meaning articulation, we have to define what we mean by meaning and articulation. The meaning has normally been approached by some exterior explanation that gives to the proposition its meaning. Wittgenstein [17 p. 4] challenged the traditional doctrines and posted 'if we had to name anything which is the life of the sign, we should have to say that it was its use'. We follow Wittgenstein and consider the meaning as the use. Recently, Canny [3] defined; 'The meaning of an action in a context is the *anticipated consequences* of that action'. We partly agree with this practical formulation of the consequences of the use, but point out that the meaning can be expressed only in the ways of using the concept in a language i.e. in its use (that itself is an activity). Wittgenstein [18] employs language games to show the right use of language as a part of an activity. Wittgenstein, citing the mathematician Gottlob Frege, provides an example of a language game, interestingly, by a fuzzy definition:

> Frege compares a concept to an area and says that an area with vague boundaries cannot be called an area at all. This presumably means that we cannot do anything with it. –But is it senseless to say: 'Stand roughly there'? Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand –as if I were indicating a particular *spot*. And this is just how one might explain to someone what a game is. [18, p. 34]

This is what we are trying to depict by DLMA, showing, by a pictorial language, how the concepts are linked into the reality and how it can be usefully i.e. meaningfully applied in a context. Further, we hope to show the right uses of fuzzy membership functions.

At first sight, language games may look simple but actually they are quite hard to formulate. Trials may end up in an endless battle with the language. We understand that just applying pictorial symbols together with the sentence does not make the sentence alive. As Wittgenstein [17, p. 5] points out:

> If the meaning of the sign is an image built up in our minds when we see or hear the sign, then first let us adopt the method we just described of replacing this mental image by some outward object seen, e.g. a painted of modelled image. Then why should the written sign plus this painted image be alive if the written sign alone was dead? - In fact, as soon as you think of replacing the mental image by, say, a painted one, and as soon as the image thereby loses its occult character, it ceases to seem to impart any life to the sentence at all. … The sign (the sentence) gets its significance from the system of signs, from the language to which it belongs.

This clarifies that the meaning of the sentence (or word) can not be expressed by another sentence but it is possible by using a group of sentences that belongs to the language. Also Vygotsky [19, p. 212] makes an important and relevant point:

> A word without meaning is an empty sound; meaning, therefore, is a criterion of 'word', its dispensable component. … the meaning of every word is a generalization or a concept. And since generalizations and concepts are undeniably acts of thought, we may regard meaning as a phenomenon of thinking. … It is a phenomenon of verbal thought, or meaningful speech-a union of word and thought.

The meaning of a word is not a static expression but an act of thought. Obviously, a single sentence can not have a meaning. Thus we employ several sentences that articulate the meaning, i.e. the use. That gives the meaning through a connection in a thought. The pictorial elements are used to clarify these moves showing the ways of thinking. This we call articulation. The moves are further explained later in this paper.

3.2 Structure

The key idea in DLMA is to represent the contextual information of the case in three different domains:

1. Pictorial (Perception of physical reality),
2. Linguistic (English) and
3. Moves (Graph).

All these elements are shown in a graph (Fig. 6). We utilize Bliss on a pictorial layer. It was selected because:

Bliss is related to the physical world and thus easier to understand
Bliss can show better the context of the sentence as it is correlated to the real world
Bliss uses simple pictorial elements
Bliss differentiates between the thing, the action and an evaluation
Bliss is easy to learn
Bliss is international, and it can be understood independently from the spoken language
Bliss is a scientifically tested language

We increase the expression power of Bliss by fuzzy membership functions to define vague concepts like *near*, *slow* or *approximately*. On a graph, the sentences are numbered because of indexing. Generally, it is good to form the problem as a hypothetical question that is answered. It makes it also easier to apply a threshold to the accuracy of the answer.

3.3 Bliss

The idea of a universal pictorial language is not new, for example, mathematician Leibniz was fascinated by the sign writing of Chinese. He did not like its complexity, but dreamed of something simpler including a simple logic that shows that 1 + 2 = 4 contains a lie. Leibniz never worked out his system, but in the 1940’s Charles Bliss created a symbolic writing system *Bliss* [2]. He learned Chinese and realized that all Chinese, either from south or north can read the same newspapers, even though they speak different languages. Furthermore, pictorial symbols remain stable and 2500-years-old Chinese writing is easily comprehensible by modern Chinese. Bliss is like a simplified Chinese, a pictorial representation of real things as we see them.

Bliss is a symbolic, graphical language that is currently composed of over 3,000 symbols [2]. These symbols can be combined and recombined in endless ways to create new symbols. Bliss-words can be sequenced to form many types of sentences, and express almost everything that is capable in a spoken natural language. Mr. Charles Bliss [2] gives many examples of Bliss used in complicated everyday and professional situations. Bliss uses simple shapes to keep it easy and fast to draw. Sonesson [15] notes that Bliss depicts relatively few perceptual properties of the objects in question. Today Bliss is mainly used for Augmentative and Alternative Communication [13] and has gained a success on the field [15].

Mr. Bliss, as a chemist, tried to form a language that works logically like the chemical symbols [2, p. 92]. Famous philosopher, Bertrand Russell thought very highly of the logical analysis of Bliss and considered the symbols smart and easy to understand [2 p. 35]. Bliss-symbols articulate the meaning of the word as a supplementary writing. Bliss argued [2, p. 66] that graphical symbols are used in extreme situations (Fig.1).

According to Crockford [4, p. 4-5], symbols of Bliss take good advantage of the symbol area. For instance, the symbols of sky and earth are simple horizontal lines on the top and on the bottom of the area; and the symbol of world consists of both the sky and the earth (Fig. 2). Additionally, simple figures of the human body are utilized. For instance, the mind symbol comes from the shape of the top of our head and the idea symbol combines the mind and an arrow down. The Bliss-symbol knowledge combines mind and a container (open box) to make the mind’s storehouse (Fig. 4).

In Bliss, indicators above the symbol (see Fig. 4) differentiate things, actions and evaluations. Examples in Fig. 5 illustrate these indicators; the first symbol means a weight of 100; the second symbol means a physical weigher; the third one the act of weighing; and the fourth one means an evaluation of weight being 100.

Fig. 6 presents a basic example of a Bliss sentence: ‘Bob lives near Berkeley’. Where ⅄ is a man, ⦶ is a life, | | is near, × is many and ××⌂ means a town.

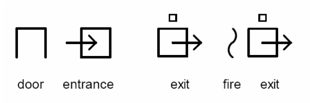


Fig. 1. Bliss Symbols used for the emergency purposes

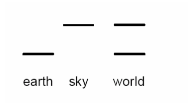


Fig. 2. The usage of the sign area

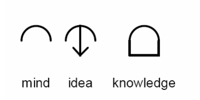


Fig. 3. Examples of Bliss symbols

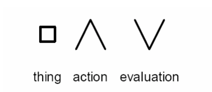


Fig. 4. The classification signs for some words

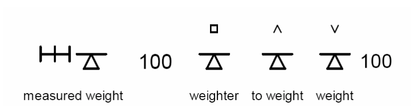


Fig. 5. Examples of indicators

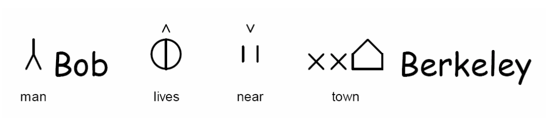


Fig. 6. 'Bob lives near Berkeley' in Bliss

3.4 Fuzzy Sets in DLMA

Bliss does not have any way to express vague concepts like *near*, *slow* or *approximately*. Thus we introduce fuzzy membership functions to Bliss. In DLMA, fuzzy membership functions are drawn on a Bliss symbol area. This function depicts a membership function that gradually changes the degree of membership to a concept on a measurement scale. In a context, the values (0,1) of membership function are drawn on the vertical axis over the relevant totality of a measurement scale on the earth line (Fig. 7). The degree of membership is useful, for instance, in human-like fuzzy reasoning [20] and in Fuzzy Arithmetic [9].

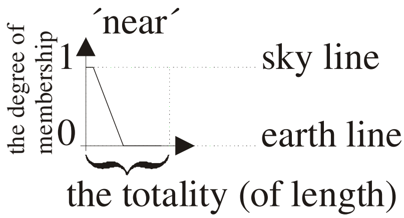


Fig. 7. An illustration of a membership function 'near' in DLMA

We well understand and are actually motivated because the definition of a fuzzy membership function is context dependent. In DLMA, the membership functions are nailed down to the physical reality depicted by the graph. DLMA is capable of defining the proper use of membership functions i.e. give the meaning to them. The membership function might not be exactly defined, for instance, we do not know the right level of membership on each point of the measurement scale, and due to this several membership functions can be drawn. All these possible (or probable) functions have to be taken into account while applying it in a graph. In a forthcoming paper, we try to conquer this key problem by the computational part of Computational Theory of Meaning Articulation (CTMA).

3.5 Moves

As the meaning of a sentence is its use, we have to be able to somehow show the use in order to define the meaning. Moves are used to show the use of the sentence. Moves emerge from the rules of reality (and thinking). The possible moves are (at least) as follows:

1. Ignore or Make up (remove/fill up) information/evaluation
2. Dump or Exaggerate (scale) an evaluation
3. Add information,
   a) Include extra sentence
   b) Logic *IF* n_1 and n_2 then n_3
   c) Fuzzy rules: *IF* μ_1 and μ_2 then μ_3
   d) Function $n_3 = f(n_1, n_2)$
4. Act, change the situation

The first alternative is that we ignore or make up some knowledge. We could, for instance, think that the whole issue is not important. The second alternative is almost similar but now we reform an evaluation ('near' in our example). We could, for instance, dump or exaggerate the nearness to get a satisfactory solution. The third alternative is adding some relevant information. We might know that Bob is a fast walker. In general, we can apply any function that is reasonable in the specific context. We might have to change the case (act) to fulfill the task. In DLMA, an act changes the case and starts an additional graph. The examples in the next part illustrate the moves further.

4 Examples

In this part, we introduce several examples to illustrate the idea of DLMA. The first two examples handle the liar's and the Sorite paradoxes. The third example illustrates the utilization of sentence: 'Bob lives near Berkeley' in answering 'How long does it take for him to reach Berkeley?'

4.1 Liar Paradox

According to Weisstein [16] a sharp version of the liar's paradox is 'This statement is false'. Fig. 8 shows an analysis of it in DLMA. The first sentence is the paradox, the second sentence adds information: a truth is an answer to a question: 'Is this statement true?' The third sentence includes the fact that the question and the answer cannot be the same. Both added facts are straightforward from the appearance of the Blissymbols *a truth* to the difference between *a question* and *an answer*. The fourth sentence argues that the paradox is both; the question and the answer, thus there is a conflict in the paradox and it is unquestionably nonsense i.e. a misusage of language.

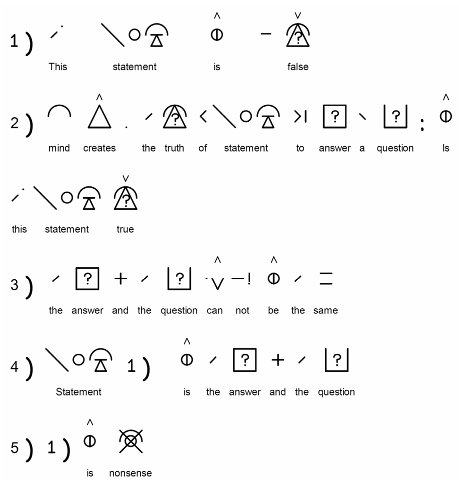


Fig. 8. An analysis of Liar's paradox: 'This statement is false'

4.2 Sorites Paradox

The Sorites (or heap) paradox is that if one piece of grain does not make a heap and the second one does not make a change, then no amount of grain makes a heap [6]. Fig. 9 shows the heap paradox in DLMA. The analysis (Fig. 10) starts from the fact that there is a relation between the amount of grain and the grade of being a heap. The relation can be figured out by a membership function. Thus, the adding of grain finally makes a heap. This apparently turns the paradox to confusing nonsense.

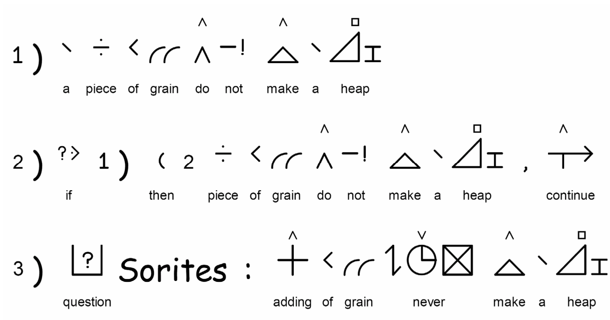


Fig. 9. The Sorites paradox rephrased by DLMA

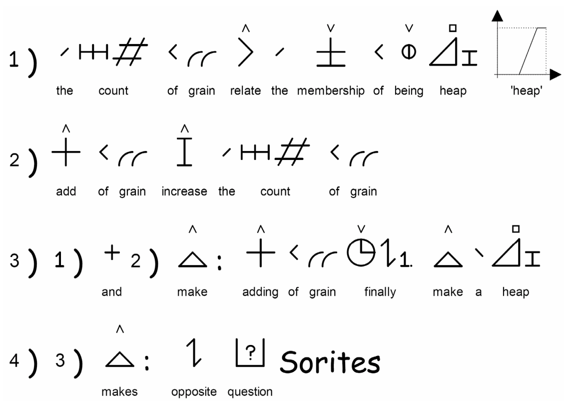


Fig. 10. Analysis for the Sorites Paradox

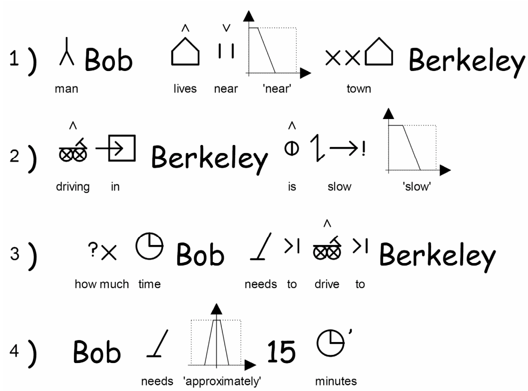


Fig. 11. Example of DLMA with fuzzy membership functions

In these analyses we use additional information to show the true nature of the paradox. As discussed earlier, this is true in all use of language; the information offered by the bare paradox (or sentence) is not enough. The human mind does all this, intuitively and subconsciously. We have to make an effort to illustrate this.

4.3 Bob Lives Near Berkeley

The third example depicts the idea of the computation in the DLMA analysis. We ask: 'How long does it take for Bob to drive to Berkeley?' if 'Bob lives near Berkeley'. The analysis is shown in Fig. 11. We include the fact that driving in Berkeley is *slow*. CTMA utilizes a fuzzy arithmetic that will be presented in forthcoming publications.

5 Conclusion and Future Work

The combination of a pictorial language Bliss, English and Fuzzy Sets offers a platform that articulates the meaning. The meanings are depicted by symbols and several kinds of moves can be used to show the use of the knowledge.

The DLMA handles vague and uncertain information. Applications of DLMA are the illustration of different processes and the design of artificial intelligence applications, like a human computer interface or process control. The next steps, the computational theory and applications, are presented in forthcoming papers.

Acknowledgments. Academy of Finland, Grant No. 112903, BISC at UCB, TEKES, and ACI at Tampere University of Technology financed this study. Thanks to Tuomo Riihimäki for the license of Symbol for Windows® and Antti Hautamäki for valuable comments.

References

1. BCI, 2004. *The fundamental rules of Blissymbolics*. Blissymbolics Communication International (BCI), 2004-09-28. Available at http://ww.blissymbolics.org/
2. Bliss, C. K., 1965. Semantography, 2nd Ed., Semantography Publ., Sydney, Australia.
3. Canny, J., 2006. A Pragmatic Approach to Context and Meaning. To appear in the *Proc. of 12th Int. Conf. on Human-Computer Interaction* 22-27 July 2007, Beijing, P.R. of China
4. Crockford, D., 2003. *Lesson One: An introduction to Blissymbols*. Blissym Language Institute. Available at http://www.blissym.com.
5. Evans, J. St. B. T., 2003. In two minds. *Trends in Cognitive Sciences*. (7)10, 454-459
6. Hyde, D., 2005. Sorites Paradox. *Stanford Encyclopedia of Philosophy*. Available at http://plato.stanford.edu/contents.html
7. Jaynes, E. T., 2003. Probability theory – The logic of Science. Cambridge Univ. Press.
8. Joronen, T., 2003. A Philosophical Study on Fuzzy Sets and Fuzzy Applications. Proc. of 2003 IEEE 12th Int. Fuzzy Systems Conf., St. Louis, MO, USA, 25th – 28th May 2003
9. Kauffman A., and Gupta M.M., 1985. *Introduction to Fuzzy Arithmetic*. Van Nostrand Reinhold, NY.
10. Muter, P., and Johns E. E., 1985. Learning Logographies and Alphabetics Codes. *Human Learning*, Vol. 4, 105-125

11. Najjar, L. J., 1997. A framework for learning from media: The effects of materials, tasks, and tests on performance (GIT-GVU-97-21). Atlanta, GA: Georgia Inst. of Tech., Graphics, Visualization and Usability Center. Available on: http://www.cc.gatech.edu/gvu/reports
12. Paivio, A., and Csapo, K., 1973. Picture superiority in free recall: Imagery or dual coding? Cognitive Psychology, 5, 176-206
13. McNaughton, S., 1985. *Communicating with Blissymbolics*. BCI, Ontario, Canada.
14. Shannon, C. E. A, 1948. Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
15. Sonesson, G., 1996. Blissymbolics. In: *The Encyclopedia of Semiotics*. Ed. Bouissac, Paul. Oxford University Press, New York. 86-88
16. Weisstein, E. W., 2006. Epimenides Paradox. From MathWorld-A Wolfram Web Resource. http://mathworld.wolfram.com/EpimenidesParadox.html
17. Wittgenstein, L., 1965. *The Blue and Brown Books*. New York: Harper Torchbooks.
18. Wittgenstein L., 1968. *Philosophical Investigations*. Basil Blackwell, Oxford, England.
19. Vygotsky, L. S., 1986. *Thought and Language*. MIT Press, Cambridge, Massachusetts.
20. Zadeh, L. A., 1975. Fuzzy Logic and Approximate Reasoning, *Synthese*, 30(1975), 407-428
21. Zadeh, L. A., 2002. Towards a Perception-based Theory of Probabilistic Reasoning with Imprecise Probabilities, *Journal of Statistical Planning and Inference*, 105, 2002, 233-264
22. Zadeh, L. A., 2003. Fuzzy Logic as a Basis for a Theory of Hierarchical Definability, *Proc. of the 33rd Int. Symp. on Multiple-Valued Logic*, 16-19 May 2003, Tokyo, Japan, 3- 4
23. Zadeh, L. A., 2004. Precisiated Natural Language (PNL), *AI Magazine*, 25(3)2004, 74-91
24. Zadeh, L. A., 2006. Generalized Theory of Uncertainty--Principal Concepts and Ideas, Computational Statistics & Data Analysis 51, 15-46

Part IV

Search Engine and Information Processing and Retrieval

FCBIR: A Fuzzy Matching Technique for Content-Based Image Retrieval

Vincent. S. Tseng, Ja-Hwung Su, and Wei-Jyun Huang

Department of Computer Science and Information Engineering,
National Cheng Kung University, Tainan, Taiwan, R.O.C.
tsengsm@mail.ncku.edu.tw

Abstract. Semantic image retrieval basically can be viewed as a pattern recognition problem. For human, pattern recognition is inherent in herself/himself by the inference rules through a long time experience. However, for computer, on the one hand, the simulated human identification of objects is impressive at its experience (training) like a baby learns to identify objects; on the other hand, the precise identification is unreasonable because the similar features are usually shared by different objects, e.g., "an white animal like cat and dog", "a structural transportation like car and truck". In traditional approaches, disambiguate the images by eliminating irrelevant semantics does not fit in with human behavior. Accordingly, the ambiguous concepts of each image estimated throughout the collaboration of similarity function and membership function is sensible. To this end, in this paper, we propose a novel fuzzy matching technique named *Fuzzy Content-Based Image Retrieval (FCBIR)* that primarily contains three characteristics: 1) conceptualize image automatically, 2) identify image roughly, and 3) retrieve image efficiently. Out of human perspective, experiments reveal that our proposed approach can bring out good results effectively and efficiently in terms of image retrieval.

Keywords: multimedia database, content-based image retrieval, data mining, fuzzy set, fuzzy search.

1 Introduction

A huge amount of images are generated in our everyday life as the fast growth of advanced digital capturing devices for multimedia, such as digital camera and mobile-photography phone. Through WWW, the collective image repository will be further bigger and bigger because of the speeding exchange of these life images. As a result, how to access the growing heterogeneous repositories effectively and efficiently has been becoming an attractive research topic for multimedia processing. Basically, semantic image retrieval can be viewed as a pattern recognition problem. For human, pattern recognition is inherent in herself/himself by the inference rules through a long time experience. However, for computer, it is hard to represent an image out of human aspect even though a number of researchers attempt to investigate a powerful identification algorithm from different visual viewpoints. It tells us the truth that there still exists a large improvement ground for image recognition. Classic approaches make use of image features like color, texture and shape to calculate the similarities among images, called visual-based or content-based image retrieval (CBIR). The

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 141–150, 2007.
springerlink.com

main drawback of this-like approaches is that it is hard to represent the diverse concepts of an image just by a set of low-level visual features. To be closer to human sense, the other ways to connect human sense and machine cognition are classification and annotation, called textual-based image retrieval. Practically, both of classification and annotation put the focus on distinguishing the specific semantics of images by computing feature dissimilarities among them. Unfortunately, couples of objects (categories) in real world always share the same features and hence they are so difficult to be identified precisely. For example, yellow color and circle shape are shared by many objects like sun, egg yolk and etc. Accordingly, for computer, on the one hand, the simulated human identification of objects is impressive at its experience (training) like a baby learns to identify objects; on the other hand, the precise identification is unreasonable because the similar features are usually shared by different objects, e.g., “an white animal like cat and dog”, “a structural transportation like car and truck”. Besides, to raise the accuracy of image retrieval up, derivative complex computations will really damage the execution time, and the poor performance cannot satisfy user’s requirement. Therefore, in this paper, we propose a novel fuzzy matching technique named *Fuzzy Content-Based Image Retrieval (FCBIR)* that primarily contains three characteristics: 1) conceptualize image automatically, 2) identify image roughly, and 3) retrieve image efficiently. The rest of this paper is organized as follows: Section 2 briefly describes the previous works on image retrieval. Section 3 introduces the proposed method in detail. Experiments on our approach are illustrated in Section 4 and conclusions and future work are stated in Section 5.

2 Related Work

Image retrieval has been a hot research issue for a long time because it can prevent the search from costing expensively by efficient image recognition. General visual-based similarity matching methods primarily take advantage of extracted features to accomplish the image retrieval, e.g., [12]. Unfortunately, this-like approaches cannot provide enough semantic support to help user get accurate results since visual features cannot supply common users sufficient information to identify the semantic they want. In addition, data mining is another way to make effective image retrieval. Chabane Djeraba [5] proposed an approach by using association mining [2] for content-based image retrieval. In this approach, it generates an efficient visual dictionary that summarizes the features in database. Each feature of visual dictionary associated with a symbolic representation help users find out the images effectively. The other way to reduce the gap between low-level features and high-level concepts is to let the images be with proper concepts, such as classification and annotation, e.g., [1][6][10][11]. Indeed, the mutual aim of existing approaches is to do image retrieval a good favor, but in vain. The similar experienced phenomenon also exists in most of the other AI research fields. The major reason is that the precise process is very difficult to deliver the exact concept in user’s mind. Hence, fuzzy set theory has been adopted by more and more recent intelligent systems due to its simplicity and similarity to human reasoning [7][8]. FIRST (Fuzzy Image Retrieval SysTem) proposed by Krishnapuram *et al.* [9] uses Fuzzy Attributed Relational Graphs (FARGs) to represent images where each node in the graph represents an image region and each edge represents a relation between two regions. Every query is converted into a FARG to compare with the FARGs in the database. Chen *et al.* [4] proposed UFM (unified feature matching) to retrieve the images. In this study, an image is represented as several segmented regions based on a fuzzy feature. Nevertheless, the effects of above two fuzzy approaches are both on the foundation of segmented regions and the region segmentation still has not been very promising until now. Therefore, in this paper, we propose a new fuzzy matching technique to touch user's mind without region segmentation.

3 Proposed Approach: FCBIR

As mentioned above, the goal of our proposed approach is to make effective and efficient image retrieval with fuzzy human concept. To achieve this goal, we integrate similarity function and membership function to assist the fuzzy image retrieval, as shown in Figure 1. The major task of proposed approach can be briefly decomposed into three following subtasks.

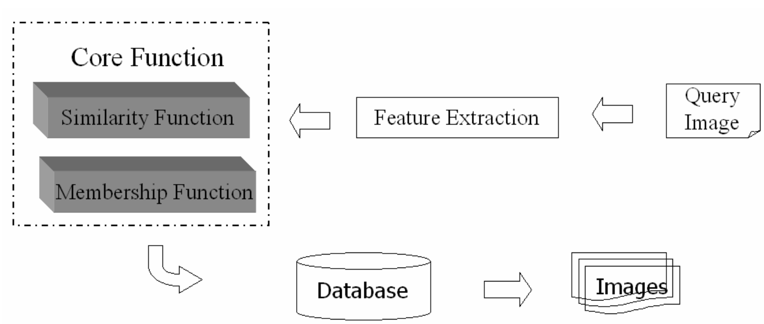


Fig. 1. The work flow diagram of FCBIR

I. Construction of Data Ontology: This phase involves some infrastructural works that include defining the data ontology, collecting all related images and clustering these collected images. In general, this phase can be regarded as an essential work for preprocessing images stored in the database.
II. Transformation of Fuzzy Sets: As those categorized images are clustered in above phase, similarity function and membership function will cooperate to let the images be with rough semantics in this phase.
III. Exploration of Images: Once the system receives a query submitted from user, the proposed matching algorithm performs a nice concept exploration of images. According to the specific concept picked by user, she/he thus can obtain the preferred images further.

In the followings, we will describe above works in great detail.

3.1 Construction of Data Ontology

Generally, this idea is motivated by natural human learning because the reason why humans can identify an object is that the viewed object can be identified by the similar objects of data ontology in her/his memory. Unlike traditional similarity matching approaches based on low-level features, the categorized images of pre-defining semantic ontology can actually facilitate the image concept retrieval like human learning. Hence, we take a look at the construction of data ontology in the beginning of designing this system. In this work, first, the frame of concept ontology has to be defined since it is projected on by the query images during the concept retrieval phase. As shown in Figure 2, it can be considered as a tree structure composing of hierarchical nodes structured by linguistic terms. Second, gather all categorized images belonging to each leaf node and store them into the database. In fact, without the exact collection, we cannot conceptualize the image excellently. Third, cluster the images of each category individually. More seriously, clustering is a fundamental but critical preprocessing operation for image identifications. In the third procedure, features, such as Color Layout, Color Structure, Edge Histogram, Homogeneous Texture and Region Shape, are extracted by the popular tool XMTool [3], and the similarities are generated by calculating Euclidean Distance d of v to u ($d = \sqrt{\sum_{i=1}^{n} (v_i - u_i)^2}$, where n is the number of feature vectors) during the clustering period. At last, the images of each category are clustered into equalized groups by famous cluster algorithm k-means. Its physical meaning is that we can discover the images with the same semantic but different views. For example, a category “car” usually contains different visual features like “color”, “texture” and “shape”, and each car may be with different visual properties like “front-view”, “back-view” and “side-view”.

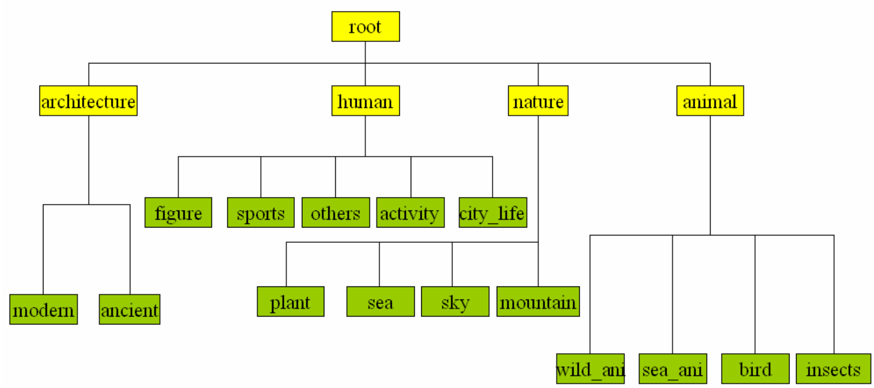


Fig. 2. Example of data ontology

Due to clustering operation does make a significant impact on the accuracy for both conceptualizing and exploiting the images, one of the important issues in our proposed method is the quality of clustering. Thus, we employ three measures to make the validation for clustering quality.

Local Density: Local Density is the density of each cluster. It delivers the entropy of each cluster. A cluster with lower density shows us a poor clustering effect because most of points in this cluster are very dissimilar.

Global Density: Global Density is the density of all clusters in the global space. In contrast with local density, good cluster dispersion is with a longer average distance among global clusters.

Local Proportion: In order to reach the presetting quantity of clusters, thereby the cardinality of images of each cluster will not be the same. The higher local proportion represents a reliable quality of clustering.

On the basis of above, we can set three thresholds to ensure the clustering being good enough to offer sufficient support for the following tasks described in the next two subsections. That is, the clustering algorithm ends while three thresholds are all satisfied.

As described above, a cluster for each category can be considered as a view with the visual-distinguishability property. Another crucial issue in this step is how to generate the sample-image for each cluster. Given $N = \{ca_1, ca_2, \ldots., ca_i\}$ denotes a set of categories and $CL_{ca_i} = \{cl_1, cl_2, \ldots., cl_j\}$ denotes a set of clusters belonging to the category ca_i, and each cluster contains m images $cl_j=\{I_1, I_2, \ldots., I_m\}$. Then the sample-image of cl_j, $SMMG$, is defined as:

$$SMMG_{cl_j}(C,S,T) = \left(\frac{\sum_{x=1}^{m} C_x}{m}, \frac{\sum_{y=1}^{m} S_y}{m}, \frac{\sum_{z=1}^{m} T_z}{m}\right) \quad (1)$$

where $\{C, S, T\}$ represents the feature set {color, shape, texture}. After calculating the $SMMG$, we can find out the sample-image of each cluster of each category. Once the sample-images are identified, the images in the database are easy to be conceptualized by projecting them onto the concept ontology with computing the minimized similarities to these clusters.

3.2 Transformation of Fuzzy Sets

Actually, the images that are conceptualized with some rough semantics in this phase will enable the retrieval to be closer to human sense. In traditional approaches, disambiguate the images by eliminating irrelevant semantics does not fit in with human behavior. For example, some objects we never see possibly get couples of linguistic terms. To this end, the ambiguous concepts of each image estimated throughout the collaboration of similarity function and membership function in this phase is sensible. As shown in Figure 3, the whole process of Algorithm Trans_Fset for transformation of fuzzy sets can be elaborated on two following steps.

I. Similarity Calculation: In line 3 of Algorithm Trans_Fset, similarity mainly depends on computing the distance between an image and the sample-images of each cluster of each category. Accordingly, the processed image can pertain to some semantics with respect to the clusters that are with shorter distances derived from the above similarity function. For example, assume that k is 20. An image with four relevant concepts can be represented as {(animal, 7), (insect, 3), (building, 5), (plant, 5)} after the similarity calculations.

Input: The images in database D, predefined categories with grouped clusters, a set of membership functions
Output: Table T containing images with fuzzy sets
1. Define cardinality k;
2. **for** each image $I_j \in D$ **do**
3. Calculate distances and discover the top k closer clusters;
4. Calculate the count cnt_{ca_i} $(0 \leq cnt_{ca_i} \leq k)$ of each category ca_i from closer k clusters;
5. **for** each category with $cnt_{ca_i} \neq 0$ **do**
6. Convert cnt_{ca_i} of ca_i into a fuzzy set $f^j_{ca_i}$ denoted as $\left(\frac{M_1^{ca_i}}{R_1^{ca_i}} + \frac{M_2^{ca_i}}{R_2^{ca_i}} + ... + \frac{M_n^{ca_i}}{R_n^{ca_i}}\right)$ by employing the given membership functions, where $R_n^{ca_i}$ is the n^{th} fuzzy region of ca_i and $M_n^{ca_i}$ is the fuzzy membership value in region $R_n^{ca_i}$;
7. $F^j = \cup f^j_{ca_i}$;
8. **end for**
9. $T = \cup F^j$;
10. **end for**
11. **return** T

Fig. 3. Algorithm Trans_Fset

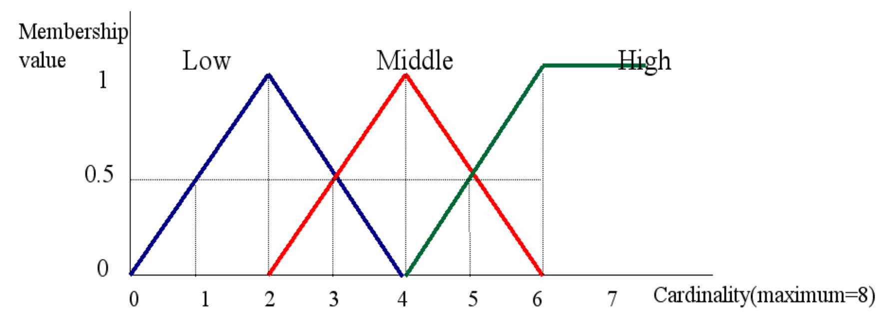


Fig. 4. Fuzzy membership functions for cardinality attribute

II. Fuzzy Set Transformation: After similarity calculations, the fuzzy sets of each image in the database can be implied by our proposed membership functions based on [7], and from line 4 to line 9 of Algorithm Trans_Fset, the transaction table T will soon be yielded by these fuzzified images. For example as above, the third concept (building, 5) of an image can be converted into the fuzzy set ($\frac{0.0}{building.low}+\frac{0.5}{building.middle}+\frac{0.5}{building.high}$) by employing the given membership functions as shown in Figure 4. The whole procedure ends while all the concepts in each image are converted into fuzzy sets. In our proposed membership functions, cardinalities are represented by three fuzzy regions: *Low*, *Middle* and *High*. Thus, *concept.term* is called a fuzzy region.

3.3 Exploration of Images

As the fuzzilized table being ready, the system will perform image search algorithm described in this subsection. This phase generally concerns the procedure that starts with while the image queried by user, the system first analyzes the query image by similarity function and membership function. Then the query image with fuzzy sets will be compared with the fuzzilized images in the database by executing the proposed matching algorithm FIM, as shown in Figure 5. At last, the system responses the ranking images for each related concept. In detail, if FIM finds out the images with fuzzy regions fully hit by the query, the most similar images for each related concept are selected. Otherwise, the top-m images with fuzzy regions partially hit by the query will be filtered by *I_SIM* and *I_DISIM* described as follows. That is to say, the top-m images are mainly selected by the higher *I_SIM*, and if the images with the same *I_SIM*, those with higher *I_DISIM* will be discarded further. Finally, the referred concepts can be derived from these selected images by using *C_SIM*.

As stated above, in this algorithm, three main functions that are devoted to calculate image similarity and concept similarity will be described below. Assume that the amount of the images with full-hit fuzzy sets to the query for category ca_i is N_{ca_i} and the amount of the images for category ca_i is Q_{ca_i}, then the similarity of the query image to ca_i is:

$$C\_SIM_{ca_i} = \frac{N_{ca_i}}{Q_{ca_i}}, \tag{2}$$

and given that the j^{th} image $I_j \in D$ with the fuzzy sets $I_j=\{\{same\}, \{diff\}\}$, where *same* is a set consisting of the same *sr* fuzzy regions to the query image and *diff* is a set consisting of different *dr* fuzzy regions to the query image, $diff=I_j \backslash same$, then the similarity of the query image to I_j is:

$$SIM_{I_j} = \sum_{sr=1}^{|same|} membership(sr), \tag{3}$$

and the dissimilarity is:

$$DISIMI_j = \sum_{dr=1}^{|diff|} membership(dr) \tag{4}$$

Apart from *C_SIM*, in order to capture higher-level concepts of the concept ontology, we can further compute the accumulated similarities of higher-level concepts to the query image by the following equation.

$$HC\_SIM_l = \frac{\sum_{i=1}^{|HC_l|} C\_SIM_{ca_i}}{|HC_l|} \quad (5)$$

where HC_l is l^{th} higher-level concept that is the ancestor of the category ca_i.

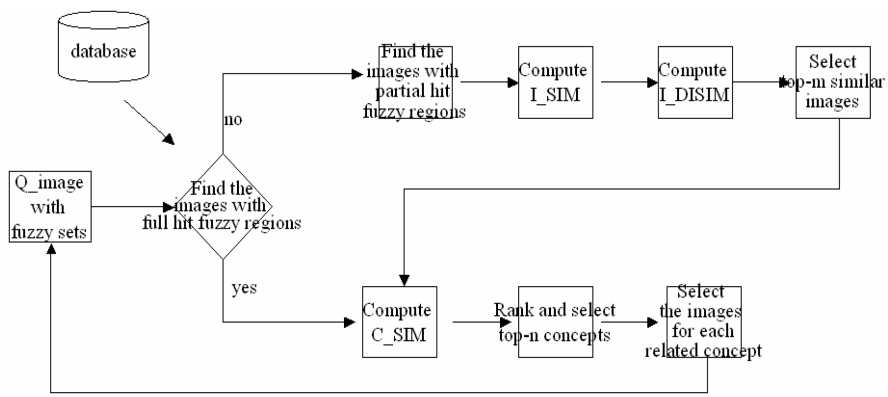


Fig. 5. Algorithm FIM (Fuzzy Image Matching)

4 Experimental Evaluation

In previous section, we have expressed the proposed approach for fuzzy content-based image retrieval. Now we describe the prototype of this system and the results in evaluating the performance of proposed method by experiments using real image data. As shown in Figure 2, there were 36867 images collected from Corel spreads in 15 categories that vary in amounts, and each category is grouped into 8 clusters. Figure 6 depicts the system prototype. In this interface, the query image displayed on the top of the left frame is detected for several ranking concepts, and the referred higher-level concepts are displayed at the bottom of the left frame. This query image is originally classified into concept "activity". In traditional approaches, its implicit concepts, such as "sea", "sports", "human" and "nature", are hard to be estimated, but our proposed approach comparatively can catch these hidden concepts. Out of these implied concepts, the most similar images of top-5 concepts are shown in the right frame. User can further pick the preferable concepts or images by clicking them. Every picked concept will represent the most similar 20 images to the query. Regarding our experiments, most of images can be detected for its correct concept or higher-level concepts. Even more implicit reasonable concepts can be also found. Furthermore, we

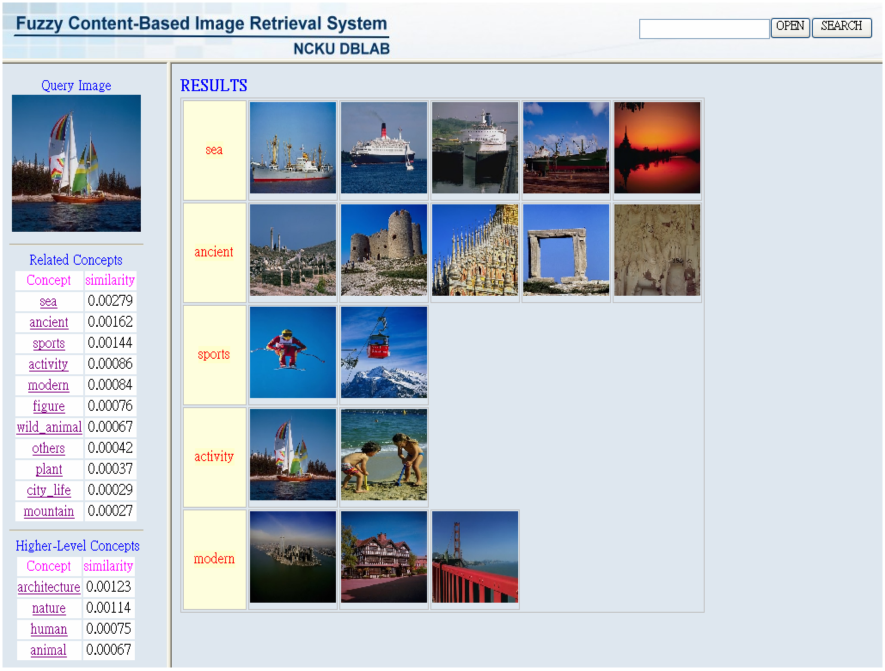


Fig. 6. Example of query results

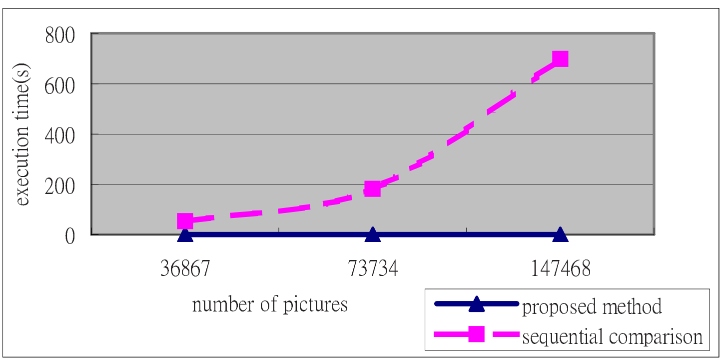


Fig. 7. Execution efficiency for image search

doubled and tripled the population of original images to evaluate its performance. In this experiment, the fuzzy sets of all images for our proposed approach are pre-generated and stored into SQL server in advance, and the features of those for sequential comparison approach are pre-extracted and stored in a feature vector list. Figure 7 illustrates that the performance of our proposed approach outperforms the sequential comparison approach significantly in terms of execution time.

5 Conclusions and Future Work

In this paper, we have represented a new fuzzy matching technique for content-based image retrieval by combining visual features and fuzzy sets. The rough concepts mined exhibits a reasonable machine learning from human aspect, and moreover it can furnish a great support to assist a common user in exploring the images from a large-scale database effectively and efficiently. Without additional segmentation operation and sequential comparison, membership function can facilitate the rough concept search together with similarity function. In the future, we will further keep an eye on the relevance feedback and develop an effective approach for the interaction with user during the retrieval phase. Besides, investigate optimal settings about cluster number, *k* clusters closer to the query, similarity function and membership functions is another critical issue that perhaps can bring out a better result.

Acknowledgments. This research was supported by Ministry of Economic Affairs, R.O.C., under grant no. 95-EC-17-A-02-S1-024 and by National Science Council, R.O.C., under grant no. NSC95-2422-H-006 -001

References

1. Zaher Aghbari and Akifumi Makinouchi, "Semantic Approach to Image Database Classification and Retrieval," *NII Journal*, no. 7, September 2003.
2. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", In *Proc. 20th VLDB Conference*, 1994, pp. 487-499.
3. Bouajila, C. Claus and A. Herkersdorf, "MPEG-7 eXperimentation Model (XM)." Available at: *http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/ e_ mpeg7.html.*
4. Yixin Chen, James Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 24, Issue 9, September 2002.
5. Chabane Djeraba, "Association and Content-Based Retrieval," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No. 1, 2003.
6. H. Feng, R. Shi and T.S. Chua "A bootstrapping framework for annotating and retrieving WWW images." In *Proc. of the 12th annual ACM international conference on Multimedia Technical session 15*, 2004.
7. T.P. Hong, K.Y. Lin, and S.L. Wang, "Mining Fuzzy Generalized Association Rules from Quantitative Data under Fuzzy Taxonomic Structures," *International Journal of Fuzzy Systems*, Vol. 5, No. 4, 2003, 239-246.
8. A. Kandel, *Fuzzy Expert Systems*, CRC Press, Boca Raton, 1992, pp. 8-19.
9. Raghu Krishnapuram, Swarup Medasani, Sung-Hwan Jung, Young-Sik Choi, and Rajesh Balasubramaniam, "Content-Based Image Retrieval Based on a Fuzzy Approach", *IEEE Transactions on Knowledge and Data Engineering* , Vol. 16, No. 10, pp. 1185-1199, October 2004.
10. J. Li, A. Najmi, and R.M.Gray, "Image classification by a two-dimension hidden Markov model," *IEEE Transaction on Signal Processing*, vol. 48, no. 2, pp. 517-533, February 2000.
11. J. R. Quinlan, "C4.5: Programs for Machine Learning", *Morgan Kaufmann Publishers*, 1993.
12. J. R. Smith and S.-F. Chang, "VisualSEEK: A fully automated content-based image query system." In *Proc. of the 4th ACM international Conference on Multimedia*, 1996.

Computing with Words Using Fuzzy Logic: Possibilities for Application in Automatic Text Summarization

Shuhua Liu

Academy of Finland and Abo Akademi University, IAMSR
Joukahaisenkatu 3-5 B, ICT Building 4th floor, FIN-20520 Turku, Finland
sliu@abo.fi

Abstract. The theory of "computing with words" offers mathematical tools to formally represent and reason with perceptive information, which are delivered in natural language text by imprecisely defined terms, concepts classes and chains of thinking and reasoning. It thus provides relevant methods for understand-based text summarization systems. In this paper we propose a framework for text summarization and examine the possibilities as well as challenges in using the concepts and methods of CW in text summarization.

Keywords: Text summarization, natural language understanding, computing with words, fuzzy logic.

1 Introduction

Text summarization is a process of distilling the most important/essential/critical information from a text source and formulating a concise text to serve a particular user and tasks. In some occasions, specific pieces of information will be picked up from a text document, when the "most important" information is regarded the same as the "most needed" and when the "needs" can be pre-specified. Key issues in automated text summarization include how to identify the most important content out of the rest of the text and how to synthesize the substance and formulate a summary text based on the identified content.

So far, there have been basically two different approaches to text summarization [3][4][5]: the selection-based approach and understanding-based approach. Text summarization systems can choose to make use of shallow text features at sentence level, discourse level, or corpus level to locate the important sentences, which then make-up the output summary. They often treat the "most important" as the "most frequent" or the "most favorably positioned" content, thus avoid any efforts on deep text understanding. On the other hand, text summarization systems may also choose to imitate human summarization process thus are based on an understanding of text meaning and content. Selection based methods are easy to implement and generally applicable to different text genres but it is usually very hard to achieve performances that exceed a generally attained level. Understanding-based methods, on the other hand, always require a reasonably complete formal specification of document content, based on which a knowledge-rich approach could help deliver a quality summary text

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 151–160, 2007.
springerlink.com

in terms of both the substance and the presentation. However, the better quality of a summary is traded over general applicability and flexibility of the system. Many summarization tools and systems have emerged as a result of the sustaining research efforts since the 1990s, but with most of the practical text summarization tools of today are selection based rather than understanding based. While selection based methods are good at quickly identifying the "important" content, it is understanding-based methods that would be good at synthesizing the selected information. The theory of fuzzy logic based "computing with words (CW)" offers mathematical tools to formally represent and reason about "perceptive information", which are delivered in natural language text by imprecisely defined terms, concepts classes and chains of thinking and reasoning. It thus provides relevant methods for understand-based summarization systems. Until recently, the application of fuzzy logic in natural language understanding has been discussed only sparsely and scattered in the literature of soft computing and computational linguistics, although the theoretical foundation has been laid out in several articles by Prof. L.A. Zadeh already decades ago [8][9][10]. It is our intension in this paper to examine the possibilities as well as challenges in using the concepts and methods of CW in understanding based text summarization.

2 Cognitive Models of Reading Comprehension and Summarization

The theoretical foundation for understanding based text summarization ssytems is found in the cognitive models of reading comprehension, among which the microstructure-macrostructure model proposed by Kintsch and van Dijk [2][7] is perhaps the most influential. The model takes as its input a list of propositions that represent the meaning of a text segment. The output is the semantic structure of the text at both micro and macro levels represented in the form of a coherence graph. Such a structure is believed will enable the full meaning of the text be condensed into its gist. A microstructure refers to the semantic structure of sentences. It reflects the individual propositions and their close neighboring relations. A macrostructure presents the same facts as the whole of microstructures, but describes them from a global point of view. They are both represented as a set of ordered and connected propositions. The order of the connection is determined particularly by their "referential relations" in the form of argument overlapping.

Microstructure and macrostructure are related by a set of semantic mapping rules called "macro-rules": detail-deletion rule, irrelevance-deletion rule, rule of generalization and rule of construction. Macro rules are applied in "macro-operations" that derive macrostructures from microstructures. Macro-operations are controlled by a "schema", which is a formalized representation of the reader's goals and it helps to determine the relevance/irrelevance and importance/unimportance of propositions.

In reading process, "a reader often also tries to produce a new text that satisfies the pragmatic conditions of a particular task context or the requirements of an effective communication context". "The new text will contain information not only remembered from the original text, but also re-constructively added explanations or comments based on his knowledge and experience." A summary is one of such new texts. Kintsch and van Dijk [2] noted four types of text production activities associated with text reading: (i) lexical and structural transformations such as lexical substitution, proposition reordering, explication of coherence relations among propositions, and perspective changes! (ii) Through the memory traces, particular contents will be retrieved so that they become part of the new text. (iii) When micro and macro information is no longer retrievable, the reader will try to reconstruct the information by applying rules of inference to the information that is still available. (iv) A new text may also be some meta-comments on the structure or content of the text such as giving comments, opinions, expressing attitudes.

Finally, knowledge is indispensable in effective reading and comprehension. A reader's knowledge determines to a large extent the meaning that he or she derives from a text. Knowledge that is important in reading comprehension can be grouped as four types: linguistic knowledge (phonetics, morphological, syntax, semantics, pragmatics knowledge, genre knowledge), general world knowledge (commonsense knowledge as well as socially known properties of some social and natural world), specialized domain knowledge and context/situational knowledge (task context, communication context, location context) [1] [6]. World knowledge is of many different types, which do not always apply in discourse processing in the same way but are instead personally and contextually variable [2].

3 Computing with Words (CW): Fuzzy Logic Based Natural Language Content Representation and Reasoning

Key to understanding-based summarization is the capability to correctly interpret word meaning and sentence meaning (i.e. to relate linguistic forms to meaning), to formally represent the text content (to map natural language expressions onto a formal representation), and to derive the most important information by appropriately operating/reasoning on the formally described content.

There have been many different ideas about how meaning and knowledge can be represented in human mind and machines: semantic networks, frames and logics. Logics are much better received by formal treatment than frames and semantic networks, and have been playing a significant role in language processing [1] [6]. In linguistics as well as psychology studies, first order predicate calculus (FOPC) or its variants has been the most popularly adopted logic in representing and analyzing text meaning. Natural language text contains rich predicate-argument assertions that can be generally treated as propositions that lend themselves to logical representation and operations.

However, it is an acknowledged fact that there is a sharp contrast and mismatch between the formality and precision of classical logic and the flexibility and variation of natural languages. Natural language abounds with perceptive information and perceptions are intrinsically imprecise and fuzzy. Although widely applied in NLP systems, classic logics such as FOPC have significant limitations in terms of expressing uncertain and imprecise information and knowledge. It has only rather limited power in expressing qualitative quantifiers, modifiers, or propositional attitudes (associated with words like believe, wants, think, dream, knows, should, and so on). Fuzzy logic, on the other hand, provides the necessary means to make qualitative values more precise by introducing the possibility of representing and operating on various quantifiers and predicate modifiers, which help to maintain close ties to natural language [8][9][10]. Thus, in principle, the imprecision and vagueness of terms, concepts and meaning in natural language text could largely be treated in a quantitative way using the method of CW.

One basic notion of CW is that imprecise, uncertain and ambiguous information needs to be precisiated first. Precise meaning can be assigned to a proposition *p* drawn from a natural language by translating it into a generalized constraint in the form $p \rightarrow$ X isr R, where X is a constrained variable and R is a constraining relation that is implied in *p* and the text context; r is an "indexing variable" whose value intensifies the way in which R constraints X. The principal types of generalized constraints include e.g. equality (r is =), possibilistic (r is blank), probabilistic (r is p), random set (r is rs), usuality (r is u), fuzzy graph (r is fg), etc. [11][12]. A natural language proposition *p* in a generalized constraint form is called precisiated natural language (PNL).

This means that, with PNL, the meaning of a lexically imprecise proposition is represented as an elastic constraint on a variable or a collection of elastic constraints on a number of variables. The translation of *p* into generalized constraint is a process of making the implicit constraints and variables in *p* explicit. An "explanatory database" needs to be constructed first, which will help to "identify" and "explicate" the constrained variable and constraining relation in different types of propositions based on test-score semantics [11]. Once the data for reasoning is ready, the reasoning process is treated as propagation of generalized constraints. Rules of deduction are equated to rules which govern the propagation of generalized constraints. Deduction rules drawn from various fields and various modalities of generalized constraints reside in a deduction database. Generalized constraints in conclusions need to be retranslated into propositions expressed in a natural language.

4 Computing with Words Using Fuzzy Logic: Possibilities for Application in Text Summarization

Text summarization as the purpose or byproduct of a reading and comprehension process means in essence three things: (i) text understanding; (ii) finding out what is important; (iii) rewrite a number of important messages to form a coherent text. "What is important" differs to different people and differs in different situations; the focus of "importance" changes with the change of the context. Important information may be "new information" to fill the cognitive gaps in reader's memory and mental models concerning world objects. Important information may be a direct or derived answer to a query. Important information may also simply refer to the core content that the author intends to convey to the readers with bulk of background details, repetitions, comments omitted. Abstraction and generalization is a key component in creative text and information summarization. Formal or informal representations of events, actions, and properties can be generalized by abstraction along the time, place, participants, other properties dimensions, and so on.

The Kintsch-Dijk model recognize that a macrostructure could be able to capture the most important or essential object denoted by a sequence of propositions and thus to represent the gist of a text segment. Thus, a summary can be generated through the deriving of a marcostructure from microstructures by deleting details, deleting irrelevant proposition, generalizing multiple propositions and constructing new propositions. This process goes on recursively on sequences of micro propositions as long as constraints on the rules are satisfied. This in certain sense resembles the process of constraint propagation in CW. Both micro and macro propositions denote both hard facts and soft perceptions that could be formalized as generalized constraints. And the process of inducting macro propositions from micro propositions can be realized as propagation of generalized constraints, as well as abstraction and generalization with protoforms.

If we look at existing content analysis systems, they are largely built upon language models for IR tools which are usually based on "words" instead of "meanings". Such language models loses a lot of syntactic and semantic information about the original document but often maintains enough information to decide if two documents are on the similar topics, thus help in document retrieval. Although they alone are not sufficient tools for performing more intellectual and cognitively demanding tasks such as language understanding, they prove to be very helpful in quickly identifying relevant/important sentences or passages through quick statistical analysis of text surface features. Thus, the trend is to develop and test more sophisticated text modeling and analysis methods that take advantage of both selection based methods and understanding based methods. On the basis of the theories introduced earlier, a framework of summarization is proposed (Figure 1).

This is a summarization process embracing two phases. In the 1st Phase, important/relevant text segments are extracted from the original text making use of only shallow features of the text. The targeted text segments may be for the purpose of composing a generic summary or query-specific summary. Topic words or phrases can also be identified quickly for each text segments, to serve as context cues useful in semantic analysis. Topics can also be used as cue words for scoring the text passages to be included the summary. There are a collection of statistical methods and techniques that are applicable. The 1st Phase also includes a support module "corpus analysis" which is an umbrella term that represents a variety of different types of learning based corpus analysis functions that will help, for example, to fine tune the methods used in topic identification, theme analysis, and passage selection, and so on. In the 2nd Phase, selected passages are input to an intensive text analysis and language understanding process. First, syntactic and semantic analysis will transform free texts (and natural language query in the case of question answering) into connected listed of natural language propositions that together can represent the microstructures. These propositions will reflect the "atomic" information (facts or perceptions) contains in the text. Two parallel processes can continue with the natural language propositions.

On one hand, macro semantic structure is to be derived from the microstructures, which is guided by pre-specified macro rules and the controlling schema determined according to text genre or derived from query description. The controlling schema determines which micro propositions to be deleted or what generalizations of micro propositions are relevant, and thus which part of the text will form its gist. On the

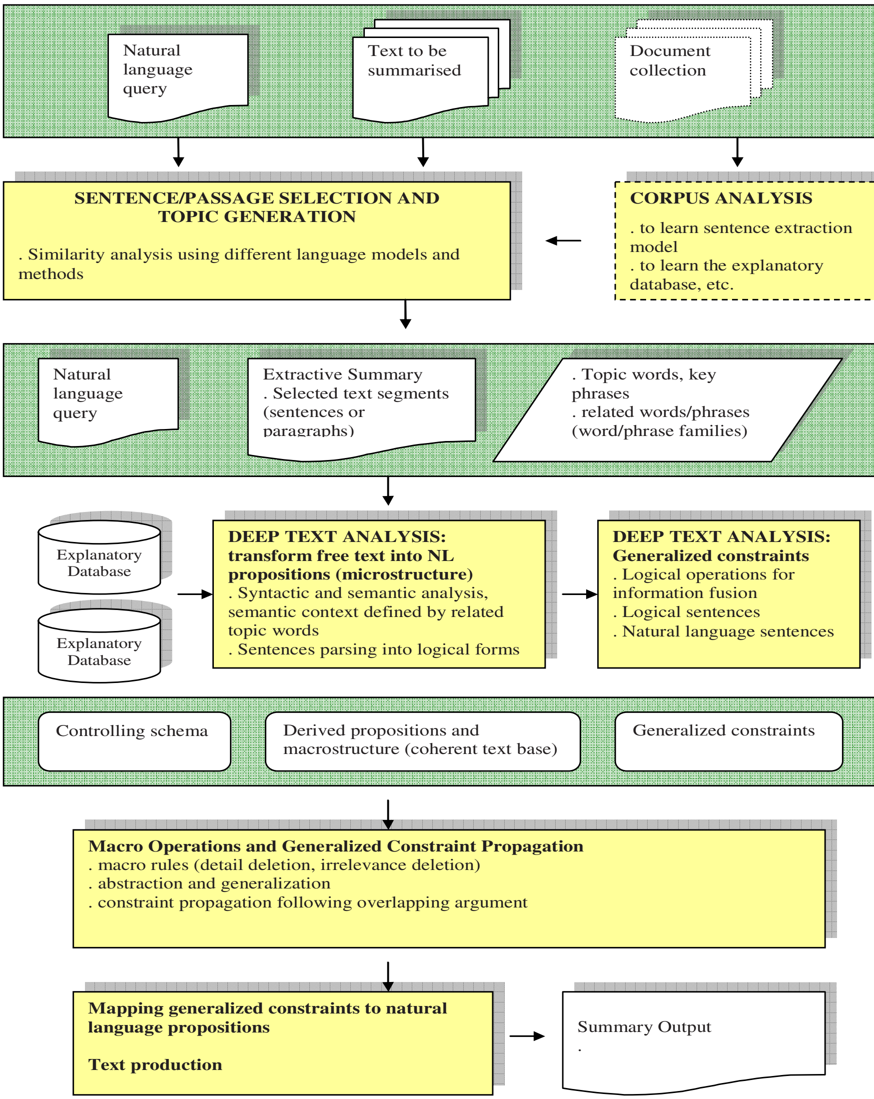


Fig. 1. A Framework for Text Summarization

other hand, natural language propositions can also be transformed into Generalized Constraints with the support of Explanatory Database. Ambiguity resolving mechanisms are the tools for the construction of the explanatory database in the CW framework. Constraint propagation will derive conclusions from facts presented by the collection of propositions. This process is driven by query or topic description propositions.

4 Discussion

One evident limitation in the Kintsch-Dijk model is that it only takes into consideration the referential relations between words/phrases while neglects any connectives or other discourse relations between sentences. The formulation of discourse meaning structure is solely based on a coherent referential relation that does not necessarily equals to meaning relations. A macrostructure must be implied by the explicit microstructure from which it is derived, while the explicit microstructures often only supply incomplete information on the meaning of the content. For human readers, natural language discourse may be connected even if the propositions expressed by the discourse are not directly connected through referential relations because a reader is often able to provide the missing links of a sequence on the basis of their general world knowledge or contextual knowledge as well as inferences [2]. For computing systems this is the hardest challenge. The topmost propositions resulted from referential relationship analysis only represent presuppositions of their subordinate propositions; there is no guarantee that they are truly the most important content and information [2].

Although it is not perfect as a model for summarization, the Kintsch-Dijk model presents a useful framework for us to understand the mental processes taking place in reading comprehension. Text understanding with the implementation of the Kintsch and Dijk model is not the best choice for the processing of long text. But it offers some useful methods for the interpretation and reasoning of shorter text segments such as paragraphs or passages. With the support of proper parsing tools and learning tools, the construction of micro and macro semantic structure of discourse (the explication of predicates as well as propositional arguments) is possible.

The fuzzy logic based theory of CW takes a very different approach from traditional approaches in computational linguistics. It suggests a relevant framework for natural language understanding system. It seems to offer a better framework and more suitable methodology for the representation of and reasoning with meaning, knowledge and strength of belief expressed in natural language than is possible within the framework of classical logic. The proposed constraint propagation reasoning mechanism can derive new constraint statements from groups of constraint statements. An advantageous application for it would be to derive from natural language text answers to natural language queries. Text summarization could benefit from such an approach when the required summary can be created as the collected answer to a collection of queries about certain topic or object. Depending on how the questions and queries are formulated, the resulted summary may represent the text content from multiple viewpoints and angles.

CW is a relevant approach to tasks of understanding based text summarization, but there seem to be considerable difficulties in applying it in practice presently. Firstly, to apply its concepts and methods in text summarization presupposes an accurate natural language processing function that can transform free text into natural language propositions and then to the form of generalized constraints. This assumes that the constrained variables and constraining relations can be reliably defined, which presents a serious obstacle.

Secondly, the specification of generalized constraints requires the appropriate granulation of attribute values and calibration of lexical constituents (adverbs, adjectives) of propositions available. In reality however, this is as difficult a task as to define the linguistic variables and relations implied in NL sentences, if not more. It is made harder by the fact that standard calibration of value terms in many cases does not exist and it is very common that different people calibrate the same value term differently. So even though the operations on membership values is consistent and can be done in a systematic manner according to standard mathematical functions, the initial definition of value terms may be inappropriate. Finding effective ways for constructing the explanatory database for natural language texts in sufficiently specific domains and facilitating the calibration of lexical constituents of propositions will be the key to the implementation of the working systems.

Thirdly, although constraint propagation will make precise reasoning about the presented generalized constraints, there are chances that the derived conclusion may be invalid because (i) the input to the reasoning process (i.e. the generalized constraints) were wrongly identified; (ii) the fact that not every proposition in natural language is precisiable [12]. Incorrect input combined with a formal reasoning process result in false conclusions.

Fourthly, to deal with perceptive information, the theory of computing with words using PNL introduced notations of sufficient generality and expressiveness, while will inevitably become complicated and computationally heavy. Such complexity does not necessarily guarantee better results in summarization due to e.g. the lack of other elements (such as sufficient world knowledge) in the process.

5 Conclusions

Natural language presents the most informal model of the world, however, also the most correct model usually. Natural language model of the world is certainly not the most precise one, but usually precise enough. By means of computational models, it is not possible to seek for a “complete” “accurate” or “ideal” state of understanding of natural language discourse. And perhaps it is not necessary either trying to attain a fully, correct understanding of a text in order to summarize it. At least for many indexing and retrieval systems there are designs that do very well without any specific linguistic analysis of the text material. Although many current systems lack the deductive reasoning capability, very useful tools has already been developed using a bit simple and mature techniques.

The Kintsch-Dijk model regards gist formation and summary production as an integral part of the cognitive process of reading comprehension and text production. It suggests one way for formulating what is important in a text as a mapping from its micro semantic structures to macro semantic structures. It also pointed out that a summarizer not only reads and interprets text meaning and content, but also actively reconstructs meaning according to his prior knowledge or with respect to his information gaps. The process of meaning construction is often the reasoning or inference that draws upon prior knowledge to fill the gaps of incoming information and that adapts the new knowledge to what is already in memory.

Zadeh's pioneer work on CW and generalized theory of uncertainty laid the theoretical foundations for representing and reasoning with information in all format, numerical or perceptive. Meaning and information is represented as a generalized constraint on the values which a variable is allowed to take. It provides a more general and encompassing conceptual framework of sets and logic that is able to handle both precise and imprecise information and meaning. Although fuzzy systems still do not necessarily outperform human in dealing with uncertainty and imprecision, it helps to reduce the real world problems to a scale that is possible for computing solutions that was impossible before.

The methodology of CW contributes to text summarization by contributing to the general process of representing and reasoning with perceptive information. Text summarization can benefit from CW by being formulated as a process of seeking answers for a collection of text based queries. It is thus one of the many tools that will be helpful in text summarization. There are many challenging issues in implementing the CW framework. Large amount of propositions contained in a text can easily prevent the framework of computing with words from being of practical use. Another major barrier has been the vast amount of linguistic and world knowledge needed in natural language understanding, as well as the flexibility and context dependence in the way the knowledge are applied. Summarization calls for a combination of shallow and deep analysis methods. Good enough and working solutions for text summarization will probably be found in the middle between an elegant model but infeasible computationally, and a more crude techniques and computationally effective solutions.

Acknowledgments. Financial support from Academy of Finland is gratefully acknowledged (Grant Number 111692).

References

1. Allen J.: Natural Language Understanding, Pearson Addison-Wesley (1987)
2. Kintsch W. and van Dijk T A.: Toward a model of text comprehension and production, Psychol. Review 85 (1978)363-394
3. Mani I. and Maybury M. T. (Eds.): Advances in Automatic Text Summarization, MIT Press. Cambridge, MA. (1999)
4. Moens M. and Szpakowicz S. (Eds.): Text Summarization Braches Out, Proceedings of the ACL-05 Workshop, Barcelona, Spain, July 25-26 (2004)
5. Radev D., Jing H., Stys M. and Tam D.: Centroid-based summarization of multiple documents. Information Processing and Management 40 (2004) 919-938.
6. Russell & Norvig : Artificial Intelligence: A Modern Approach, Prentice Hall (1995)
7. van Dijk, T. A. and Kintsch, W.: Strategies of discourse comprehension. New York: Academic Press (1983)
8. Zadeh L A.: PRUF-a meaning representation language for natural languages, Int. J. Man-Machine Studies 10, 395-460, 1978.
9. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages, Computers and Mathematics 9 (1983) 149-184

10. Zadeh L.A.: Knowledge representation in fuzzy logic, IEEE Trans. on Knowledge and Data Engineering 1 (1989) 89-100
11. Zadeh L A.: From Computing with Numbers to Computing with Words -- From Manipulation of Measurements to Manipulation of Perceptions, IEEE Transactions on Circuits and Systems 45 (1999) 105-119
12. Zadeh L A.: Toward a generalized theory of uncertainty (GTU)--an outline, Information Sciences 172 (2005) 1-40

Concept-Based Questionnaire System

Masoud Nikravesh

EECS Department, University of California, Berkeley
Nikravesh@eecs.berkeley.edu

Abstract. In this paper, we go beyond the traditional web search engines that are based on keyword search and the Semantic Web which provides a common framework that allows data to be shared and reused across application,. For this reason, our view is that "Before one can use the power of web search the relevant information has to be mined through the concept-based search mechanism and logical reasoning with capability to Q&A representation rather than simple keyword search". In this paper, we will focus on development of a framework for reasoning and deduction in the web. A new web search model will be presented. One of the main core ideas that we will use to extend our technique is to change terms-documents-concepts (TDC) matrix into a rule-based and graph-based representation. This will allow us to evolve the traditional search engine (keyword-based search) into a concept-based search and then into Q&A model. Given TDC, we will transform each document into a rule-based model including it's equivalent graph model. Once the TDC matrix has been transformed into maximally compact concept based on graph representation and rules based on possibilistic relational universal fuzzy--type II (pertaining to composition), one can use Z(n)-compact algorithm and transform the TDC into a decision-tree and hierarchical graph that will represents a Q&A model. Finally, the concept of semantic equivalence and semantic entailment based on possibilistic relational universal fuzzy will be used as a basis for question-answering (Q&A) and inference from fuzzy premises. This will provide a foundation for approximate reasoning, language for representation of imprecise knowledge, a meaning representation language for natural languages, precisiation of fuzzy propositions expressed in a natural language, and as a tool for Precisiated Natural Language (PNL) and precisation of meaning. The maximally compact documents based on Z(n)-compact algorithm and possibilistic relational universal fuzzy--type II will be used to cluster the documents based on concept-based query-based search criteria.

1 Introduction

Semantic Web is a mesh or network of information that are linked up in such a way that can be accessible and be processed by machines easily, on a global scale. One can think of Semantic Web as being an efficient way of representing and sharing data on the World Wide Web, or as a globally linked database. It is important to mention that Semantic Web technologies are still very much in their infancies and there seems to be little consensus about the likely characteristics of such system. It is also important to keep in mind that the data that is generally hidden in one way or other is often useful in some contexts, but not in others. It is also difficult to use on a large scale such information, because there is no global system for publishing data in such a way as it can be easily processed by anyone. For example, one can think of information about local hotels, sports events, car or home sales info, insurance data, weather information

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 161–170, 2007.
springerlink.com

stock market data, subway or plane times, Major League Baseball or Football statistics, and television guides, etc.. All these information are presented by numerous web sites in HTML format. Therefore, it is difficult to use such data/information in a way that one might wanted to do so. To build any semantic web-based system, it will become necessary to construct a powerful logical language for making inferences and reasoning such that the system to become expressive enough to help users in a wide range of situations. This paper will try to develop a framework to address this issue, Concept-Based Questionnaire Systems. In the next sections, we will describe the components of such system.

2 Mining Information and Questionnaire Systems

The central tasks for the most of the search engines can be summarize as 1) query or user information request- do what I mean and not what I say!, 2) model for the Internet, Web representation-web page collection, documents, text, images, music, etc, and 3) ranking or matching function-degree of relevance, recall, precision, similarity, etc. One can use clarification dialog, user profile, context, and ontology, into an integrated frame work to design a more intelligent search engine. The model will be used for intelligent information and knowledge retrieval through conceptual matching of text. The selected query doesn't need to match the decision criteria exactly, which gives the system a more human-like behavior. The model can also be used for constructing ontology or terms related to the context of search or query to resolve the ambiguity. The new model can execute conceptual matching dealing with context-dependent word ambiguity and produce results in a format that permits the user to interact dynamically to customize and personalized its search strategy. It is also possible to automate ontology generation and document indexing using the terms similarity based on Conceptual-Latent Semantic Indexing Technique (CLSI). Often time it is hard to find the "right" term and even in some cases the term does not exist. The ontology is automatically constructed from text document collection and can be used for query refinement. It is also possible to generate conceptual documents similarity

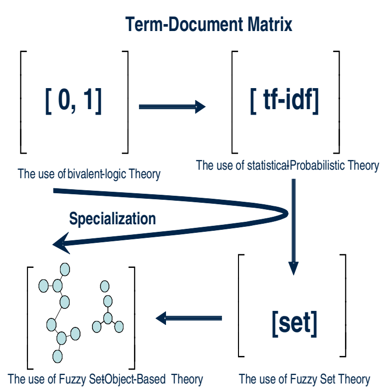


Fig. 1. Evolution of Term-Document Matrix representation

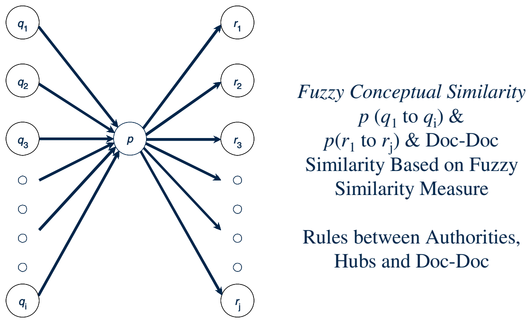


Fig. 2. Fuzzy Conceptual Similarity

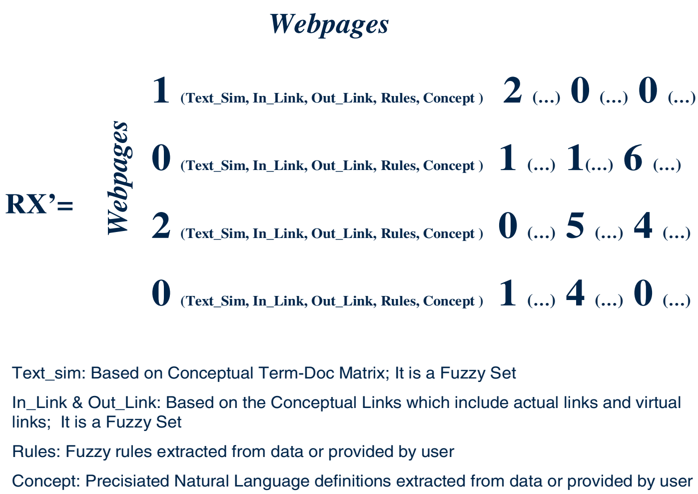


Fig. 3. Matrix representation of Fuzzy Conceptual Similarity model

map that can be used for intelligent search engine based on CLSI, personalization and user profiling.

The user profile is automatically constructed from text document collection and can be used for query refinement and provide suggestions and for ranking the information based on pre-existence user profile. Given the ambiguity and imprecision of the "concept" in the Internet, which may be described by both textual and image information, the use of Fuzzy Conceptual Matching (FCM) is a necessity for search engines.

In the FCM approach (Figures 1 through 4), the "concept" is defined by a series of keywords with different weights depending on the importance of each keyword. Ambiguity in concepts can be defined by a set of imprecise concepts. Each imprecise concept in fact can be defined by a set of fuzzy concepts. The fuzzy concepts can then be related to a set of imprecise words given the context. Imprecise words can then be translated into precise words given the ontology and ambiguity resolution through clarification dialog. By constructing the ontology and fine-tuning the strength of links (weights), we could construct a fuzzy set to integrate piecewise the imprecise concepts and precise words to define the ambiguous concept. To develop FCM, a series of new tools are needed. The new tools are based on fuzzy-logic-based method of computing with words and perceptions (CWP [1-4]), with the understanding that perceptions are described in a natural language [5-9] and state-of-the-art computational intelligence techniques [10-15]. Figure 1 shows the evolution of Term-Document Matrix representation. The [0,1] representation of term-document matrix (or in general, storing the document based on the keywords) is the simplest representation. Most of the current search engines such as Google™, Teoma, Yahoo!, and MSN use this technique to store the term-document information. One can extend this model by the use of ontology and other similarity measures. This is the core idea that we will use to extend this technique to FCM. In this case, the existence of any keyword that does not exist directly in the document will be decided through the connection weight to other keywords that exist in this document. For example consider the followings:

• if the connection weight (based on the ontology) between term “i” (i.e. Automobile) and term “j” (i.e. Transportation) is 0.7; and the connection weight between term “j” (i.e. Transportation) and term “k” (i.e. City) is 0.3; wij=0.7, and wjk=0.3, and term “i” doesn’t exist in document “I”, term “j” exists in document “I”, and term “k” does not exist in document I

TiDI=0, TiDI=1, TkDI=0

Given the above observations and a threshold of 0.5, one can modify the term-document matrix as follows:

TiDI’=1, TiDI’=1, TkDI’=0

In general one can use a simple statistical model such as the concurrence matrix to calculate wij [5, 12].

An alternative to the use of [1,0] representation is the use of the tf-idf (term frequency-inverse document frequency) model. In this case, each term gets a weight given its frequency in individual documents (tf, frequency of term in each document) and its frequency in all documents (idf, inverse document frequency). There are many ways to create a tf-idf matrix [7]. In FCM, our main focus is the use of fuzzy set theory to find the association between terms and documents. Given such association, one can represent each entry in the term-document matrix with a set rather then either [0,1] or tf-idf. The use of fuzzy-tf-idf is an alternative to the use of the conventional tf-idf. In this case, the original tf-idf weighting values will be replaced by a fuzzy set representing the original crisp value of that specific term. To construct such value, both ontology and similarity measure can be used. To develop ontology and similarity, one can use the conventional Latent Semantic Indexing (LSI) or Fuzzy-LSI [7]. Given this concept (FCM), one can also modify the link analysis (Figure 2) and in general Webpage-Webpage similarly (Figure 3). More information about this project and also a Java version of Fuzzy Search Tool (FST) that uses the FCM model is

Table 1. Terms-Documents-Concepts (TDC) Matrix

Terms-Documents-Concepts					
Documents	key_1	key_2	key_3		Concepts
D_1	k_1^1	k_1^2	k_1^3		c_1
D_2	k_1^1	k_2^2	k_1^3		c_1
D_3	k_2^1	k_2^2	k_1^3		c_1
D_4	k_3^1	k_2^2	k_1^3		c_1
D_5	k_3^1	k_1^2	k_2^3		c_1
D_6	k_1^1	k_2^2	k_2^3		c_1
D_7	k_2^1	k_2^2	k_2^3		c_1
D_8	k_3^1	k_2^2	k_2^3		c_1
D_9	k_3^1	k_1^2	k_1^3		c_2
D_{10}	k_1^1	k_1^2	k_2^3		c_2
D_{11}	k_2^1	k_1^2	k_1^3		c_2
D_{12}	k_2^1	k_1^2	k_2^3		c_2

Table 2. Intermediate results for Z(n)- compact algorithm

The Intermediate Results/ Iterations					
Documents	key_1	key_2	key_3		C
D_1	k_1^1	k_1^2	k_1^3		c_1
D_2	k_1^1	k_2^2	k_1^3		c_1
D_3	k_2^1	k_2^2	k_1^3		c_1
D_4	k_3^1	k_2^2	k_1^3		c_1
D_5	k_3^1	k_1^2	k_2^3		c_1
D_6	k_1^1	k_2^2	k_2^3		c_1
D_7	k_2^1	k_2^2	k_2^3		c_1
D_8	k_3^1	k_2^2	k_2^3		c_1
D_2,D_3, D_4	*	k_2^2	k_1^3		c_1
D_6,D_7, D_8	*	k_2^2	k_2^3		c_1
D_5,D_8	k_3^1	*	k_2^3		c_1
D_2,D_6	k_1^1	k_2^2	*		c_1
D_3,D_7	k_2^1	k_2^2	*		c_1
D_4,D_8	k_3^1	k_2^2	*		c_1
D_2,D_6	*	k_2^2	*		c_1
D_2,D_3, D_4,D_6,D_7, D_8					
D_1	k_1^1	k_1^2	k_1^3		c_1
D_5,D_8	k_3^1	*	k_2^3		c_1
D_2,D_3, D_4,D_6,D_7, D_8	*	k_2^2	*		c_1

available at: http://www.cs.berkeley.edu/~nikraves/fst//SFT and a series of papers by the author at Nikravesh, Zadeh and Kacprzyk [12].

Currently, we are extending our work to use the graph theory to represent the term-document matrix instead of the use of fuzzy set. While, each step increases the complexity and the cost to develop the FCM model, we believe this will increase the performance of the model given our understanding based on the results that we have analyzed so far. Therefore, our target is to develop a more specialized and personalized model with better performance rather than a general, less personalized model with less accuracy. In this case, the cost and complexity will be justified.

One of the main core ideas that we will use to extend our technique is to change terms-documents-concepts (TDC) matrix into a rule and graph. In the following section, we will illustrate how one can build such a model.

2.1 Questionnaire Systems

Consider a terms-documents-concepts (TDC) matrix presented as in Table 1 where (Note that TDC entries (kiijj) could be crisp number, tf-idf values, set or fuzzy-objects (including the linguistic labels of fuzzy granular) as shown in Figure 1) .

Di : Documents; where i = 1… m (in this example 12)
Keyj : Terms/Keywords in documents; where j = 1… n (in this example 3)
Cij : Concepts; where ij = 1…l (in this example 2)

Table 3. Maximally Z(n)-compact representation of TDC matrix

Documents	key_1	key_2		key_n	Concepts
D_1	k_1^1	k_1^2		k_1^3	c_1
D_5,D_8	k_3^1	*		k_2^3	c_1
D_2,D_3, D_4,D_6,D_7, D_8	*	k_2^2		*	c_1
D_9	k_3^1	k_1^2		k_1^3	c_2
D_{10}	k_1^1	k_1^2		k_2^3	c_2
D_{11}, D_{12}	k_2^1	k_1^2		*	c_2

Table 4. Rule-based representation of Z(n)-compact of TDC matrix

Documents	Rules
D_1	If key_1 is k_1^1 and key_2 is k_1^2 and key_3 is k_1^3 THEN Concept is c_1
D_5,D_8	If key_1 is k_3^1 and key_3 is k_2^3 THEN Concept is c_1
D_2,D_3, D_4,D_6,D_7, D_8	If key_2 is k_2^2 THEN Concept is c_1
D_9	If key_1 is k_3^1 and key_2 is k_1^2 and key_3 is k_1^3 THEN Concept is c_2
D_{10}	If key_1 is k_1^1 and key_2 is k_1^2 and key_3 is k_2^3 THEN Concept is c_2
D_{11}, D_{12}	If key_1 is k_2^1 and key_2 is k_1^2 THEN Concept is c_2

One can use Z(n)-compact algorithm to represent the TDC matrix with rule-base model. Table 2 shows the intermediate results based on Z(n)-compact algorithm for concept 1. Table 3 shows how the original TDC (Table 1) matrix is represented in final pass with a maximally compact representation. Once the TDC matrix represented by a maximally compact representation (Table 3), one can translate this compact representation with rules as presented in Tables 4 and 5. Table 6 shows the Z(n)-compact algorithm. Z(n) Compact is the basis to create web-web similarly as shown in Figure 3.

As it has been proposed, the TDC entries could not be crisp numbers. The following cases would be possible:

A-- The basis for the kiijjs are [0 and 1]. This is the simplest case and Z(n)-compact will work as presented.

B-- The basis for the kiijjs are tf-idf or any similar statistical based values or GA-GP context-based tf-idf, ranked tf-idf or fuzzy-tf-idf. In this case, we use fuzzy granulation to granulate tf-idf into series of granular, two ([0 or 1] or [high and low]), three (i.e. low, medium, and high), etc. Then the Z(n)-compact will work as it is presented.

C-- The basis for the kiijjs are set value created based on ontology which can be created based on traditional statistical based methods, human made, or fuzzy-ontology. In this case, the first step is to find the similarities between set values using statistical or fuzzy similarly measures. BISC-DSS software has a set of similarity measures, T-norm and T-conorm, and aggregator operators for this purpose. The second step is to use fuzzy granulation to granulate the similarities values into series of granular, two ([0 or 1] or [high and low]), three (i.e. low, medium, and high), etc. Then the Z(n)-compact will work as it is presented.

D-- It is also important to note that concepts may also not be crisp. Therefore, steps B and C could also be used to granulate concepts as it is used to granulate the keyword entries values (kiijjs).

Table 5. Rule-based representation of Maximally Z(n)-compact of TDC matrix (Alternative representation for Table 4)

Document	Rules
D_1	If key_1 is k_1^1 and key_2 is k_1^2 and key_3 is k_1^3
	OR
D_5, D_8	If key_1 is k_3^1 and key_3 is k_2^3
	OR
$D_2, D_3, D_4, D_6, D_7, D_8$	If key_2 is k_2^2
	THEN Concept is c_1
D_9	If key_1 is k_3^1 and key_2 is k_1^2 and key_3 is k_1^3
	OR
D_{10}	If key_1 is k_1^1 and key_2 is k_1^2 and key_3 is k_2^3
	OR
D_{11}, D_{12}	If key_1 is k_2^1 and key_2 is k_1^2
	THEN Concept is c_2

Table 6. Z(n)-Compactification Algorithm

Z(n)-Compact Algorithm:

The following steps are performed successively for each column j; j=1 … n

1. Starting with k_{ii}^{jj} (ii=1, jj=1) check if for any k_{ii}^1 (ii=1, ..,3 in this case) all the columns are the same, then k_{ii}^1 can be replaced by *
   - For example, we can replace k_{ii}^1 (ii=1, …,3) with * in rows 2, 3, and 4 . One can also replace k_{ii}^1 with * in rows 6, 7, and 8. (**Table 1,** first pass).
2. Starting with k_{ii}^{jj} (ii=1, jj=2) check if for any k_{ii}^2 (ii=1, ..,3 in this case) all the columns are the same, then k_{ii}^2 can be replaced by *
   - For example, we can replace k_{ii}^2 (ii=1, …,3) with * in rows 5 and 8. (**Table 1,** first pass).
3. Repeat steps one and 2 for all jj.
4. Repeat steps 1 through 3 on new rows created Row* (Pass 1 to Pass nn, in this case, Pass 1 to Pass 3).
   - For example, on Rows* 2,3, 4 (Pass 1), check if any of the rows given columns jj can be replaced by *. In this case, k_{ii}^3 can be replaced by *. This will gives: * k_2^2 *.
   - For example, on Pass 3, check if any of the rows given columns jj can be replaced by *. In this case, k_{ii}^1 can be replaced by *. This will gives: * k_2^2 *
5. Repeat steps 1 through 4 until no compactification would be possible

E-- Another important case is how to select the right keywords in first place. One can use traditional statistical or probabilistic techniques which are based on tf-idf techniques, non traditional techniques such as GA-GP context-based tf-idf, ranked tf-idf or fuzzy-tf-idf, or clustering techniques. These techniques will be used as first pass to select the first set of initial keywords. The second step will be based on feature selection technique based on to maximally separating the concepts. This techniques are currently part of the BISC-DSS toolbox, which includes the Z(n)-Compact-Feature-Selection technique (Z(n)-FCS)).

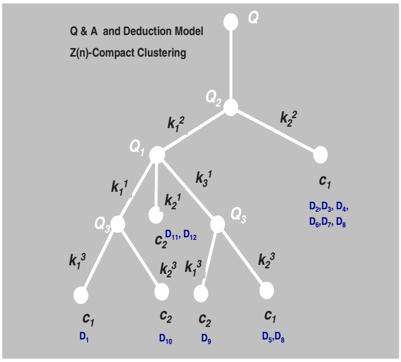


Fig. 4. a. Q & A model of TDC matrix and maximally Z(n)-compact rules and concept-based query-based clustering

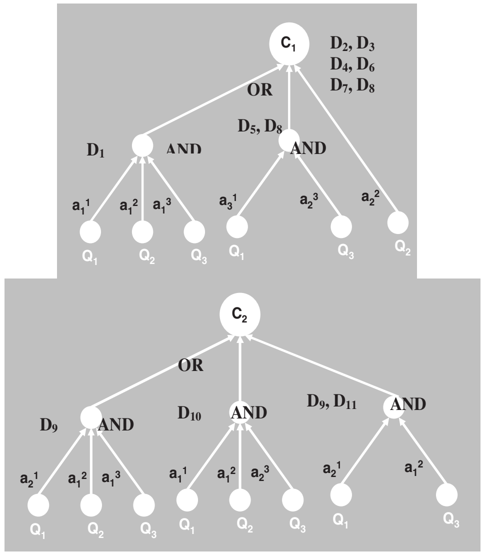


Fig. 4. b. Tree model of TDC matrix for Classes 1 and 2, where a_i^j in Figure 4 b is ideally equal to k_i^j in Figure 4 a

F-- Other possible cases: These include when the keywords are represented based on a set of concepts (such as Chinese-DNA model) or concepts are presented as a set of keywords (traditional techniques). One can use Neu-FCS model to create concepts automatically and relate the keywords to the concepts through a mesh of networks of neurons.

G-- Another very important case is when the cases are not possibilistic and are in general form " IF Keyi isri Ki and … Then Class isrc Cj ; where isr can be presented as:

r: = — *equality constraint: X=R is abbreviation of X is=R*
r: $\leq$ — *inequality constraint:* $X \leq R$
r: $\subset$ — *subsethood constraint:* $X \subset R$
r: blank — *possibilistic constraint; X is R; R is the possibility distribution of X*
r: v — *veristic constraint; X isv R; R is the verity distribution of X*
r: p — *probabilistic constraint; X isp R; R is the probability distribution of X*
r: rs — *random set constraint; X isrs R; R is the set-valued probability distribution of X*
r: fg — *fuzzy graph constraint; X isfg R; X is function and R is its fuzzy graph*
r: u — *usuality constraint; X isu R means usually (X is R)*
r: g — *group constraint; X isg R means that R constrains the attribute-values of the group*

- *Primary constraints: possibilistic, probabilisitic and veristic*
- *Standard constraints: bivalent possibilistic, probabilistic and bivalent veristic*

H-- It is important to note that Keys also can be presented in from of grammar and linguistic, such as subjects 1 to n; verbs 1 to m, objects 1, l, etc. In this case, each sentence can be presented in form of "isr" form or "IF … Then " rules with "isr" or "is" format.

Once the TDC matrix has been transformed into maximally compact concept based on graph representation and rules based on possibilistic relational universal fuzzy--type I, II, III, and IV (pertaining to modification, composition, quantification, and qualification), one can use Z(n)-compact algorithm and transform the TDC into a decision-tree and hierarchical graph that will represents a Q&A model. Finally, the concept of semantic equivalence and semantic entailment based on possibilistic relational universal fuzzy will be used as a basis for question-answering (Q&A) and inference from fuzzy premises. This will provide a foundation for approximate reasoning, language for representation of imprecise knowledge, a meaning representation language for natural languages, precisiation of fuzzy propositions expressed in a natural language, and as a tool for Precisiated Natural Language (PNL) and precisation of meaning. The maximally compact documents based on Z(n)-compact algorithm and possibilistic relational universal fuzzy--type II will be used to cluster the documents based on concept-based query-based search criteria. For more information regarding technique based on possibilistic relational universal fuzzy--type I, II, III, and IV (pertaining to modification, composition, quantification, and qualification) refer to refrences [3-4 and 11-13]

Once the TDC matrix (Table 1) has been transformed into maximally Z(n)-compact representation (Table 3) and Rules (Tables 4 and 5), one can use Z(n)-compact algorithm and transform the TDC into a decision-tree/hierarchical graph which represent a Q&A model as shown in Figure 4.

4 Conclusions

Intelligent search engines with growing complexity and technological challenges are currently being developed. This requires new technology in terms of understanding, development, engineering design and visualization. While the technological expertise of each component becomes increasingly complex, there is a need for better integration of each component into a global model adequately capturing the imprecision and deduction capabilities. In addition, intelligent models can mine the Internet to conceptually match and rank homepages based on predefined linguistic formulations and rules defined by experts or based on a set of known homepages. The FCM model can be used as a framework for intelligent information and knowledge retrieval through conceptual matching of both text and images (here defined as "Concept"). The FCM can also be used for constructing fuzzy ontology or terms related to the context of the query and search to resolve the ambiguity. This model can be used to calculate conceptually the degree of match to the object or query.

Acknowledgements. Funding for this research was provided by the British Telecommunication (BT), the BISC Program of UC Berkeley, and Imaging and Informatics group at Lawrence Berkeley National Lab. The author would like to thanks Prof.

Zadeh for his feedback, comments and allowing the authors to use the published and his unpublished documents, papers, and presentations to prepare this paper.

References

1. L. A. Zadeh, From Computing with Numbers to Computing with Words -- From Manipulation of Measurements to Manipulation of Perceptions, IEEE Transactions on Circuits and Systems, 45, 105-119, 1999.
2. L. A. Zadeh, "A new direction in AI: Towards a Computational Theory of Perceptions," AI magazine, vol. 22, pp. 73--84, 2001.
3. L.A. Zadeh, Toward a Perception-based Theory of Probabilistic Reasoning with Imprecise Probabilities, Journal of Statistical Planning and Inference, 105 233–264, 2002.
4. L. A. Zadeh and M. Nikravesh, Perception-Based Intelligent Decision Systems; Office of Naval Research, Summer 2002 Program Review, Covel Commons, University of California, Los Angeles, July 30th-August 1st, 2002. (L.A. PRUF-a meaning representation language for natural languages, Int. J. Man-Machine Studies 10, 395-460, 1978.)
5. M. Nikravesh and B. Azvine; New Directions in Enhancing the Power of the Internet, Proc. Of the 2001 BISC Int. Workshop, University of California, Berkeley, Report: UCB/ERL M01/28, August 2001.
6. V. Loia , M. Nikravesh, L. A. Zadeh, Journal of Soft Computing, Special Issue; fuzzy Logic and the Internet, Springer Verlag, Vol. 6, No. 5; August 2002.
7. M. Nikravesh, et. al, "Enhancing the Power of the Internet", Volume 139, published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (2004).
8. M. Nikravesh, Fuzzy Logic and Internet: Perception Based Information Processing and Retrieval, Berkeley Initiative in Soft Computing, Report No. 2001-2-SI-BT, September 2001a.
9. M. Nikravesh, BISC and The New Millennium, Perception-based Information Processing, Berkeley Initiative in Soft Computing, Report No. 2001-1-SI, September 2001b.
10. M. Nikravesh, V. Loia,, and B. Azvine, Fuzzy logic and the Internet (FLINT), Internet, World Wide Web, and Search Engines, International Journal of Soft Computing-Special Issue in fuzzy logic and the Internet , 2002
11. M. Nikravesh, Fuzzy Conceptual-Based Search Engine using Conceptual Semantic Indexing, NAFIPS-FLINT 2002, June 27-29, New Orleans, LA, USA
12. M. Nikravesh and B. Azvin, Fuzzy Queries, Search, and Decision Support System, International Journal of Soft Computing-Special Issue in fuzzy logic and the Internet , 2002
13. M. Nikravesh, V. Loia, and B. Azvine, Fuzzy logic and the Internet (FLINT), Internet, World Wide Web, and Search Engines, International Journal of Soft Computing-Special Issue in fuzzy logic and the Internet, 2002
14. M. Nikravesh, Fuzzy Conceptual-Based Search Engine using Conceptual Semantic Indexing, NAFIPS-FLINT 2002, June 27-29, New Orleans, LA, USA
15. V. Loia, M. Nikravesh and Lotfi A. Zadeh, "Fuzzy Logic and the Internet", Volume 137, published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (2004)

A Hybrid Model for Document Clustering Based on a Fuzzy Approach of Synonymy and Polysemy

Francisco P. Romero[1], Andrés Soto[2], and José A. Olivas[1]

[1] Department of Information Technologies and Systems, University of Castilla La Mancha, Paseo de la Universidad 4, 13071 Ciudad Real
{Francisco.Romero, JoseAngel.Olivas}@uclm.es

[2] Department of Computer Science, Universidad Autónoma del Carmen, CP 24160, Ciudad del Carmen, Campeche, México
soto_andres@yahoo.com

Abstract. A new model for document clustering is proposed in order to manage with conceptual aspects. To measure the presence degree of a concept in a document (or even in a document collection), a concept frequency formula is introduced. This formula is based on new fuzzy formulas to calculate the synonymy and polysemy degrees between terms. To solve the several shortcomings of classical clustering algorithm a soft approach to hybrid model is proposed. The clustering procedure is implemented by two connected and tailored algorithms with the aim to build a fuzzy-hierarchical structure. A fuzzy hierarchical clustering algorithm is used to determine an initial clustering and the process is completed using an improved soft clustering algorithm. Experiments show that using this model, clustering tends to perform better than the classical approach.

Keywords: Web Information Retrieval, Fuzzy Set, synonymy, polysemy.

1 Introduction

Clustering could be considered as the unsupervised learning process of organizing objects into groups whose members are similar in some way, such that unobvious relations and structures can be revealed. Document clustering methods have been widely applied in Information Retrieval, IR, supported by the hypothesis that documents relevant to a given query should be more similar to each other than to irrelevant documents, so they would be clustered together [1].

There are many algorithms for automatic clustering such as partitioning algorithm and hierarchical clustering that can be applied to a set of vectors to form the clusters. Partitioning algorithms are famous for their simplicity and efficiency in large data sets clustering. However, these algorithms have one shortcoming: clustering results are heavily dependent on the user-defined variants [2], i.e., the selection of the initial centroid seeds. Another limitation of these algorithms comes from the necessity of specifying a priori the desired number of clusters to generate. In the document clustering task, the optimum number is not known a priori. Hierarchical clustering algorithms provide a representation with different levels of granularity, which are good for visualizing and browsing through document collections. The main shortcoming of these algorithms is their high complexity that increases with the number of documents [3].

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 171–179, 2007.
springerlink.com

The document representation model is very important to obtain good results in the clustering process. Traditionally, a document is considered as a bag of words, once the stop words have been removed. Term frequency is a statistical measure often used in IR to evaluate the relative importance of a word in a document. Term frequency is the number of times the word appears in a document divided by the total number of words in it. Each document is represented as a vector of term frequency values. The main problem of these approaches is that they only consider lexicographical aspects and do not consider the semantic relations between words [4].

In this paper a hybrid clustering algorithm to obtain good quality results without loss of effectiveness is presented. In order to evaluate conceptual aspects of documents, fuzzy measures of synonymy and polysemy are introduced, based on WordNet [5] synonym sets. Then, concept frequency is calculated as a statistical measure of the relative importance of a concept or meaning in a document. According to concept frequency the similarity degree between documents is estimated [6]. At last, a relevant terms selection process to create a meaningful representation for the clusters is presented. According to the experiments, this method provides a better representation and results than the classical approaches.

The rest of the paper is structured as follows: Section 2 will provide details of the representation method. Section 3 explains the clustering model. In Section 4 metrics and experiments are described and experimental results are discussed. Finally, in Section 5 conclusions and future direction of the work are outlined.

2 Fuzzy Models of Synonymy and Polysemy

From the ancient times a controversy has existed about how to consider synonymy: as a relation between expressions with identical or similar meaning. In this paper, synonymy will be understood as a gradual, fuzzy relation between terms as in [7], which is closer to its behavior in dictionaries, where it is possible to find cases in which equivalence relation do not hold. For example, auto and automobile share a common meaning: “a motor vehicle with four wheels; usually propelled by an internal combustion engine” [5]. But automobile has another meaning: as a verb, it means “to travel in an automobile”. Therefore, auto and automobile are not identical or equivalent, but similar.

The degree of synonymy between words is closely related with their polysemy level. Polysemy means the ambiguity of a word that can be used (in different contexts) to express two or more different meanings. Words with only one meaning are considered strong as long as words with several meanings are considered weak. In the above example, auto is a strong word, while automobile is weaker than auto, and car is even weaker because it has 5 meanings according to [5].

Let introduce some general definitions:

- V a set of terms which belongs to a particular dictionary
- M the set of meanings associated to the terms in V.
- Therefore, each term in V has one or more meanings in M and each meaning in M has one or more terms associated in V.
- $M(t)$ the set of different meanings associated with a certain term t.
- $N_m(t)$ the number of meanings associated with the term t
- $T(m)$ the set of terms that share a meaning m

Then a fuzzy relation *S* between two terms $t_1, t_2 \in V$ is introduced such that $S(t_1,t_2)$ express the degree of synonymy between both terms:

$$S(t_1, t_2) = \frac{|M(t_1) \cap M(t_2)|}{|M(t_1)|} \tag{1}$$

In this way, the degree of synonymy between **auto** and **automobile** will be 1, which means that the concept **auto** totally corresponds with the concept **automobile**. But, in the other way, the degree of synonymy between **automobile** and **auto** is just 0.5 because **automobile** correspond with **auto** just in half of the meanings.

In order to measure the weakness of a term an index $I_p(t)$ is defined to represent the polysemy degree of term *t* where $I_p(\text{auto})=0$, $I_p(\text{automobile})= 0.5$ and $I_p(\text{car}) = 0.8$.

$$I_p : V \rightarrow [0, 1]$$

where

$$I_p(t) = 1 - \frac{1}{N_m(t)} \ . \tag{2}$$

Furthermore, based on this degree of polysemy, concept frequency measure is introduced, such that $0 \leq Cf_j(m) \leq 1$, where n_{ij} is the number of occurrences of term t_i in the document D_j and n_{*j} is the number of terms in the same document.

$$Cf_j(m) = \frac{\sum_{t_i \in T(m)} \left(n_{ij} \cdot \left(1 - I_p(t_i)\right)\right)}{n_{*j}} = \frac{1}{n_{*j}} \cdot \sum_{t_i \in T(m)} \left(\frac{n_{ij}}{N_m(t_i)}\right) \ . \tag{3}$$

That way, it is easy to compare the relative importance of different meanings in a document D_j. This measure is equivalent to term frequency which is one of the most referenced in IR, but for meanings, which is not just lexicographic but include some semantic interpretation of the terms.

This way, it is easy to calculate the concept frequency for a meaning *m* for two documents D_1 and D_2 and to measure the degree of similarity between them.

By calculating $Cf_1(m)$ and $Cf_2(m)$ for all $m \in M$, two vectors Cf_1^M and Cf_2^M are obtained. For those vectors, using the widely used cosine similarity coefficient [8], the similarity relation between both documents could be defined:

$$similar^M(D_1, D_2) = \frac{\sum_{m \in M} \left(Cf_1(m) * Cf_2(m)\right)}{\sqrt{\sum_{m \in M} Cf_1(m)^2 * \sum_{m \in M} Cf_2(m)^2}} \ . \tag{4}$$

By this similarity relation, it is easy to organize documents in clusters

3 Clustering Algorithm

Using the clustering process the collection of documents will be split up in a reduced number of groups made up of documents with enough conceptual similarity. Each group will contain one or more relevant meanings which will make it different from the rest.

For our clustering algorithm documents are represented using an extended vector-space model using the fuzzy approach to synonymy and polysemy explain in the previous section. The similarity between two documents must be measured in some way if a clustering algorithm is to be used. The formula (4) provides a fuzzy way to calculate the similarity degree according with the representation method.

In this work, a hierarchical fuzzy clustering approach is presented. The clustering procedure is implemented by two connected and tailored algorithms.

With the aim of detecting the initial relationships between the documents to be able to later ascertain several conceptual document clusters, an agglomerative hierarchical clustering process will be carried out. Every document is initially assigned to its own cluster and then pairs of clusters are repeatedly merged until the whole tree is formed using the similarity function. A modified Repertory Grids like technique [9] is used with different aggregation, density and normalization functions to choose the cluster to merge in the agglomerative process.

The process of merging should finish when the documents are clustered into sets that correspond to different concepts. Therefore, the termination criterion will be a threshold on cohesion of the common meaning in the clusters, i.e. a threshold on the similarity measure between documents in a cluster.

In order to improve the quality of the hierarchical results, a specific method for hierarchy fuzzyfication is used. In this way the cluster cardinality will be corrected. The method described in [10] is used, with several changes to fit the method for finding hierarchical features.

This initial result is always the same and the number of initial clusters is optimal, but the obtained organization is not concluding. For the next clustering step we need a fast algorithm, which is able to deal with large datasets, and provide a reasonable accuracy. In this work, this result is refined using the SISC [11] clustering algorithm improved in FISS meta-searcher structure [12].

This modified algorithm is characterized by using the obtained clusters on the previous step, followed by an iterative process that moves each document in the clusters whose average similarity is greater than the threshold of similarity. This threshold is automatically calculated using the same density and normalization functions defined in the agglomerative step.

The algorithm also considers merging clusters and removing documents for clusters when their average similarities decrease under the threshold. In order to get a hierarchical structure, big clusters and the bag cluster (formed by the less similar documents) are reprocessed with the same method.

All these processes dynamically produce a fuzzy hierarchical structure of groups of "conceptually related" documents.

The resulting organization is hierarchical, so, from a large repository of documents we will obtain a tree folder organization. The resulting clusters can be considered as fuzzy sets, so each one of the retrieved documents has a membership degree (obtained from an average similarity degree) to each one of these clusters.

3.1 Describing the Document Clusters

The clustering algorithms normally do not use data labels for classification. However, labels can be used for evaluating the performance of partition result. Therefore, the analysis of clustering results and the representation of document clusters will be the last step. Then the problem of summarizing the cluster content arises, i.e., a textual description or a set of terms representing the cluster topics must be identified. Good clusters must have concise descriptions.

Each document cluster is represented using its relevant concept set. The relevant concept set is a collection of meanings with an important presence in all the cluster documents. Therefore, meanings should exceed a percentage threshold to be considered important enough to be included in that set. Then, they will be extracted with the algorithm modification explained in [13]. An example of cluster description using the most significant word of a concept explanation is shown in the Table 1.

Table 1. Example of cluster descriptions. The cluster description is conformed to the most relevant meanings. Each meaning has a most significant word.

Tag	Description
trade	dollar (13481061), market (01082610), inflation (13325078), policy (06567622), exchange (01078424)
money-supply	money-supply: loan (13226412), interest (13147070),cash (13214226)
income	price (05084251), economy (08252295), money (13212169) , account(06430339)

4 Evaluation

Evaluation of the hierarchy quality generated by a particular algorithm is an important and non-trivial task. This section describes the data sets used for performing our practical experiments. The experimental results are also presented. In this work, vectors of documents were created using the fuzzy model of synonymy and polysemy explain in the section 2. Then they were clustered using the Hybrid Model presented in this work. After the clustering process, the documents get labelled with their relevant terms, identifying in this way the document classes. Then comparing the actual classes with the found classes, the quality measures can be obtained.

4.1 Metrics

Some of the most relevant evaluation measures [14] will be used to compare and analyse the performance of clustering methods.

1. Measures of the document representation method:
    a. Mean Similarity (MS): Average of similarity of each element with the rest of the set.

   b. Number of Outliers (NO): An outlier is an object that is quite different from the majority of the objects in a collection.
2. Measures of the clustering results
   3. Internal quality measures that depend on the representation
      a. Cluster Self Similarity (CSS): the average similarity between the documents in a cluster
      b. Size of Noise Cluster (SNC): number of elements unclassified in the hierarchical structure.
   4. External quality measures based on a known categorization.
      a. F Measure [15]: combines the precision (p) and recall (r) values from IR [16][17].

The F measure of cluster j and class i is given by:

$$F(i,j) = \frac{2 * r_{ij} * p_{ij}}{r_{ij} + p_{ij}} \quad . \tag{5}$$

For an entire cluster hierarchy the F measure of any class is the maximum value obtained at any node in the tree. An overall value for the F measure is computed by taking the weighted average of all values for the F measure as follows, where n is the number of documents and the maximum is calculated over all clusters at all levels:

$$F = \sum \frac{n_i}{n} * \max\{F(i,j)\} \ . \tag{6}$$

4.2 Experiments

To evaluate the effectiveness of a classification and representation method, a test collection is needed. Therefore, we use, in this initial experiment, two different datasets which are widely used in other publications and reflect the conditions in a broad range of real life applications. These data sets are (experiment details are shown in Table 2):

- SMART: The SMART collections contains 1400 CRANFIELD documents from aeronautical systems papers, 1033 MEDLINE documents from medical journals and 1460 CISI documents from information retrieval papers (ftp://ftp.cs.cornell.edu/pub/smart)
- Reuters: This test data set consists of 21578 articles from the Reuters news service in the year 87 [19]). We used all documents from the specified categories according to the “ModApte” split.

Table 2. Technical Note of the Experiment

Property	SMART	REUTERS
Number of documents	3893	8654
Number of meanings	5068	12594
Number of ranked meanings	1062	1254

4.3 Results

The results obtained by this model are compared with those obtained by the classical methods, such as the tf-idf representation method [19] and the fuzzy c-means clustering algorithm [20].

The experimental results are shown in the Table 3, expressed in percent. In the first part of the table are grouped the results corresponding to metrics of type "higher is better". Then, in the second part, are grouped the results corresponding to metrics of type "lower is better".Fig.1 shows a graphic with the results.

Table 3. Experimental Results

	TF-IDF & FCM	**Hybrid Model**	**TF-IDF & FCM**	**Hybrid Model**
Metric	**SMART**	**SMART**	**REUTER**	**REUTER**
MS	37	49	29	45
CSS	24	55	22	43
F-measure	43	63	45	54
NO	22	10	25	15
SNC	15	8	28	10

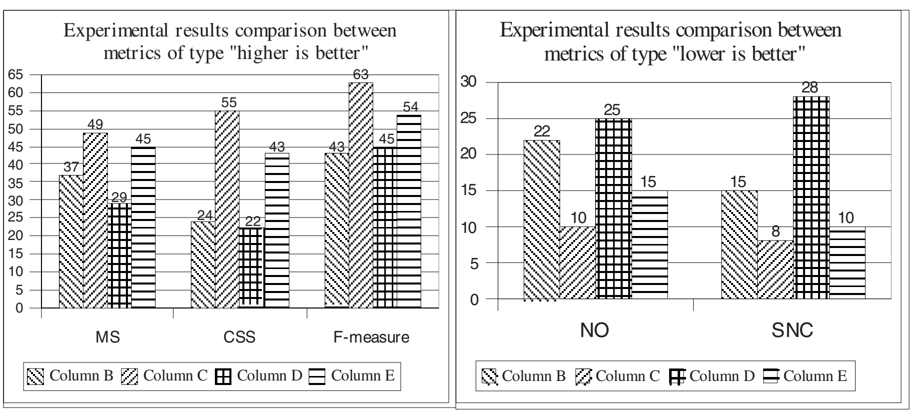


Fig. 1. Experimental results comparison

5 Conclusions and Future Work

In this work, a new method for document clustering is proposed. The method is based on clustering by taking into account word meanings in spite of clustering by term frequency. A new model to calculate the degree of synonymy between terms in order to manage with their meanings has been applied. Several new fuzzy formulas have been introduced to calculate the degree of word polysemy in order to manage with their meanings. Using these formulas, even though a certain term does not appear in a document, it is possible to estimate some degree of its presence according to the degree of synonymy shared with terms that do appear in the document.

By this formula, it is possible not only to measure the frequency of use of terms in documents, but to measure the frequency of use of meanings or concepts in documents. A coefficient was introduced in order to quantify the presence of a concept in a document, which was denoted concept frequency coefficient.

With the concept frequency coefficient, it is possible to measure how similar are two or more documents depending on their use of some concept. In this approach, this coefficient could also be used to document collection in relation with the use made of the concept by the different documents in the collection. We use this coefficient to perform a clustering algorithm with the aim to divide a set of documents into a specified number of clusters, so that documents within a cluster have high conceptual similarity.

The experiments demonstrate that the quality and effectiveness of clustering using this method is better than the usual tf-idf representation. The clustering hybrid model allows a good quality of results without loss of effectiveness. According to the experiments, this approach performs better than the methods based on the classical clustering algorithms.

At this moment this approach is capable of being improved in some aspects such as managing multilingual document collections and studying more mechanisms to determine and exploit the context.

In future, more experiments will be conducted with other collections and different methods of clustering. Further studies about the efficiency of the clustering algorithm are essential. All the steps in this approach might be replaced by other methods that digest similar input and produce similar output without changing our principal approach. We could use another algorithm such as Kohonen Self Organizing Maps (SOM) [21]. This is a neural network based technique, which takes the initial hierarchical results as input and build a fuzzy document map in which similar documents are mapped to the same or nearby neurons.

Acknowledgments. Partially supported by SCAIWEB PAC06-0059 project, JCCM, Spain.

References

1. Zamir, O., Etzioni, O.: Grouper: A dynamic clustering interface to web search results, Proceedings of the WWW8, (1999).
2. Spath, H.: Clustering Analysis Algorithms for Data Reduction and Classification of Objects, Ellis Horwood, Chichester, (1980).
3. Zhao, Y. Karypis, G.: Evaluation of hierarchical clustering algorithms for document datasets. In Proceedings of CIKM, ACM Press, (2002) 515–524.
4. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. Addison-Wesley, ACM Press, New York, (1999).

5. WordNet: An Electronic Lexical Database, Edited by Christiane Fellbaum, MIT Press, May (1998).
6. Soto, A.; Olivas J.A., Prieto M. E.: Fuzzy Approach of Synonymy and Polysemy for Information Retrieval, Proceedings International Symposium on Fuzzy and Rough Sets (ISFUROS '06) Santa Clara, Cuba (2006).
7. Fernandez, S., Grana, J., Sobrino, A.: A Spanish e-dictionary of synonyms as a fuzzy tool for information retrieval. In Actas de las I Jornadas de Tratamiento y Recuperación de Información (JOTRI 2002), León, Spain, September (2002).
8. Mendes, M.E.S., Sacks L.: A Scalable Hierarchical Fuzzy Clustering Algorithm for Text Mining, in Proc. of the 4th International Conference on Recent Advances in Soft Computing, RASC'2004, Nottingham, UK, (2004) 269-274.
9. El-Hamdouchi A., Willet P.: Comparison of Hierarchic Agglomerative Clustering Methods for Document Retrieval, The Computer Journal, Vol. 32, No. 3, (1989).
10. Akrivas G., Wallace M., Andreou G., Stamou G., Kollias S.: Context - Sensitive Semantic Query Expansion, Proceedings of the IEEE International Conference on Artificial Intelligence Systems (ICAIS), Divnomorskoe, Russia (2002).
11. King-ip L., Ravikumar K.: A similarity-based soft clustering algorithm for documents. Proc. of the Seventh Int. Conf. on Database Sys. for Advanced Applications (2001).
12. Olivas J.A., Garcés P., Romero F.P.: An application of the FIS-CRM model to the FISS metasearcher: Using fuzzy synonymy and fuzzy generality for representing concepts in documents. International Journal of Approximate Reasoning (Soft Computing in Recognition and Search) 34, (2003) 201- 219
13. Beil F., Ester M., Xu X.: Frequent Term-Based Clustering, Proceedings of the SIGKDD'02, Edmonton, Canada, (2002).
14. Y. Yang , An Evaluation of Statistical Approaches to Text Categorization, Journal of Information Retrieval, Vol 1, No. 1/2, (1999) 67-88.
15. Larsen B., Aone C.: Fast and Effective Text Mining Using Linear-time Document Clustering, KDD-99, San Diego, California, (1999).
16. Van Rijsbergen C. J.: Information Retrieval, Buttersworth, London, Second edition. (1989).
17. Kowalski G., Information Retrieval Systems – Theory and Implementation, Kluwer Academic Publishers, (1997).
18. Lewis, D.: Reuters-21578 text categorization text collection 1.0. http://www.research.att.com/~lewis.
19. Salton, G. McGill, M. J.: Introduction to Modern Information Retrieval. McGraw-Hill, (1983).
20. Pedrycz, W.: Conditional Fuzzy C-Means, Pattern Recognition Letters, Vol.17, (1996) 625-631.
21. Kohonen, T.: Self-organizing Maps, Series in Information Sciences, vol. 30, Springer, (1995).

Morphic Computing

Germano Resconi[1] and Masoud Nikravesh[2]

[1] Catholic University, Brescia , Italy
resconi@numerica.it
[2] University of California, Berkeley, CA 94720, US
Nikravesh@eecs.berkeley.edu

Abstract. In this paper, we introduce a new type of computation called "*Morphic Computing*". *Morphic Computing* is based on *Field Theory* and more specifically *Morphic Field*s. *Morphic Fields* were first introduced by Rupert Sheldrake [1981] from his hypothesis of formative causation that made use of the older notion of *Morphogenetic Fields*. In this paper, we introduce the basis for this new computing paradigm.

Keywords: Morphic Computing, Morphogenetic Computing, Morphic Fields; Morphogenetic Fields, Quantum Computing, DNA Computing.

1 Introduction

Inspired by the work of the French philosopher Henri Bergson, Rupert Sheldrake [1981] developed his famous theory of Morphic Resonance. Rupert Sheldrake's [1981] theory of *Morphic Fields* is based on his hypothesis of formative causation that makes use of the older notion of *Morphogenetic Fields*. *Morphogenetic Fields* is a hypothetical biological fields and it has been used by environmental biologists since 1920's which deals with living things. However, *Morphic Fields* are more general than *Morphogenetic Fields* and are defined as universal information for both organic (living things) and abstract forms [Sheldrake 1988]. We claim that Morphic Fields reshape multidimensional space to generate local contexts. For example, the gravitational fields in general relativity are the *Morphic Fields* that reshape space-time space to generate local context where particles move. Morphic Computing reverses the ordinary N input basis fields of possible data to one field in the output system. The set of input fields form an N dimension space or context. The N dimensional space can be obtained by a deformation of an Euclidean space and we argue that the *Morphic Fields* is the cause of the deformation. In line with Rupert Sheldrake [1981] our *Morphic Fields* is the formative causation of the context.

Now the output field of data is the input field X in *Morphic Computing*. To compute the coherence of the X with the context, we project X into the context. Our computation is similar to a quantum measure where the aim is to find coherence between the behaviour of the particles and the instruments. So the quantum measure projects the physical phenomena into the instrument as a context. The logic of our projection is the same as quantum logic in the quantum measure. In conclusion, we compute how a context can implement desired results based on *Morphic Computing*.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 180–189, 2007.
springerlink.com

Our new computation paradigm – *Morphic Computing*- is based on *Field Theory* and more specifically *Morphic Fields.* Resconi and Nikravesh [2007 a and b] claim that *Morphic Computing* is a natural extension of Holographic Computation, Quantum Computation.. In this paper, we will introduce the basis for our new computing paradigm – *Morphic Computing based on Field Theory*.

2 Morphic Computing and Field Theory: Basic Concepts

In classical physics, we represent the interaction among the particles by local forces that are the cause of the movement of the particles. Also in classical physics, it is more important to know at any moment the individual values of the forces than the structure of the forces. This approach considers that the particles are independent from the other particles under the effect of the external forces. But with further development of the theory of particle physics, the researchers discovered that forces are produced by intermediate entities that are not located in one particular point of space but are at any point of a specific space at the same time. These entities are called "*Fields*". Based on this new theory, the structure of the fields is more important than the value itself at any point. In this representation of the universe, any particle in any position is under the effect of the fields. Therefore, the fields are used to connect all the particles of the universe in one global entity. However, if any particle is under the effect of the other particles, every local invariant property will disappear because every system is open and it is not possible to close any local system. To solve this invariance problem, scientist discovered that the local invariant can be conserved with a deformation of the local geometry and the metric of the space. While the form of the invariant does not change for the field, the action is changed. However, these changes are only in reference to how we write the invariance. In conclusion, we can assume that the action of the fields can be substituted with deformation of the space. Any particle is not under the action of the fields, the invariance as energy, momentum, etc. is true (physical symmetry). However, the references that we have chosen change in space and time and in a way to simulate the action of the field. In this case, all the reference space has been changed and the action of the field is only a virtual phenomena. In this case, we have a different reference space whose geometry in general is non Euclidean. With the quantum phenomena, the problem becomes more complex because the particles are correlated one with one other in a more hidden way without any physical interaction with fields. This correlation or entanglement generates a structure inside the universe for which the probability to detect a particle is a virtual or conceptual field that covers the entire Universe.

Morphic Computing is based on the following concepts:

1. The concept of field in the reference space
2. The fields as points or vectors in the N dimension Euclidean space of the objects (points)
3. A set of $M \leq N$ basis fields in the N dimensional space. The set of M fields are vectors in the N dimensional space. The set of M vectors form a non Euclidean subspace H (context) of the space N. The coordinates S^{α} in M of the field X are the

contro-variant components of the field X. The components of X in M are also the intensity of the sources of the basis field. The superposition of the basis field with different intensity give us the projection Q of X or Y = QX into the space H When M < N the projection operator of X into H define a constrain or relation among the components of Y.
4. With the tensor calculus with the components S^{α} of the vector X or the components of more complex entity as tensors , we can generate invariants for any unitary transformation of the object space or the change of the basis fields.
5. Given two projection operators Q_1 , Q_2 on two spaces H_1 , H_2 with dimension M_1 and M_2 we can generate the $M = M_1 M_2$, with the product of Y_1 and Y_2 or $Y = Y_1 Y_2$. Any projection Q into the space H or Y = QX of the product of the basis fields generate Y. When $Y \neq Y_1 Y_2$ the output Y is in entanglement state and cannot separate in the two projections Q_1 and Q_2 .
6. The logic of the Morphic Computing Entity is the logic of the projection operators that is isomorphic to the quantum logic

The information can be coded inside the basis fields by the relation among the basis fields. In *Morphic Computing,* the relation is represented by a non Euclidean geometry which metric or expression of the distance between two points shows this relation. The projection operator is similar to the measure in quantum mechanics. The projection operator can introduce constrains inside the components of Y. The sources are the instrument to control the image Y in *Morphic Computing*. There is a strong analogy between *Morphic Computing* and computation by holography and computation by secondary sources [Jessel 1973] in the physical field. The computation of Y by X and the projection operator Q that project X into the space H give his result when the Y is similar to X. In this case, the sources S are the solution of the computation. We see the analogy with the neural network where the solution is to find the weights w_k at the synapse. In this paper, we show that the weights are sources in *Morphic Computing*. Now, it is possible to compose different projection operators in a network of *Morphic Systems*. It is obvious to consider this system as a *System of Systems*.

Any *Morphic Computation* is always context dependent where the context is H. The context H by the operator Q define a set of rules that are comparable with the rules implemented in a digital computer. So when we change the context with the same operations, we obtain different results. We can control the context in a way to obtain wanted results. When any projection operator of X or QX is denoted as a measure, in analogy with quantum mechanics, any projection operator depends on the previous projection operator. In the measure analogy, any measure depends on the previous measures. So any measure is dependent on the path of measures or projection operators that we realise before or through the history. So we can say that different projection operators are a story (See Roland Omnès [1994] in quantum mechanics stories).

The analogy of the measure also gives us another intuitive idea of *Morph Computing*. Any measure become a good measure when gives us an image Y of the real phenomena X that is similar, when the internal rules to X are not destroyed. In the measure process, the measure is a good measure. The same for *Morphic Computing*, the computation is a good computation when the projection operator does not destroy the internal relation of the field in input X. The analogy with the measure in quantum mechanics is also useful to explain the concept of *Morphic Computing* because the instrument in the quantum measure is the fundamental context that interferes with the physical phenomena as H interferes with the input field X. A deeper connection exists between the Projection operator lattice that represents the quantum logic and *Morphic Computing* processes (see Eddie Oshins 1992).

3 Reference Space, Space of the Objects, Space of the Fields in the Morphic Computing

Given the n dimensional reference space $(R_1, R_2, \ldots, R_n)$, any point

$$P = (R_1, R_2, \ldots, R_n) \text{ is an object.}$$

Now we create the space of the objects which dimension is equal to the number of the points and the value of the coordinates in this space is equal to the value of the field in the point. We call the space of the points "space of the objects".

Any field connect all the points in the reference space and is represented as a point in the object space. The components of this vector are the value of the field in the different points. We know that each of the two points connected by a link assume the value of one of the connections. All the other points assume zero value. Now any value of the field in a point can be considered as a degree of connection of this point with all the others. Therefore, in one point where the field is zero, we can consider this point as non-connected to the others. In fact, because the field in this point is zero the other points cannot be connected by the field to the given point. In conclusion, we consider the field as a global connector of the objects in the reference space. Now inside the space of the objects, we can locate any type of field as vectors or points. In field theory, we assume that any complex field can be considered as a superposition of prototype fields whose model is well known.

The prototype fields are vectors in the space of the objects that form a new reference or field space. In general, the field space is a non Euclidean space. In conclusion, any complex field Y can be written in this way

$$Y = S_1 H_1 (R_1, \ldots, R_n) + S_2 H_2 (R_1, \ldots, R_n) + \ldots +$$

$$S_n H_n (R_1, \ldots, R_n) = H(R)\, S \tag{1}$$

In equation (1), $H_1, H_2, \ldots, H_n$ are the basic fields or prototype fields and $S_1, S_2, \ldots, S_n$ are the weights or source values of the basic fields. We assume that any basic field is generated by a source. The intensity of the prototype fields is proportional to the intensity of the sources that generates the field itself.

3.1 Example of the Basic Field and Sources

In Figure 1, we show an example of two different basic *fields in a two dimensional reference space (x, y).* The general equation of the fields is

$$F(x, y) = S\ [e^{-h((x-x_0)^2+(y-y_0)^2)}] \tag{2}$$

the parameters of the field F_1 are S=1 h =2 and x_0 = -0.5 and y_0 = -0.5,
the parameters of the field F_2 are S=1 h =2 and x_0 = 0.5 and y_0 = 0.5

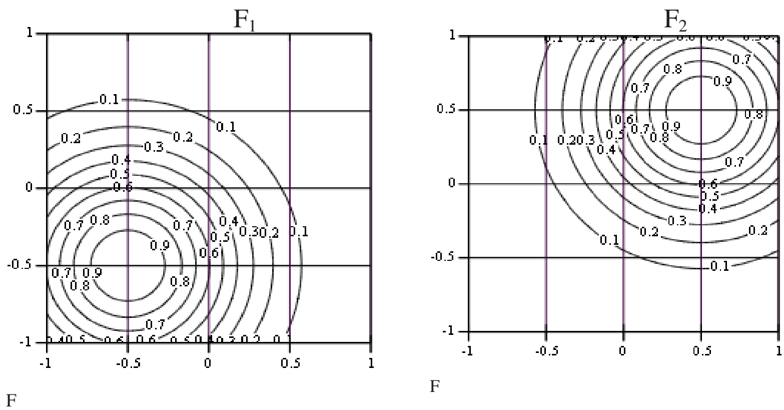


Fig. 1. Two different basic *fields* in the two dimensional reference space (x,y)

For the sources S_1 = 1 and S_2 = 1 the superposition field F that is shown in Figure 2 is F = F_1 + F_2. For the sources S_1 = 1 and S_2 = 2, the superposition field F that is shown again in Figure 2 is F = F_1 + 2 F_2 .

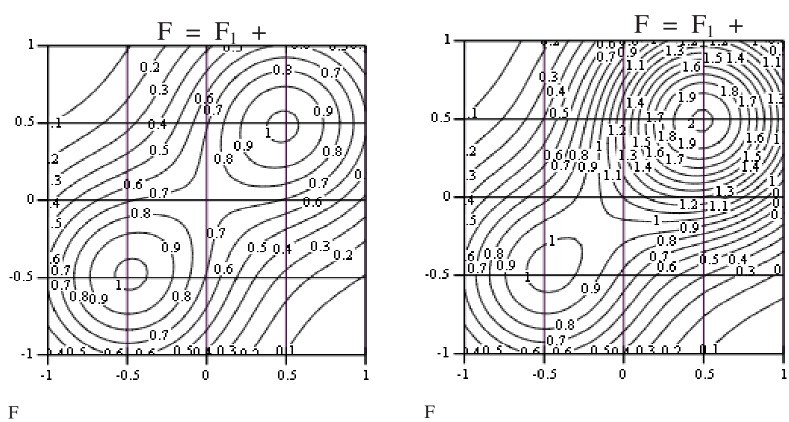


Fig. 2. Example of superposition of elementary fields F_1, F_2

3.2 Computation of the Sources

To compute the sources S_k, we represent the prototype field H_k and the input field X in a Table 1 where the objects are the points and the attribute are the fields.

Table 1. Fields values for M points in the reference space

	H_1	H_2	…	H_N	Input Field X
P_1	$H_{1,1}$	$F_{1,2}$	…	$H_{1,N}$	X_1
P_2	$H_{2,1}$	$H_{2,2}$	...	$H_{2,N}$	X_2
…	…	…	…	…	...
P_M	$H_{M,1}$	$H_{M,2}$	…	$H_{M,N}$	X_M

The values in Table 1 are represented by the following matrices

$$H = \begin{bmatrix} H_{1,1} & H_{1,2} & ... & H_{1,N} \\ H_{2,1} & H_{2,2} & ... & H_{2,N} \\ ... & ... & ... & ... \\ H_{M,1} & H_{M,2} & ... & H_{M,N} \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ ... \\ X_M \end{bmatrix}$$

The matrix H is the relation among the prototype fields F_k and the points P_h. At this point, we are interested in the computation of the sources S by which they give the best linear model of X by the elementary field values. Therefore, we have the *superposition* expression

$$Y = S_1 \begin{bmatrix} H_{1,1} \\ H_{2,1} \\ ... \\ H_{M,1} \end{bmatrix} + S_2 \begin{bmatrix} H_{1,2} \\ H_{2,2} \\ ... \\ H_{M,2} \end{bmatrix} + ... + S_n \begin{bmatrix} H_{1,n} \\ H_{2,n} \\ ... \\ H_{M,n} \end{bmatrix} = HS \tag{3}$$

Then, we compute the best sources S in a way the difference $|Y - X|$ is the minimum distance for any possible choice of the set of sources. It is easy to show that the best sources are obtained by the expression

$$S = (H^T H)^{-1} H^T X \tag{4}$$

Given the previous discussion and field presentation, the ***elementary Morphic Computing*** element is given by the input-output system as shown in Figure 3.

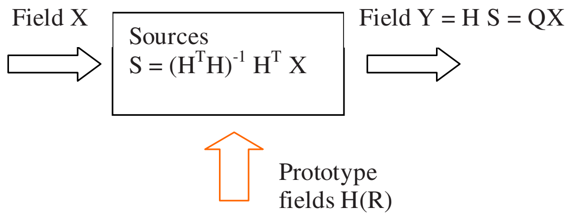


Fig. 3. Elementary Morphic Computing

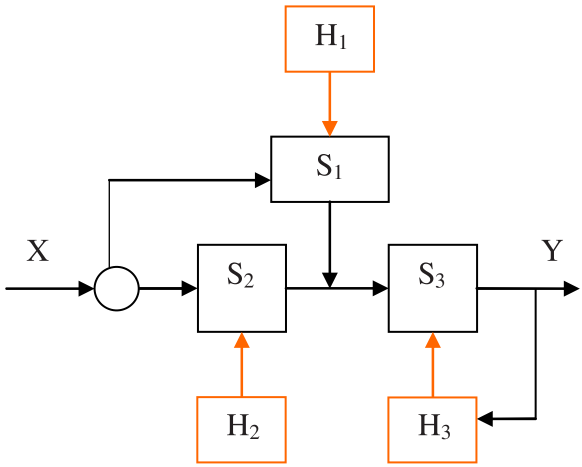


Fig. 4. Shows the Network of Morphic Computing

Figure 4 shows network of elementary Morphic Computing with three set of prototype fields and three type of sources with one general field X in input and one general field Y in output and intermediary fields from X and Y. When H is a square matrix, we have Y = X and

$$S = H^{-1} X \text{ and } Y = X = H\, S \tag{5}$$

Now for any elementary computation in the Morphic Computing, we have the following three fundamental spaces.

1. The reference space
2. The space of the objects (points)
3. The space of the prototype fields

Figure 5 shows a very simple geometric example when the number of the objects are three (P_1 , P_2 , P_3) and the number of the prototype fields are two (H_1 , H_2). The space which coordinates are the two fields is the space of the fields.

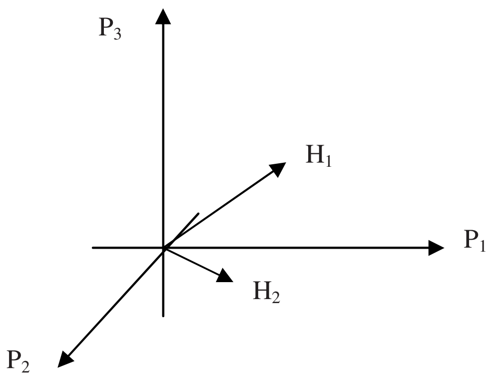


Fig. 5. The fields H_1 and H_2 are the space of the fields. The coordinates of the vectors H_1 and H_2 are the values of the fields in the three points P_1 , P_2 , P_3

Note that the output Y = H S is the projection of X into the space H

$Y = H (H^T H)^{-1} H^T X = Q X$

With the property $Q^2 X = Q X$

Therefore, the input X can be separated in two parts

$X = Q X + F$

where the vector F is perpendicular to the space H as we can see in a simple example given in Figure 6.

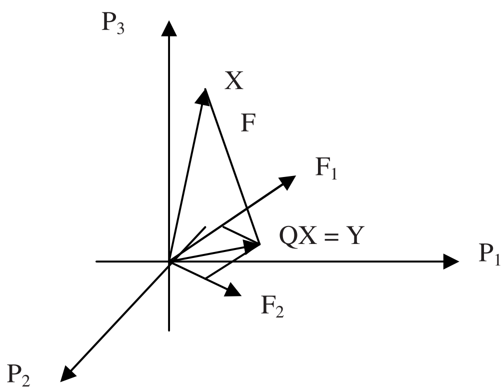


Fig. 6. Projection operator Q and output of the elementary Morphic Computing. We see that X = Q X + F , where the sum is the vector sum. Now, we try to extend the expression of the sources in the following way

Given $G(\Gamma) = \Gamma^T \Gamma$ and $G(H) = H^T H$ and

$S^* = [\, G(H) + G(\Gamma)\,]^{-1} H^T X$ and

$$[\,G(H) + G(\Gamma)\,]\,S^* = H^T X$$

So for $S^* = (H^T H)^{-1} H^T X + \Omega^\alpha = S^\alpha + \Omega^\alpha$

we have

$$([\,G(H) + G(\Gamma)\,]\,(S^\alpha + \Omega^\alpha) = H^T X$$

$$(H^T H)\,(H^T H)^{-1} H^T X + ([\,G(H) + G(\Gamma)\,]\,\Omega^\alpha = H^T X$$

$$G(\Gamma)\,S^\alpha + [\,G(H) + G(\Gamma)\,]\,\Omega^\alpha = 0$$

and

$$S^* = S + \Omega = \left(H^T H + \Gamma^T\ \Gamma\right)^{-1}\ H^T\ X$$

where for Ω is function of S by the equation

$$G(\Gamma)\,S + [\,G(H) + G(\Gamma)\,]\,\Omega = 0$$

For non-square matrix and/or singular matrix, we can use the generalized model given by Nikravesh [1994] as follows;

$$S^* = \left(H^T\ \Lambda^T \Lambda\, H\right)^{-1}\ H^T\ \Lambda^T \Lambda\, X = \left((\Lambda H)^T\ (\Lambda\, H)\right)^{-1}\ (\Lambda H)^T \Lambda\, X$$

Where we transform by Λ the input and the references H.

The value of the variable D (metric of the space of the field) is computed by the expression (6)

$$D^2 = (H\,S)^T (H\,S) = S^T H^T H\,S = S^T G\ S = (Q\,X)^T Q\,X \qquad (6)$$

For the unitary transformation U for which, we have $U^T U = I$ and H' = U H the prototype fields change in the following way

$$H' = U\,H$$

$$G' = (U\,H)^T (U\,H) = H^T U^T U\,H = H^T H$$

And

$$S' = [\,(U\,H)^T (U\,H)\,]^{-1} (U\,H)^T Z = G^{-1} H^T\ U^T Z = G^{-1} H^T (U^{-1} Z)$$

For $Z = U\,X$ we have S' = S and

the variable D is invariant. for the unitary transformation U .

We remark that $G = H^T H$ is a quadratic matrix that gives the metric tensor of the space of the fields. When G is a diagonal matrix the entire elementary field are independent one from the other. But when G has non diagonal elements, in this case *the elementary fields are dependent on from the other*. Among the elementary fields there is a *correlation* or a *relationship* and the geometry of the space of the fields is a non Euclidean geometry.

4 Conclusion

In this paper, we introduced a holistic interpretation of the computation denoted Morphic Computing. We showed the possibility of going beyond the traditional computation based on a step-by-step process. Our type of computation is similar to a quantum measure where the instrument is the context with proper rules. Input fields are introduced in the instrument as a context and write in internal parameters or sources of internal prototype fields. Morphic Computing projects the external fields to the internal context. In the projection process, the input is reshaped in a way to select part of the input that is coherent with the context. Because the projection operator uses the internal prototype fields and the computed sources, we can give a context dependent model of the external input. This is the meaning of the new computation.

References

1. M. Jessel , Acoustique Théorique , Masson et Cie Editours 1973
2. M. Nikravesh, Intelligent Computing Techniques for Complex systems, in *Soft Computing and Intelligent Data Analysis in Oil Exploration*, pp. 651-672, Elsevier, 2003.
3. R. Omnès (1994), The Interpretation of Quantum Mechanics , Princeton Series in Physics, 1994
4. Oshins, E., K.M. Ford, R.V. Rodriguez and F.D. Anger (1992). A comparative analysis: classical, fuzzy, and quantum logic. Presented at 2nd Florida Artificial Intelligence Research Symposium, St. Petersburg, Florida, April 5, 1989. In Fishman, M.B. (Ed.), Advances in artificial intelligence research, volume II, Greenwich, CT: JAI Press. *Most Innovative Paper Award, 1989, FLAIRS-89,* Florida AI Research Symposium.
5. G. Resconi and M. Nikravesh, Morphic Computing Part I and Part II, to be published in Forging New Frontiers, Series in Soft Computing and fuzziness, 2007a and 2007b.
6. G. Resconi and M. Nikravesh, Field Theory and Computing with Words, , FLINS 2006.
7. R. Sheldrake (1981), A New Science of Life: The Hypothesis of Morphic Resonance (1981, second edition 1985), Park Street Press; Reprint edition (March 1, 1995)
8. R. Sheldrake (1988), The Presence of the Past: Morphic Resonance and the Habits of Nature (1988), Park Street Press; Reprint edition (March 1, 1995)
9. P.Z. Wang and Sugeno M. (1982) , The factor fields and background structure for fuzzy subsets. Fuzzy Mathematics , Vol. 2 , pp.45-54

Morphic Computing: Web and Agents

Germano Resconi[1] and Masoud Nikravesh[2]

[1] Catholic University, Brescia, Italy
resconi@numerica.it
[2] University of California, Berkeley, CA 94720, US
Nikravesh@eecs.berkeley.edu

Abstract. *Morphic Computing* is based on *Field Theory* and more specifically *Morphic Field*s. In this paper, we introduce extensions of Morphic Computing to non classical fields and logic. *Morphic Computing* change or compute non physical conceptual fields. One example is in representing the semantics of words. In this paper, application to the field of computation by words as an example of the *Morphic Computing, Morphic Fields - concepts and Web search, and agents and fuzzy in Morphic Computing will be discussed.*

Keywords: Morphic Computing, Morphogenetic Computing, Morphic Fields; Morphogenetic, Soft Computing, Computing with Words.

1 Introduction

Morphic Fields and it's subset *Morphogenetic Fields* have been at the center of controversy for many years in mainstream science and the hypothesis is not accepted by some scientists who consider it a pseudoscience. Sheldrake defined *Morphic* and *Morphogenetic Fields* in his book, *The Presence of the Past* [1988] as follows:

> *"The term [Morphic Fields] is more general in its meaning than Morphogenetic Fields, and includes other kinds of organizing fields in addition to those of morphogenesis; the organizing fields of animal and human behaviour, of social and cultural systems, and of mental activity can all be regarded as Morphic Fields which contain an inherent memory." – Sheldrake [1988].*

Resconi and Nikravesh introduced a new type of computation called "Morphic Computing" [Resconi and Nikravesh, 2006 and 2007]. They assumed that computing is not always related to symbolic entities such as numbers, words or other symbolic entities. Fields as entities are more complex than any symbolic representation of the knowledge. For example, Morphic Fields include the universal database for both organic (living) and abstract (mental) forms. Morphic Computing can also change or compute non physical conceptual fields. The basis for *Morphic Computing* is *Field Theory* and more specifically *Morphic Field*s. *Morphic Fields* were first introduced by Rupert Sheldrake [1981] from his hypothesis of formative causation [Sheldrake, 1981 and 1988] that made use of the older notion of *Morphogenetic Fields*. Rupert

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 190–198, 2007.
springerlink.com

Sheldrake developed his famous theory, Morphic Resonance [Sheldrake 1981 and 1988], on the basis of the work by French philosopher Henri Bergson [1896, 1911]. Resconi and Nikravesh [2006 and 2007] claim that *Morphic Computing* is a natural extension of Optical Computation by holograms and holistic systems, Quantum Computation, Soft Computing [Zadeh 1991], and DNA Computing. All natural computation bonded by the Turing Machine such as classical logic and AI can be formalized and extended by our new type of computation model – *Morphic Computing*.

In this paper, we will first introduce the basis of Morphic Computing for non physical fields such conceptual fields, theory of generalized constraint for natural languages, and non classical logic for agents, Then *Morphic Computing*'s applications Then Morphic Computing's applications to the field of computation by words [Zadeh and Kacprzyk 1999a and 1999b, Zadeh and Nikravesh 2002will be given. Finally, we present *Morphic Fields* - concepts and Web search, and Agents and fuzzy [Zadeh, 1965] in *Morphic Computing*.

1.1 Morphic Computing and Conceptual Fields – Non Physical Fields

Morphic Computing change or compute non physical conceptual fields. One example is in representing the semantics of words. In this case, a field is generated by a word or a sentence as sources. For example, in a library the reference space would be where the documents are located. At any given word, we define the field as a map of the position of the documents in the library and the number of the occurrences (values) of the word in the document. The word or source is located in one point of the reference space (query) but the field (answer) can be located in any part of the reference. Complex strings of words (structured query) generate a complex field or complex answer by which the structure can be obtained by the superposition of the fields of the words as sources with different intensity. Any field is a vector in the space of the documents. A set of basic fields is a vector space and form a concept. We break the traditional idea that a concept is one word in the conceptual map. Internal structure (entanglement) of the concept is the relation of dependence among the basic fields. The ambiguous word is the source (query) of the fuzzy set (field or answer).

1.2 Morphic Computing and Natural Languages – Theory of Generalized Constraint

In a particular case, we know that a key assumption in computing with words is that the information which is conveyed by a proposition expressed in a natural language or word may be represented as a generalized constraint of the form "X isr R", where X is a constrained variable; R is a constraining relation; and r is an indexing variable whose value defines the way in which R constrains X. Thus, if p is a proposition expressed in a natural language, then "X isr R" representing the meaning of p, equivalently, the information conveyed by p. Therefore, the generalised constraint model can be represented by field theory in this way. The meaning of any natural proposition p is given by the space X of the fields that form a concept in the reference space or objective space, and by a field R in the same reference. We note that a concept is not only a word, but is a domain or context X where the propositions p represented by the field R is located. The word in the new image is not a passive entity but is an active entity. In fact, the word is the source of the field. We can also use the idea that the word as an abstract entity is a query and the field as set of instances of the query is the answer.

1.3 Morphic Computing and Agents – Non Classical Logic

In the agent image, where only one word (query) as a source is used for any agent, the field generated by the word (answer) is a Boolean field (the values in any points are true or false). Therefore, we can compose the words by logic operations to create complex Boolean expression or complex Boolean query. This query generates a Boolean field for any agent. This set of agents creates a set of elementary Boolean fields whose superposition is the fuzzy set represented by a field with fuzzy values. The field is the answer to the ambiguous structured query whose source is the complex expression p. The fields with fuzzy values for complex logic expression are coherent with traditional fuzzy logic with a more conceptual transparency because it is found on agents and Boolean logic structure. As points out [Nikravesh, 2006] the Web is a large unstructured and in many cases conflicting set of data. So in the World Wide Web, fuzzy logic and fuzzy sets are essential parts of a query and also for finding appropriate searches to obtain the relevant answer. For the agent interpretation of the fuzzy set, the net of the Web is structured as a set of conflicting and in many case irrational agents whose task is to create any concept. Agents produce actions to create answers for ambiguous words in the Web. A structured query in RDF can be represented as a graph of three elementary concepts as subject, predicate and complement in a conceptual map. Every word and relationship in the conceptual map are variables whose values are fields which their superposition gives the answer to the query. Because we are more interested in the meaning of the query than how we write the query itself, we are more interested in the field than how we produce the field by the query. In fact, different linguistic representations of the query can give the same field or answer. In the construction of the query, we use words as sources of fields with different intensities. With superposition, we obtain the answer for our structured query. We structure the text or query to build the described field or meaning. It is also possible to use the answer, as a field, to generate the intensity of the words as sources inside a structured query. The first process is denoted READ process by which we can read the answer (meaning) of the structured query. The second process is the WRITE process by which we give the intensity or rank of the words in a query when we know the answer. As an analogy to holography, the WRITE process is the construction of the hologram when we know the light field of the object. The READ is the construction of the light field image by the hologram. In the holography, the READ process uses a beam of coherent light as a laser to obtain the image. Now in our structured query, the words inside of text are activated at the same time. The words as sources are coherent in the construction by superposition of the desired answer or field. Now the field image of the computation by words in a crisp and fuzzy interpretation prepares the implementation of the *Morphic Computing* approach to the computation by words. In this way, we have presented an example of the meaning of the new type of computation, "*Morphic Computing*".

2 Field Theory, Concepts and Web Search

To search in the web, we use the term-document matrix to obtain the information retrieved. In Table 1, we show the data useful in obtaining the desired information in the Web.

Table 1. Term (word), document and complex text G

	$Word_1$	$Word_2$	…	$Word_N$	Concept X
$Document_1$	$K_{1,1}$	$K_{1,2}$	…	$K_{1,N}$	X_1
$Document_2$	$K_{2,1}$	$K_{2,2}$	...	$K_{2,N}$	X_2
…	…	…	…	…	...
$Document_M$	$K_{M,1}$	$K_{M,2}$	…	$K_{M,N}$	X_M

Where $K_{i,j}$ is the value of the $word_j$ in the $document_i$. The word in Table 1 is a source of a field which values are the values in the position space of the documents. Any document is one of the possible positions of the word. In a Web search, it is useful to denote the words and complex text in a symbolic form as *queries*, the *answers* are the fields generated by the words or text as sources. Any ontology map is a conceptual graph in RDF language where we structure the query as a structured variable. The conceptual map in Figure 1 is the input X of a complex field obtained by the superposition of the individual words in the map.

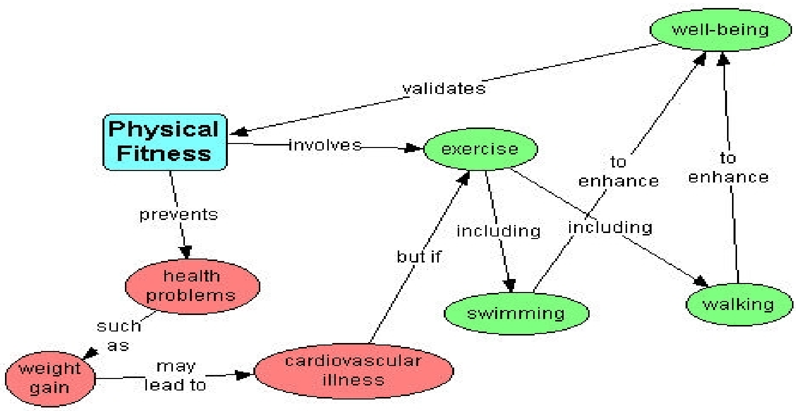


Fig. 1. Conceptual map as structured query. The map is a structured variable whose answer or meaning is the field G in the documents space located in the Web.

With the table two we have that the library H is the main reference for which we have

$$H = \begin{bmatrix} K_{1,1} & K_{1,2} & \dots & K_{1,N} \\ K_{2,1} & K_{2,2} & \dots & K_{2,N} \\ \dots & \dots & \dots & \dots \\ K_{M,1} & K_{M,2} & \dots & K_{M,N} \end{bmatrix}$$

Where M > N.

For any new concept in input $X = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_M \end{bmatrix}$ we can found the suitable sources S by which we can project X into the knowledge inside the library H.

The library can be represented in a geometric as shown in Figure 2.

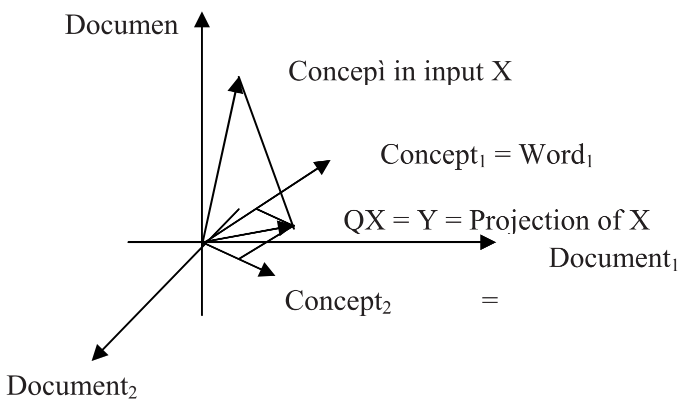


Fig. 2. Geometric representation of the Library (Documents) and concepts (words)

Recently much work has been done to realise conceptual maps of lexicon for natural language. This work has a foundation in Description Logic. With a conceptual map of natural language and Description Logic, we can create a semantic web. The problem with a semantic web is the integration of different concepts in a geometric frame. We are in the same situation as the quantum computer where the Geometric Hilbert space is the geometric tool to show the global properties of the classical logic and quantum logic. We propose in this paper to give a geometric image of knowledge. The meaning of the words and concepts are not located in the axiomatic description logic, in the conceptual map or in the lexical structure. The meaning is inside the space of the fields of instances of the words or documents. A prototype set of concepts, inside the space of the documents as object, generate an infinite number of other concepts in a space. Any concept inside the space of the concepts is obtained by an integration process of the prototype concepts. Now any new concept can be compared with the concepts that belong to the space of the concepts. The MS can reshape, with the minimum change the new concept X into another concept Y. Concept Y belongs to the set of concepts obtained by the integration of the prototype concepts. Y is inside the space of the prototype fields that represent prototype concepts.

MS can compute the part of X that cannot be integrated in the space of the concepts. Transformation of the space of the concepts generate a set of new prototype concepts by which we can represent the dynamical process of the concepts and the possible conceptual invariance. In this way a huge set of concepts can be reduced to a more simple set of prototype concepts and invariance. In the semantic web based on description logic and the conceptual map, we cannot reduce the concepts to a simple set of basic concepts which. With the description of invariance conceptual frame and the type of geometry (metric , Euclidean and non -Eucliden) we believe we can improve the primitive semantic structure base on the lexical description of the concepts.

3 Agents and Fuzzy in Morphic Computing

Given the word "warm" and three agents that are named Carlo , Anna and Antonio, we assume that any agent is a source for a Boolean fields in Figure 3.

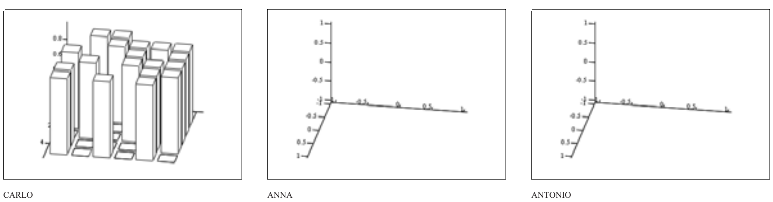


Fig. 3. Three Boolean fields $F_1(A,B)$, $F_2(A,B)$, $F_3(A,B)$ for three agents

The superposition of the three Boolean fields in Figure 3 is the fuzzy set

$$\mu(A,B) = S_1\ F_1(A,B) + S_2\ F_2(A,B) + S_3\ F_3(A,B) \tag{1}$$

At the word "warm" that is the query we associate the answer $\mu(A,B)$ that is a field. With the superposition process, we can introduce the logic of the fuzzy fields in this way.

When the logic variables values for p agents assume the usual 1 and 0 (boolean fields), we have the operations

$$\begin{aligned}
&(X_1, X_2, ..., X_p)(Y_1, Y_2, ..., Y_p) = (Y_1 X_1, Y_2 X_2, ..., Y_p X_p) \\
&(X_1, X_2, ..., X_p) + (Y_1, Y_2, ..., Y_p) = (Y_1 + X_1, Y_2 + X_2, ..., Y_p + X_p) \\
&\overline{(X_1, X_2, ..., X_p)} = (\overline{X_1}, \overline{X_2}, ..., \overline{X_p})
\end{aligned} \tag{2}$$

That are the vector product , the vector sum and the vector negation.

For any vector we can compute the scalar product

$$\mu(X_1, X_2, ..., X_p) = \mu(X) = \frac{m_1 X_1 + m_2 X_2 + ... + m_p X_p}{m_1 + m_2 + ... + m_p} =$$

$$S_1 X_1 + S_2 X_2 + ... + S_p X_p$$

where μ is the membership function and m_k where k = 1 ,.., p are the weights of any component in the Boolean vector. So the membership function is the weighted average of the 1 or 0 value of the variables X_k

We can show that

$$\begin{aligned} &when\ \mu(Y_1, ..., Y_p) \le \mu(X_1, ..., X_p) \text{ we have} \\ &\mu(Y_1 X_1, ..., Y_p X_p) = \min[\mu(X), \mu(Y)] - \mu(Y\overline{X}) \end{aligned} \tag{3}$$

$$\mu(Y_1 + X_1, ..., Y_p + X_p) = \max[\mu(X), \mu(Y_1)] + \mu(Y\overline{X}) \tag{4}$$

In fact we have $X_k Y_k + \overline{X_k} Y_k = Y_k$ and $X_k Y_k = Y_k - \overline{X_k} Y_k$ in the same way we have

$$X_k + Y_k = X_k + \overline{X_k} Y_k + X_k Y_k = X_k (1 + Y_k) + \overline{X_k} Y_k = X_k + \overline{X_k} Y_k$$

because in the fuzzy set for the Zadeh rule we have

$$\mu[(X_1, ..., X_p)\overline{(X_1, ..., X_p)}] = \min[\mu(X_1, ..., X_p), \mu\overline{(X_1, ..., X_p)}] > 0$$

$$\mu[(X_1, ..., X_p)\overline{(X_1, ..., X_p)}] = \max[\mu(X_1, ..., X_p), \mu\overline{(X_1, ..., X_p)}] < 1$$

the negation in (11) is not compatible with the other operations. So in the decomposition of the vector

$$\overline{(X_1, X_2, ..., X_p)} = (F(X_1), F(X_2), ..., F(X_p)) \tag{5}$$

at the negation $\overline{X_k}$ we substitute the negation

$$F(X_k) = \overline{X_k} \oplus \Gamma_k \text{ and } \oplus \text{ is the XOR operation}$$

We remark that when $\Gamma_k = 1$ we have $F(X_k) = X_k$ when $\Gamma_k = 0$ we have $F(X_k) = \overline{X_k}$ For the negation F we have

$$F(F(X_k) = X_k ..$$

In fact,

$$F(F(X)) = F(\overline{X} \oplus \Gamma(X)) = \overline{(\overline{X} \oplus \Gamma(X))} \oplus \Gamma(X) = (\overline{X}\Gamma(X) + X\overline{\Gamma(X)}) \oplus \Gamma(X) =$$

$$[X + \overline{\Gamma(X)}][\overline{X} + \Gamma(X)]\Gamma(X) + (\overline{X}\Gamma(X) + X\overline{\Gamma(X)})\Gamma(X) = X\Gamma(X) + X\overline{\Gamma(X)} = X$$

So the function F is an extension of the negation operation in the classical logic.

We remark that in (15) we introduce a new logic variable S_k. The vector $\Gamma = (\Gamma_1, \Gamma_2, ..., \Gamma_p)$ is the inconsistent vector for which the Tautology is not always true and the contradiction is not always false as in the Zadeh rule. In fact we have for the contradiction expression

$$C = F(X_k)X_k = (\overline{X_k} \oplus \Gamma_k)X_k = X_k\Gamma_k$$

For the previous case the fuzzy contradiction C is not always equal to zero. We remember that the classical contradiction is $\overline{X_k}X_k = 0$ always. When $\Gamma = (\Gamma_1, \Gamma_2, ..., \Gamma_p) = (0, 0, .., 0)$ we came back to the classical negation and $C_k = 0$.

For the tautology we have

$$T = F(X) + X = (\overline{X_k} \oplus \Gamma_k) + X_k = \overline{X_k}\,\overline{\Gamma_k} + X_k$$

The reference space is the space of all possible propositions and any agent is a Boolean field or prototype field. Given a Boolean field X in input, external agent, by the projection operator we generate the fuzzy field Y in output by the sources S as shown in Figure 4.

Because in the projection operation we loss information, the original boolean field for the input agent X is transformed in the fuzzy field

$$QX = Y = \mu(X) = S_1X_1 + S_2X_2 + ... + S_pX_p$$

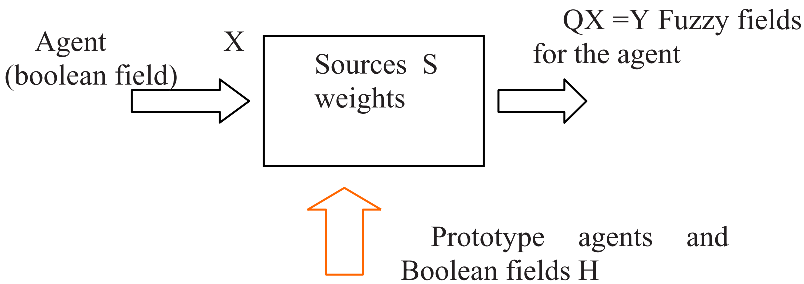


Fig. 4. Prototype Agents

4 Conclusion

Any fuzzy set is a scalar field of the membership values on the factors (reference space) (Wang and Sugeno 1982). We remember that any concept can be viewed as a fuzzy set in the factor space. So at the fuzzy set, we can introduce all the processes and concepts that we utilise in *Morphic Computing*. For the relation between concept and field, we introduce in the field theory an intrinsic fuzzy logic. So in *Morphic Computing*, we have an external logic of the projection or measure (quantum logic) and a possible internal fuzzy logic of the fuzzy interpretation of the fields. In the end, because we also use agents' superposition to define fuzzy sets and fuzzy rules, we can again use *Morphic Computing* to compute the agents inconsistency and irrationality. Thus, fuzzy set and fuzzy logic are part of the more general computation denoted as *Morphic Computing*.

References

1. G. Resconi and M. Nikravesh, Morphic Computing Part I and Part II, to be published in Forging New Frontiers, Series in Soft Computing and fuzziness, 2007.
2. G. Resconi and M. Nikravesh, Field Theory and Computing with Words, , FLINS 2006, 7th International FLINS Conference on Applied Artificial Intelligence, August 29-31, 2006 Italy
3. R. Sheldrake (1981), A New Science of Life: The Hypothesis of Morphic Resonance (1981, second edition 1985), Park Street Press; Reprint edition (March 1, 1995)
4. R. Sheldrake (1988), The Presence of the Past: Morphic Resonance and the Habits of Nature (1988), Park Street Press; Reprint edition (March 1, 1995)
5. L. A. Zadeh, Fuzzy Logic, Neural Networks and Soft Computing, Communications of the ACM, 37(3):77-84, 1994.
6. L. A. Zadeh and M. Nikravesh (2002), Perception-Based Intelligent Decision Systems; Office of Naval Research, Summer 2002 Program Review, Covel Commons, University of California, Los Angeles, July 30th-August 1st, 2002.
7. L. A. Zadeh, and J. Kacprzyk, (eds.) (1999a), Computing With Words in Information/Intelligent Systems 1: Foundations, Physica-Verlag, Germany, 1999a.
8. L. A. Zadeh, L. and J. Kacprzyk (eds.) (1999b), Computing With Words in Information/Intelligent Systems 2: Applications, Physica-Verlag, Germany, 1999b.
9. L.A. Zadeh, Fuzzy sets, Inf. Control 8, 338-353, 1965.

Part V

Perception Based Data Mining and Decision Making

Looking for Dependencies in Short Time Series Using Imprecise Statistical Data

Olgierd Hryniewicz

Systems Research Institute, Newelska 6,
01-447 Warsaw, Poland
Olgierd.Hryniewicz@ibspan.waw.pl

Abstract. In the paper we propose a very simple method for the analysis of dependencies between consecutive observations of a short time series when individual observations are imprecise (fuzzy). For this purpose we propose to apply a fuzzy version of the Kendall's τ statistic. The proposed methodology can be used for the analysis of a short series of opinion polls when answers of individual respondents are presented in an imprecise (fuzzy) form.

Keywords: time series, fuzzy observations, opinion polls, Kendall's τ.

1 Introduction

Analysis of dependencies in time series is a well established field of research in mathematical statistics, and recently has attracted interest among people working in data mining and knowledge discovery. Numerous methods have been proposed for the analysis of linear and non-linear trends that exist in real time series. There are two main aims of such analysis: to predict future values of the time series, and to describe complex mechanisms that govern its behavior. In the development of appropriate methods it is often assumed that the number of observations in the analyzed time series is sufficiently large. In other words, it is assumed that the time series is characterized by high frequency of observations. If this assumption is fulfilled we can use methods developed, for example, in the theory of automatic control, described in well known textbooks, such as a famous book by Box and Jenkins [1] and many others. However, in many practical situations the number of available observations is small, as – for example – in case of economic data. The situation becomes even more difficult if the available data come from statistical experiments with raw data delivered by human beings in imprecise, e.g. linguistic, form. Consider, for example, the case of a political opinion poll where potential voters are asked about their voting preferences. In order to simplify the problem let us assume that they have to choose from among two alternatives. If they do so, the results of a single poll are described by the binomial probability distribution, and we can use known statistical methods for the analysis of a time series consisted of several consecutive results of such polls. However, in reality the situation seems to be much more complicated. Usually many people are not eager to provide simple unequivocal answers. They prefer to present their opinions in a more soft way using different linguistic terms like "I may vote for ..", "I rather prefer to vote for ..", etc. instead of simple answers like "I will vote on ..". To describe

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 201–208, 2007.
springerlink.com

such results using the language of the theory of probability seems to be possible but prohibitively complicated. Instead, we can use the methodology proposed by Hryniewicz [2] where linguistic imprecision of individual answers is merged with stochastic uncertainty that stems from the statistical character of a poll. We briefly present this methodology in the second section of this paper. The next step of the analysis is to look for possible dependencies between the results of few consecutive polls. In the third section of the paper we sketch a non-parametric statistical methodology for the analysis of short time series in case of precise data. This methodology is based on the well known Kendall's τ statistics applied for dependent pairs of observations. In this section we also present a new application of this methodology for the case of a time series consisted of fuzzy-random observations. Finally, we present some conclusions and propose some interesting topics for further investigation.

2 Mathematical Model for Imprecise (Fuzzy) Binomial Data

Let us consider a statistical experiment when a sample of n items is investigated, and individual observations of sampled items are described by a zero-one valued random variable. Interpretations of zeroes and ones depend upon the context, but without loosing generality we can assume that 1 means "in favor", and 0 means "against". When the probability of observing 1 is equal to p and is constant for all items in the sample, the total number K of observations "in favor" in the sample is governed by the binomial distribution

$$P(K=k)=\binom{n}{k}p^k(1-p)^{n-k}, k=0,\ldots,n\,. \tag{1}$$

Hence, a pair (n,k) is a sufficient statistic for the estimation of the unknown parameter p, and the estimated value of p is equal to k/n. However, in many statistical experiments, when the input data comes from human beings, experimenters face difficulties when they attempt to receive clear answers, either "in favor" or "against". For example, people asked if they are "in favor" of buying certain goods refrain from giving unequivocal answers "yes" or "not". For many of them the answer is simple "yes" represented by 1 in our mathematical model, or "no" represented by 0, but from time to time their answers are imprecise such as "I may buy", "I might buy", "I rather don't buy", or simply "I don't know". First attempts to cope with the statistical analysis of such imprecise data can be found in Hryniewicz [2], who - in the context of quality control - assumed that each inspected item is described by a family of fuzzy subsets of a set $\{0,1\}$, with the following membership function

$$\mu_0\,|\,0+\mu_1\,|\,1, 0\le\mu_0,\mu_1\le 1, max\{\mu_0,\mu_1\}=1\,. \tag{2}$$

where the notation $\mu|k$ stands for <*membership grade*>|<*value*>. Thus, equation (2) defines a possibility distribution on a binary set representing the result of the observation of a sampled item.

When the result of the observation is "in general, against", it is expressed as a fuzzy set with the membership function $1|0+\mu_1\,|\,1$. Crisp observations must be represented by crisp sets, so in the case of crisp "against" the membership function is given by $1|0+0|1$. On the other hand, if the result of the observation is "in general, in favor", it is represented by a fuzzy set with the membership function $\mu_0|0+1|1$. Crisp answers of the "in favor" type are described by crisp sets with the membership function $0|0+1|1$.

The fuzzy probabilistic model for the imprecise data described above is given in Hryniewicz [3] who noticed that on a given α level, where $\alpha \in (0,1]$, the result of the observation is described by a three-point probability distribution on the set of possible observations: $A_1^\alpha = \{0\}$, $A_2^\alpha = \{0,1\}$, and $A_3^\alpha = \{1\}$, where $\{0\}$ means that the observation is of the "against" type, $\{1\}$ means that the observation is of the "in favor" type, and $\{0,1\}$ means that (at the given α level) it is not possible to distinguish if the observation is "against" or "in favor". Let X be a random variable that describes a single observation, and let X be defined by the following probabilities: $p_{00}^\alpha = P(X = A_1^\alpha)$, $p_{01}^\alpha = P(X = A_2^\alpha)$, and $p_{11}^\alpha = P(X = A_3^\alpha)$, where $p_{00}^\alpha + p_{01}^\alpha + p_{11}^\alpha = 1$. If we are interested in the probability of the observation of "in favor" type, it has been shown in [3] that this probability, on a given α level, is represented as an interval $p_1^\alpha = \lfloor p_{11}^\alpha, p_{11}^\alpha + p_{01}^\alpha \rfloor$. Thus, in case of imprecise observations, the probability of obtaining the result of the "in favor" type is fuzzy, and is defined by the α-cuts given by the intervals mentioned above.

It has been assumed in [2], that in the sample of n items in n_1 cases the sample items are characterized by fuzzy sets described by the membership function $\mu_{0,i}|0+1|1, i=1,\ldots,n_1$, and in the remaining $n_2 = n - n_1$ cases by fuzzy sets described by the membership function $1|0+\mu_{1,i}|1, i=1,\ldots,n_2$. Without loss of generality we can assume that $0 \le \mu_{0,1} \le \ldots \le \mu_{0,n_1} \le 1$, and $1 \ge \mu_{1,1} \ge \ldots \ge \mu_{1,n_2} \ge 0$. Hence, the fuzzy total number of observations of the "in favor" type in this sample, calculated using Zadeh's extension principle, is given by Hryniewicz [2]:

$$\begin{aligned} \tilde{d} = & \quad \mu_{0,1}|0+\mu_{0,2}|1+\ldots+1|n_1 + \\ & + \mu_{1,1}|(n_1+1)+\ldots+\mu_{1,n_2}|(n_1+n_2) \end{aligned} \qquad (3)$$

If we are interested rather in probabilities than in the numbers of observations we can use the following representation

$$\begin{aligned} \tilde{p}^* = & \quad \mu_{0,1}|0+\mu_{0,2}|(1/n)+\ldots+1|(n_1/n)+ \\ & + \mu_{1,1}|((n_1+1)/n)+\ldots+\mu_{1,n_2}|((n_1+n_2)/n) \end{aligned} \qquad (4)$$

If the results of statistical experiments, represented either by (3) or by (4), have been observed in a chronological order, then they constitute a fuzzy time series. In the next section we will analyze this series in order to find stochastic dependencies that are characteristic for certain trends of important practical consequences.

3 Statistical Analysis of Dependencies in Time Series Consisted of Few Fuzzy-Binomial Observations

The analysis of time series is a well established area of statistics. Moreover, it has also attracted interest of scientists working in data mining and other areas of artificial intelligence and soft computing. In the majority of cases, the main driving force of research in this area is the analysis of financial data. Typical data sets often consist of hundreds of observations, and the proposed analytical and computer-oriented methods profit from this fact. When the number of observations in a time series is small, some classical statistical methods, based on the regression analysis, are useful, when certain additional assumptions are fulfilled. However, when the number of available data points is small it is often even impossible to verify those assumptions. We also face a similar situation when the time series is long, but its short intervals are obviously different due to varying external circumstances. In all such cases nonparametric statistical methods seem to be more appropriate.

Specialists who analyze time series are usually interested in finding some dependencies between consecutive observations. Consider, for example, a series of consecutive opinion polls that follow a certain political event. It is of interest to find whether this event has triggered continuous decline (growth) of public support for certain politician or political party. Political analysts are usually not aware of stochastic nature of such results and of the errors involved. On the other hand, application of simple statistical methods, like linear regression, seems to be not fully justified, as precisely defined linear (or nonlinear) models of dependence are usually unidentifiable. Therefore, what seems to be more interesting to an analyst (even if he/she is not aware of) is to find whether there exists positive stochastic dependence between consecutive points of the observed time series. Positive dependence means in practice that a large value of an observation is followed by another large value rather than by a small one. Symmetrically, small values of the observed time series are accompanied by other small values, and not by large ones. This feature is characteristic for autoregressive models [1] when the data sets consist of many short intervals characterized by trends of varying directions. It has to be noted, however, that the existence of positive dependence is only a sufficient condition for the presence of trends. One can easily show examples when obvious trends are observed for data which are not positively dependent. However, in case of negative dependence the formation of trends seems to be impossible.

There exist many statistical methods for the analysis of stochastic dependence between vectors of observations. In order to apply one of them let us notice that the series of n consecutive observations $(X_1, X_2, \ldots, X_n)$ can be represented by two shorter vectors $(Y_1, Y_2, \ldots, Y_{n-1})$ and $(Z_1, Z_2, \ldots, Z_{n-1})$ whose elements are defined as: $Y_i = X_i, i = 1, \ldots, n-1$, and $Z_i = X_{i+1}, i = 1, \ldots, n-1$. Positive dependence between consecutive observations of the original time series $(X_1, X_2, \ldots, X_n)$ is thus equivalent to the positive dependence between vectors $(Y_1, Y_2, \ldots, Y_{n-1})$ and $(Z_1, Z_2, \ldots, Z_{n-1})$.

For verification of the positive dependence we can use many statistical methods. In this paper we propose to use a well known Kendall's τ statistic defined as

$$\tau = \frac{4}{n-1}\sum_{i=1}^{n-1} V_i - 1 \tag{5}$$

where

$$V_i = card\{(Y_j, Z_j) : Y_j < Y_i, Z_j < Z_i\}/(n-2), i = 1,\ldots,n-1 \tag{6}$$

Statistical properties of the Kendall's τ are known when the pairs of observations $(Y_i, Z_i), i = 1,\ldots,n-1$ are mutually independent. However, in the considered case they are obviously dependent. Therefore, in order to verify the statistical hypothesis that consecutive observations are mutually independent against the hypothesis that they are positively dependent we should compute new critical values of the test statistic which will reflect the obvious dependence between observed pairs. Fortunately, in case of interesting us short time series it can be easily done either by simple analytical methods or by computer simulations. Some typical critical values of the τ statistic, together with respective probabilities calculated under the assumption of independence, are presented in Table 1.

Table 1. Critical vales of the τ statistic when pairs of observations are dependent

n=4		*n*=5		*n*=6		*n*=7	
τ_{crit}	$P(\tau \ge \tau_{crit})$	τ_{crit}	$P(\tau \ge \tau_{crit})$	τ_{crit}	$P(\tau \ge \tau_{crit})$	τ_{crit}	$P(\tau \ge \tau_{crit})$
1	0,0833	1	0,0168	1	0,00267	1	0,00042
0,333	0,250	0,666	0,050	0,8	0,00834	0,866	0,00119
		0,333	0,167	0,6	0,03056	0,733	0,00477
				0,4	0,0865	0,600	0,01356
				0,2	0,275	0,466	0,0399
						0,333	0,1157

When precise data are observed, the Kendall's τ statistic is calculated and compared with the appropriate critical value taken from Table 1. This value is chosen in such a way that the probability of the erroneous decision $P(\tau \ge \tau_{crit})$ is sufficiently small.

The situation becomes different when the observed data points are imprecisely defined, and are described by fuzzy sets. In the considered case of fuzzy binomial data it is very unlikely that we will be able to present an unequivocal ordering of all data points. Therefore, the value of the Kendall's τ statistic has to be also fuzzy in this case.

In order to compute the fuzzy version of the Kendall's τ statistic for the considered fuzzy time series let us approximate the fuzzy set $\tilde{p}^*$ given by (4) by an equivalent fuzzy number. Let us notice that this fuzzy set is defined by *n*+1 pairs $(p_k, \mu_k), k = 0,\ldots,n$, where values $p_k = k/n$ are ordered. Moreover, from the assumption that precedes (3) it turns out that there exists a convex hull of $\tilde{p}^*$ in a form of a fuzzy number $\tilde{p}$ whose membership function is the following piece-wise linear function.

$$\mu(p)=\begin{cases} 0 & p<p_0 \\ \dfrac{p(\mu_{k+1}-\mu_k)+(p_{k+1}\mu_k-p_k\mu_{k+1})}{p_{k+1}-p_k} & p_k\le p\le p_{k+1} \quad k=0,\dots,n \\ 0 & p\ge p_n \end{cases} \tag{7}$$

Now, let us the Zadeh's extension principle for the definition of the fuzzy equivalent of the Kendall's τ in the considered case. First, let us rewrite (6) in the following form

$$V_i=\underset{j\neq i}{card}\left\{(p_j,p_{j+1}):\ p_j<p_i,\ p_{j+1}<p_{i+1}\right\}\big/(n-2),\ i=1,\dots,n-1 \tag{8}$$

We have to write the fuzzy form of (8) for the fuzzy time series $\tilde{p}_1,\dots,\tilde{p}_n$, where membership functions of all fuzzy data points have the form given by (7). Let us notice now that each fuzzy data point is completely defined by the set of its α-cuts $\left[p_{i,L}^{\alpha},p_{i,U}^{\alpha}\right], 0<\alpha\le 1$. Hence, the fuzzy equivalent of V_i, denoted by $\tilde{V}_i$, is defined by the set of its α-cuts $\left[V_{i,L}^{\alpha},V_{i,U}^{\alpha}\right], 0<\alpha\le 1$, where

$$V_{i,L}^{\alpha}=\min_{\substack{p_i\in\left[p_{i,L}^{\alpha},p_{i,U}^{\alpha}\right]\\ i=1,\dots,n}}\ \underset{j\neq i}{card}\left\{(p_j,p_{j+1}):\ p_j<p_i,\ p_{j+1}<p_{i+1}\right\}\big/(n-2),\ i=1,\dots,n-1 \tag{9}$$

and

$$V_{i,U}^{\alpha}=\max_{\substack{p_i\in\left[p_{i,L}^{\alpha},p_{i,U}^{\alpha}\right]\\ i=1,\dots,n}}\ \underset{j\neq i}{card}\left\{(p_j,p_{j+1}):\ p_j<p_i,\ p_{j+1}<p_{i+1}\right\}\big/(n-2),\ i=1,\dots,n-1 \tag{10}$$

Having the α-cuts $\left[V_{i,L}^{\alpha},V_{i,U}^{\alpha}\right], 0<\alpha\le 1$ for all $i=1,\dots,n$ we can straightforwardly calculate the α-cuts of the fuzzy Kendall's τ statistic $\left[\tau_L^{\alpha},\tau_U^{\alpha}\right], 0<\alpha\le 1$, and thus we can obtain its membership function.

Positive dependence between consecutive fuzzy observations of a short time series will be identified if we reject the statistical hypothesis that these observations are mutually independent. For solving this problem in the considered case we can use different approaches known from the literature on fuzzy statistics. For example, we can use a possibilistic approach proposed in Hryniewicz [3]. We have to reject the hypothesis of independence on the significance level δ and for the given necessity degree γ if the following inequalities hold: $\tau_L^{1-\gamma}\le\tau_{crit}$, and $P(\tau\ge\tau_{crit})\le\delta$ **.**

In the previous paragraphs we have presented a general description of the proposed fuzzy Kendall's τ statistic which is based on the Zadeh's extension principle. However, in real applications we have to propose an efficient algorithm for finding the α-cuts of the fuzzy Kendall's τ. The construction of this algorithm will be apparent if look at Figure 1.

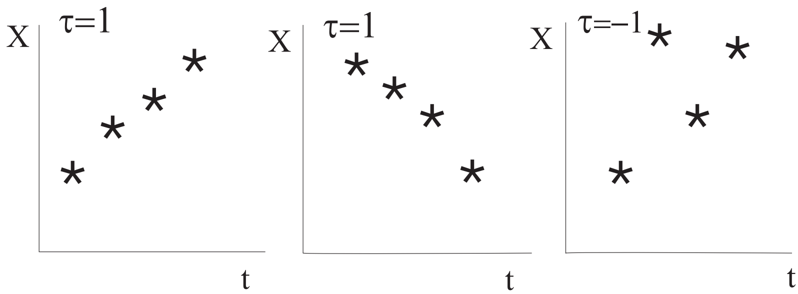


Fig. 1. Configurations of observations of a time series for the values of the Kendall's τ statistic are either the largest or the lowest

For the given value of α the largest value of τ is attained for a series of values $\hat{p}_i \in \left[p^{\alpha}_{i,L}, p^{\alpha}_{i,U}\right], 0 < \alpha \leq 1$ that form a monotone (or nearly monotone) increasing (decreasing) series. To find such a series we can start with the series $\hat{p}_i = p^{\alpha}_{i,L}$, $0 < \alpha \leq 1$. In the next step we can increase certain values of this series in order to arrive at a monotone (or nearly monotone) increasing series. The same procedure should be repeated in search of a monotone (or nearly monotone) decreasing series. In this case we can start with the series $\hat{p}_i = p^{\alpha}_{i,U}$, $0 < \alpha \leq 1$, and in the next step we should decrease certain values of this series in order to arrive at a monotone (or nearly monotone) decreasing series.

The lowest value of τ is attained for a series of values $\hat{p}_i \in \left[p^{\alpha}_{i,L}, p^{\alpha}_{i,U}\right], 0 < \alpha \leq 1$ that form an alternating series of values such that the observations with odd (even) indices form a decreasing series, and the observations with even (odd) indices form an increasing series. To find such a series we can start with the series $p^{\alpha}_{1,L}, p^{\alpha}_{2,U}, p^{\alpha}_{3,L}, \ldots$ or with the series $p^{\alpha}_{1,U}, p^{\alpha}_{2,L}, p^{\alpha}_{3,U}, \ldots$. In the next step we can increase certain values initially defined by the lower limits of the α-cuts and decrease certain values initially defined by the upper limits of the α-cuts in order to arrive at an alternating (or nearly alternating) series.

4 Conclusions

We have proposed a simple procedure for the verification if a short time series consisted of fuzzy data points manifest a tendency to form trends. For this purpose we use a fuzzy version of the Kendall's τ statistic. We have illustrated this methodology with a possible application for the analysis of opinion polls with imprecise answers of respondents. The proposed algorithm for finding the fuzzy τ is presented in a descriptive way. Thus, the future work on this problem should be focused on the construction of formal algorithms for the construction of the Kendall's τ statistic. Moreover, some research has to be done in order to find possible applications of the fuzzy versions of other nonparametric tests (for example, the Spearman's ρ test) that can be useful for the detection of the directions of trends.

References

1. Box, G.E.P., Jenkins, G.M., Reinsel G.C.: Time Series Analysis, Forecasting and Control. Prentice-Hall, Englewood Cliffs (1994)
2. Hryniewicz, O.: Statistical Decisions with Imprecise Data and Requirements. In: Kulikowski, R, Szkatula, K, Kacprzyk, J (eds) Systems Analysis and Decisions Support in Economics and Technology. Omnitech Press, Warszawa (1994), pp 135-143.
3. Hryniewicz, O.: Possiblistic decisions and fuzzy statistical tests. Fuzzy Sets and Systems 157 (2006), 2665-2673
4. Hryniewicz, O.: Statistics with fuzzy data in Statistical Quality Control. Soft Computing (submitted)

Perception Based Time Series Data Mining for Decision Making

Ildar Batyrshin and Leonid Sheremetov

Research Program in Applied Mathematics and Computing (PIMAyC),
Mexican Petroleum Institute, Eje Central Norte Lazaro Cardenas 152, 07730, D.F., México
{batyr,sher}@imp.mx

Abstract. In this paper, several aspects of perception based time series data mining based on the methodology of computing with words and perceptions are discusses. First, we consider possible approaches to precisiate perception based patterns in time series data bases and types of fuzzy constraints used in such precisiation. Next, several types of associations in time series data bases and the possible approaches to convert these associations in generalized constraint rules are discussed. Finally, we summarize the methods of translation of expert knowledge and retranslation of solutions.

1 Introduction

Time series data mining is the rapidly growing area of research [2, 4, 15, 19, 23, 24, 26, 28, 34, 37]. Large time series data bases are collected in different areas of human activity such as economics, finance, meteorology, geophysics, petroleum industry, medicine etc. The analysis of this information is very important for solving many real problems. The goal of data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [18]. The following list contains the main time series data mining tasks [23, 28]:

- **Segmentation:** Split a time series into a number of “meaningful” segments.
- **Clustering:** Find natural groupings of time series or time series patterns.
- **Classification:** Assign given time series or time series patterns to one of several predefined classes.
- **Indexing:** Realize efficient execution of queries.
- **Summarization:** Give short description of a time series which retains its essential features in considered problem.
- **Anomaly Detection:** Find surprising, unexpected patterns.
- **Motif Discovery:** Find frequently occurring patterns.
- **Forecasting:** Forecast time series values based on time series history.
- **Discovery of association rules:** Find rules relating patterns in time series.

Data mining is often declared as a methodology for knowledge extraction from data bases but what formal methods to use to integrate this knowledge with knowledge base in intelligent decision support systems is still an open problem. To solve

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 209–219, 2007.
springerlink.com

this problem it is necessary to represent the results of data mining in the form used in human knowledge and human decision making procedures. Human knowledge and decisions are often perception based and formulated linguistically. Computing with words and perceptions (CWP) gives methodology to handle perception based information and to develop knowledge based systems [38-42]. Perception based data mining integrating methods of data mining with CWP will give the possibility to extract information from data bases in linguistic form suitable for their use in perception based decision making procedures. Perception based time series data mining methods should operate with perception based time series patterns, find perception based associations between them and extract knowledge in time series data bases suitable for supplementation and precisiation knowledge base about problem area used in perception based decision making procedures. The centerpiece of fuzzy logic is the concept of a generalized constraint [40-42]. It is important to develop the methods of transformation of association rules generated by data mining procedures in generalized constraint rules that can be used in perception based knowledge representation.

In this work we discuss different aspects of perception based time series data mining based on the methodology of computing with words and perceptions and on results obtained in data mining, time series analysis, fuzzy logic and soft computing. In Sections 2 and 3 we consider possible approaches to precisiate perception based shape patterns in time series and discuss the main types of fuzzy constraints used in precisiation of perceptions. In Sections 4 and 5 we discuss the types of associations that can be looked for in time series data bases and the possible approaches to convert obtained associations in generalized constraint rules that can be used for knowledge supplementation in perception based decision making procedures. In section 6 we summarize methods of translation of expert knowledge and retranslation of solutions. In Conclusion we discuss the possible directions of research in perception based time series data mining.

2 Precisiation of Perception Based Patterns

Human perceptions about time, time series values, patterns and shapes, about associations between patterns and time series can be represented by words whose meaning is defined on the following domains of time series data bases [10]:

- time domain: time intervals (*several days*), absolute or relative position on time scale (*near future*), periodic or seasonal time intervals (a *week before Christmas*);
- range of time series values (*large price, very low level of production*);
- a set of time series shape patterns (*quickly increasing and slightly concave*);
- a set of time series, attributes or system elements (stocks of *new* companies);
- a set of relations between TS, attributes or elements (*highly associated*);
- a set of possibility or probability values (*unlikely, very probable*).

Such perceptions can be represented as fuzzy sets defined on corresponding domain. Some methods of precisiation of perception based shape patterns are considered below.

2.1 Shape Patterns Defined by Signs of First and Second Derivatives

Triangular episodes representation language was formulated in [14] for representation and extraction of temporal patterns. These episodes are defined by the signs of the first and second derivatives of time dependent variable and can be described as: *A: Increasing and Concave; B: Decreasing and Concave; C: Decreasing and Convex; D: Increasing and Convex;* etc. These episodes can code time series patterns as a sequence of symbols like ABCDAB in process monitoring, diagnosis, control etc.

In [24] the profile of measured variable $x_j(t)$ is transformed into the qualitative form as a result of approximation of $x_j(t)$ by a proper analytical function from which the signs of the first derivative *sd1* and the second derivative *sd2* are extracted. Each shape pattern on the interval where a pattern is defined is described in a qualitative form as follows: qshape= *{sd1; sd2}*. For example, the following patterns are considered: *IncreasingConcavelyConvexly:* {(+);(-,+)}, *ConcaveMaximumConcave:* {(+,-);(+,-,+)}, etc. Such patterns are used in control of fermentation processes, e.g.: *IF (DuringTheLast1hr Dissolved Oxygen has been DecreasingConcavelyConvexly) THEN (Report: Foaming) and (Feed antifoam agent).*

2.2 Granulation of Linear Trend Patterns

A scaling and granulation of linear trend patterns was considered in many papers. The linear shape patterns can be described by words: *up, rose, dropped, dropped sharply, quickly increasing, very slowly decreasing* etc [2,4,5]. Such patterns can be used in linguistic descriptions of time series in the form of the rules: R_k: *If T is* T_k *then Y is* A_k, where T_k are fuzzy intervals, like *Large*, *Between A and B* etc, and A_k are linguistic descriptions of linear trends, like *Quickly Increasing* etc [12]. The methods of fuzzy granulation of linear shape patterns based on an extension of a fuzzy set in given direction are discussed in [5]. They are based on extension principle of Zadeh and on cylindrical extension in direction [38].

2.3 Granulation of Convex-Concave Shape Patterns

The paper [4] introduces the parametric functions describing perception based shape patterns *falling less steeply, falling more steeply, rising more steeply, rising less steeply,* respectively.

Parametric methods of convex-concave modification of linear functions are discussed in [6,11]. These methods give possibility to combine linear trend patterns with convex-concave patterns and to precisiate linguistic descriptions like *"Slowly Decreasing and Strongly Convex"*. A fuzzy granulation of convex-concave shape patterns can be done similarly to the granulation of linear patterns [5]. Such patterns are used in [11] in modeling qualitative expert forecasting [13].

2.4 Ontology of Shape Patterns

The use of ontology can facilitate the reasoning with shape patterns and text generation. In [37] the ontology of sensor data from gas turbines is presented as a hierarchy of *Disturbance Patterns* divided into three types: *Spikes, Steps* and *Oscillations.* These types contain *Upward, Downward* and/or *Erratic* patterns. The last is divided on two specific types of patterns.

2.5 Shape Definition Language

A Shape Definition Language (SDL) was developed in [2] for retrieving objects based on shapes. SDL allows a variety of queries about the shapes of histories. The alphabet of SDL contains a set of patterns like *up (slightly increasing transition), Up (highly increasing transition), down (slightly decreasing transition)* etc. It can be used for the definition of shapes as follows: (*shape name (parameters) descriptor*). Complex shapes can be derived by recursive combination of elementary and previously defined shapes.

3 Perception Based Constraints

Each perception defines some crisp or fuzzy constraint on a corresponding domain. Reasoning procedure of CWP should be able to operate with perceptions defined on different domains and in constraint propagation [38, 40, 42].

3.1 Point-Wise Constraints

A fuzzy constraint is called a point-wise (pw-) constraint if it is defined on time domain (domain of time points) or can be propagated on time domain such that a degree of fulfillment of this constraint can be determined in each time point [8,9]. Fig. 1 depicts an example of pw-constraint propagation on time domain. by extension principle of Zadeh.

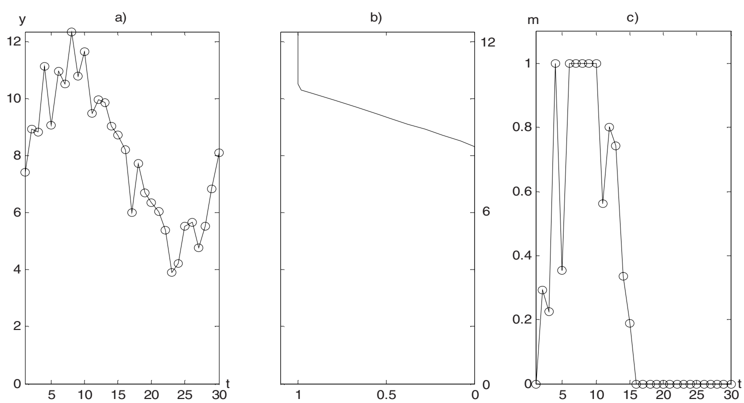


Fig. 1. a) a time series of *oil production* values; b) fuzzy constraint defined on a range of oil production values by perception *high level of oil production*; c) fuzzy constraint *days with high level of oil production* induced on the time domain by extension principle of Zadeh

3.2 Window-Wise Constraints

A fuzzy constraint is called a window-wise (ww-) constraint if it is defined on interval of time points or can be propagated on interval of time points [8,9]. Most of time series shape patterns like *increasing, convex* etc considered above are given as ww-constraints.

3.3 Composite Constraints Defined by Logical Connectives

Composite constraints can arise as a result of aggregation of simple constraints or as result of constraint propagation during inference procedure. Generally, composite constraints can be defined by some relation *R(A,B,…)* where *A, B, …* are simple constraints. The example illustrates a composite constraint obtained from pw-constraints:

Days of winter months with high level of oil production in well N and low level of gas production in well M.

Here, the verbal connectives *with* and *and* can be replaced by logical connective *and* which will act point-wise on membership values of fuzzy setspropagated on the time domain by simple pw-constraints: 1) *days of winter months;* 2) *days with high level of oil production in well N;* 3) *days with low level of gas production in well M*.

A composite constraint can aggregate pw- and ww-constraints like *several days with high level of oil production* as result of transformation of pw-constraints in ww-constraints [9].

3.4 Composite Constraints Defined by Temporal Relations

Consider example:
Stable level of water production in the well followed by a rapid increase of water production in this well.

Perceptions *Stable level* and *Rapid increase* give examples of ww-perceptions which are based on the analysis of time series values in some window of time points. Temporal relations can describe partial ordering of events and their occurrence within a time interval [15, 22, 32, 34]. Such relations can include the temporal relations of Allen's interval logic [3, 19]: *before*, *meets, overlaps, contains* etc.

3.5 Composite Constraints Defined by Spatial Relations

Below is an example of composite constraint defined by spatial relations:

Rapid increase of water production in neighbouring wells.

Spatial relations can describe position, distance or direction [22].

4 Association Rules

Association rules give examples of composite constraints provided by evaluations of the strength of association between simple or composite perceptions.

4.1 Association Rules with Support and Confidence Measures

Fuzzy association rules are studied in [16,17]. An extension of association rules on TSDB with fuzzy pw- and ww- constraints is considered in [9]. Suppose we have *m* time series $Y=\{Y_i\}$, $(i=1,\ldots,m)$, taking values $Y_i=(y_{i1},\ldots,y_{iN_i})$ in time points $T_i=(t_{i1},\ldots,t_{iN_i})$. Suppose fuzzy pw-constraint A is defined on time domains $T_A=\bigcup_{i\in I_A} T_i$, where $I_A\subseteq\{1,\ldots,m\}$, such that $\mu_A(t_{iq})$ denotes a degree of fulfillment of constraint A in time point $t_{iq}\in T_A$. Similarly, denote T_B a time domain for fuzzy constraint B. Suppose $D\subseteq T_A\times T_B$ is a set of pairs of time points (t_{iq},t_{jp}) used for evaluating association rule and $|D|$ is a number of elements in D. For each pair (t_{iq},t_{jp}) from D put $\mu_A(t_{iq},t_{jp})=\mu_A(t_{iq})$ and $\mu_B(t_{iq},t_{jp})=\mu_B(t_{jp})$. Define support s and confidence c measures of association rule

$$R: A\Rightarrow B,\ (s,c),$$

with point-wise constraints A and B as follows:

$$s=\frac{1}{|D|}\sum_{(t_{iq},t_{jp})\in D} T(\mu_A(t_{iq}),\mu_B(t_{jp})),\qquad c=\frac{\sum_{(t_{iq},t_{jp})\in D} T(\mu_A(t_{iq}),\mu_B(t_{jp}))}{\sum_{(t_{iq},t_{jp})\in D} \mu_A(t_{iq})}$$

where T is a t-norm, e.g. $T(x,y)=min(x,y)$. Association measures for association rules with ww-constraints and special cases of these measures are considered in [9].

In time series domains, association rules can be written as follows [9]:

$$R: \mathit{If}\ C\ \mathit{then}\ A\Rightarrow B,\ (s,c),$$

where C is some perception based condition which serves as a filter applying fuzzy constraints on time points where the association between time series patterns A and B is looked for. For example, A,B and C can be given as *C= Days of winter months*, *A= High level of oil production in well N* and *B = High level of gas production in well M*.

4.2 Association Rules with Perception Based Frequencies

Frequencies used in support and confidence measures of association rules can be based on perception based evaluations [39]. Consider the balls-in-box example of Zadeh [40]:

A box contains balls of various sizes and weights. Most are large. Many large balls are heavy.

Such perceptions can be transformed in a form of an association rule [9]:

large ball $\Rightarrow$ *heavy ball, (s=many*$\times$*most, c=many).*

4.3 Association Rules Based on Correlation

Association rules based on correlation can have the form [8]:

R: If C then A associated with B, (W),

where C is a constraint on time series, association is calculated as a correlation coefficient between constrained time series A and B and W is a significance of the rule given by the values of correlation coefficient and *t*-test. In fuzzy case a granular correlation coefficient studied in [30] can be considered.

4.4 Local Trend Association

Moving approximation transform (MAP) replaces time series by a sequence of slopes ("local trends") of linear approximations of time series in a sliding window of size *k* [7]. A local trend association between time series is defined by measure of similarity of MAPs of these time series.

5 Transformation of Associations in Generalized Constraint Rules

The methods of transformation of associations in generalized constraint rules [38, 40] were considered in [8,9].

5.1 Association Rules

Association rule $A \Rightarrow B, (s,c)$ can be transformed into a generalized constraint rule

$$A \rightarrow B \;\; isr\; P(s,c),$$

with a perception based evaluation $P(s,c)$ of probability, possibility, true or utility of the rule depending on confidence and support values, where a parameter *r* in *isr* specifies a mining of $P(s,c)$ [38,40]. Such transformation can be defined by suitable mapping [9] of pair (s,c) into $P(s,c)$. For a constrained association rule *If C then* $A \Rightarrow B, (s,c)$ a corresponding inference rule have a form *If C then* $A \rightarrow B \;\; isr\; P(c,s)$.

5.2 Association Rules with Perception Based Frequencies

Association rules with perception based frequencies [39] can be transformed into inference rules similarly to association rules with measurement based frequencies. For the ball-in-box example of Zadeh [40] the following generalized-constraint inference rule can be generated:

$$large\ ball \rightarrow heavy\ ball \;\; isr\; P(\,many \times most\,, many).$$

5.3 Correlation Rules

A correlation coefficient is usually considered as a measure of linear dependence between variables. Correlation rules can be transformed into inference rules as follows:

$$A\ is\ Increasing \rightarrow B\ is\ Increasing \;\; isr\; P(W),$$
$$A\ is\ Increasing \rightarrow B\ is\ Decreasing \;\; isr\; P(W), \text{ etc},$$

where *Increasing* and *Decreasing* describe linear shape patterns of time series and $P(W)$ is a perception based evaluation of the inference rule depending on the values of the correlation coefficient and *t*-test. For conditional correlation we have, e.g.:

$$If\ C\ then\ A\ is\ Increasing \rightarrow B\ is\ Increasing \;\; isr\; P(W).$$

5.4 Local Trend Associations

Local trend associations can be transformed into the generalized-constraint rules

$$A \text{ is } ID_A \rightarrow B \text{ is } ID_B \text{ isr } P(mlta(A,B),|w|),$$

where ID_A and ID_B take values like *increasing, slowly decreasing* etc, *mlta(A,B)* is a value of a measure of local trend associations calculated for window sizes $|w|$.

6 Translation of Expert Knowledge and Retranslation of Solutions

The important steps in development of perception based decision making system are translation of expert knowledge on the input of the system into a formalized form and retranslation of obtained solutions on the output in a linguistic form [38]. Here we discuss approaches to generation of translation and retranslation procedures.

6.1 Translation of Expert Knowledge in Formalized Form

Representation of expert knowledge in the form of the rules is used in most of expert systems. A rule-based fuzzy expert system WXSYS realized in FuzzyClips uses expert knowledge defined on TSDB domains to predict local weather [29]. Below is an example of expert perceptions translated further in a formalized form: *"If the current wind is blowing from S to SW and the current barometric pressure is rising from 30.00 or below, then the weather will be clearing within a few hours, then fair for several days"*.

6.2 Clustering and Linguistic Translation of Shape Patterns

In [15] it is used a clustering of complex shape patterns and generation of association rules in the form: $A \overset{T}{\Rightarrow} B$, i.e. *if A occurs, then B occurs within time T,* where *A* and *B* are identifiers of discovered patterns (clusters of patterns). Further, experts give linguistic interpretation of shapes representing obtained clusters. As an example, from the daily closing share prices of 10 database companies traded on the NASDAG it was found an association rule with linguistic interpretation: *A stock which follows a 2.5-week declining pattern of s18 "sharper decrease and then leveling out", will likely incur "a short sharp fall within 4 weeks (20 days) before leveling out again" (the shape of s4).*

6.3 Generation of Summaries Based on Grammars

System called Ana [25] uses special grammar for generation of summaries based on patterns retrieved from the summaries generated by experts. Ana generates descriptions of stock market behavior like the following: *"Wall Street's securities markets rose steadily through most of the morning, before sliding downhill late in the day. The stock market posted a small loss yesterday, with the indexes finishing with mixed results in active trading."*

The more extended system called StockReporter is discussed in [31]. It reports on the behavior of any one of 100 US stocks and on how that stock's behavior compares with the overall behavior of the Dow Jones Index or the NASDAQ. StockReporter can generate a text like the following: *"Microsoft avoided the downwards trend of the Dow Jones average today. Confined trading by all investors occurred today…"*.

An architecture for generating short summaries of large time-series data sets is discussed in [37]. Fuzzy linguistic summaries and descriptions are discussed in [12, 21, 27, 35]. Frequency based trend summaries are discussed in [20].

6.4 Perception Based Forecasting

Trend fuzzy sets like *falling less steeply, rising more steeply* etc are used in [4] to build fuzzy rules in the form:

If trend is F then next point is Y,

where F is a trend fuzzy set and Y is fuzzy time series values. A prediction using these trend fuzzy sets is performed using the Fril evidential logic rule.

Qualitative forecasting methods use the opinions of experts to subjectively predict future events [13]. These methods are usually used when historical data either are not available or are scarce, for example to forecast sales for a new product. For example, a "Growth" stage in the product life cycle can be given as follows [13]: *Start Slowly, then Increase Rapidly, and then Continue to Increase at a Slower Rate.* This perception based description is subjectively represented as *S*-curve, which could then be used to forecast sales during this stage. In [11] it was shown how to apply the technique of granulation of convex-concave patterns and methods of reconstruction of fuzzy perception based functions [5] to modeling such qualitative expert forecasting. The results of fuzzy forecasting can be retranslated further in linguistic form [6,12].

Perception based terms are extensively used in weather forecasting. Below are examples of forecasting texts [33, 36]: *Scattered thunderstorms ending during the evening; Occasional showers possible after midnight.*

7 Conclusions

Human decision making in meteorology, economics, finance, petroleum industry etc is often based on analysis of time series data bases. The important types of problems related with analysis of time dependent systems are real time system monitoring, summarization and forecasting. Human decision making procedures use expert knowledge usually formulated linguistically and containing perceptions defined on TSDB domains. A general methodology of computing with words and perceptions developed by Zadeh [38-40] can serve for development of perception based decision making systems in TSDB domains. For development of such systems it is necessary to develop models and methods of perception based time series data mining supporting decision making procedures of CWP. Data mining is often declared as a methodology for knowledge extraction from data bases but what formal methods to use for integration of this knowledge with knowledge based decision support systems is an open problem. Perception based time series data mining can give the methods of extraction of linguistic information from time series data bases and representation it in the form of generalized constraints to integrate it with the models of computing with words and perceptions. Different features of such methods were outlined in this paper.

Acknowledgements. The research work was partially supported by the IMP within the projects D.00322 and D.00413.

References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proc. ACM SIGMOD Conf., Washington, USA, 1993, 207-216.
[2] R. Agrawal, G. Psaila, E. L. Wimmers, M. Zait, Querying shapes of histories, in: Proc. VLDB '95, Zurich, Switzerland, 1995, 502-514.
[3] J. F. Allen, Maintaining knowledge about temporal intervals, Comm. ACM 26(11), 1983, 832-843.
[4] J.F. Baldwin, T.P. Martin, J.M. Rossiter, Time series modeling and prediction using fuzzy trend information, in: Proc. Int. Conf. SC Information/Intelligent Syst., 1998, 499-502.
[5] I. Batyrshin, Construction of granular derivatives and solution of granular initial value problem, in: Fuzzy Partial Differential Equations and Relational Equations. Studies in Fuzziness and Soft Computing, Vol. 142, Springer-Verlag, 2004, 285-307.
[6] I.Z. Batyrshin, On reconstruction of perception based functions with convex-concave patterns, in: Proc. Int. Conf. Comp. Intell. ICCI 2004, Nicosia, North Cyprus, 2004, 30 – 34.
[7] Batyrshin I., Herrera-Avelar R., Sheremetov L., Panova A. Moving approximation transform and local trend associations in time series data bases. In: Perception-based Data Mining and Decision Making in Economics and Finance. Series: Studies in Computational Intelligence , Vol. 36. 2007, Springer, 55-83.
[8] I.Z. Batyrshin, L.B. Sheremetov, Perception based associations in time series data bases, in Proceedings of NAFIPS 2006.
[9] I.Z. Batyrshin, L.B. Sheremetov, Perception based constraints and associations in time series data bases, Int. J. Approximate Reasoning (submitted).
[10] I. Batyrshin, L. Sheremetov, Towards perception based time series data mining, in: BISCSE'2005, University of California, Berkeley, USA, 2005, 106-107.
[11] Batyrshin I., Sheremetov L. Perception based functions in qualitative forecasting. In: Perception-based Data Mining and Decision Making in Economics and Finance. Series: Studies in Computational Intelligence , Vol. 36. 2007, Springer, 119-134.
[12] I. Batyrshin, M. Wagenknecht, Towards a linguistic description of dependencies in data, Int. J. Appl. Math. Comp. Science 12, 2002, 391–401, http://matwbn.icm.edu.pl/ksiazki/amc/amc12/amc1235.pdf
[13] B.L. Bowerman, R.T. O'Connell, Time series and forecasting: an applied approach, Mass. Duxbury Press, 1979.
[14] J.T. Cheung, G. Stephanopoulos, Representation of process trends. Part I. A formal representation framework, Computers and Chemical Engineering, Vol. 14, 1990, 495-510.
[15] G. Das, K.I. Lin et al, Rule discovery from time series, in: KDD, 1998, 16-22.
[16] M. De Cock, C. Cornelis, E.E. Kerre, Elicitation of fuzzy association rules from positive and negative examples, Fuzzy Sets and Systems 149 (1) (2005) 73-85.
[17] D. Dubois, H. Prade, T. Sudkamp, On the representation, measurement, and discovery of fuzzy associations, IEEE Trans. Fuzzy Systems, Vol. 13, 2005, 250-262.

[18] D. Hand, H. Manilla, P. Smyth, Principles of Data Mining, MIT Press, Cambridge, 2001.
[19] F. Höppner, Learning temporal rules from state sequences, in: IJCAI Workshop on Learning from Temporal and Spatial Data, Seattle, USA, 2001, 25-31.
[20] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summarization of time series by using the Choquet integral, In this book.
[21] J. Kacprzyk, S. Zadrożny, Data mining via linguistic summaries of data: an interactive approach, in: IIZUKA'98, Iizuka, Japan, 1998, 668-671.
[22] E.Ju. Kandrashina, L.V. Litvinzeva, D.A. Pospelov, Representation of Knowledge About Time and Space in Intelligent Systems, Moscow: Science, 1989 (in Russian).
[23] KDnuggets: Polls: Time-Series Data Mining (Nov 2004). What Types of TSDM You've Done? http://www.kdnuggets.com/polls/2004/ time_series_data_mining.htm
[24] K.B. Konstantinov, T. Yoshida, Real-time qualitative analysis of the temporal shapes of (bio) process variables, American Inst. Chem. Eng. Journal, Vol. 38 (11), 1992, 1703-1715.
[25] K. Kukich, Design of a knowledge-based report generator, in: Proc. 21st Annual Meeting of the Association for Computational Linguistics, 1983, 145-150.
[26] M. Last, Y. Klein, A. Kandel, Knowledge discovery in time series databases, IEEE Trans. SMC, Part B, Vol. 31, 1, 2001, 160-169.
[27] V. Novak, I. Perfilieva, A. Dvorak, G. Chen, Q. Wei, P. Yan, Mining pure linguistic associations from numerical data, Int. J. Approximate Reasoning (submitted).
[28] Lin, J., Keogh, E., Lonardi, S., Chiu, B. (2003) A symbolic representation of time series, with implications for streaming algorithms. Proc. 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA.
[29] W. Maner and S. Joyce, WXSYS: Weather Lore+ Fuzzy Logic = Wether Forecasts, Present. CLIPS Virtual Conf., 1997, http://web.cs.bgsu.edu/maner/wxsys/wxsys.htm
[30] W. Pedrycz, M.H. Smith, Granular correlation analysis in data mining, Proc. IEEE Int. Fuzzy Systems Conf., Korea, 1999, III-1235 – III-1240.
[31] StockReporter, http://www.ics.mq.edu.au/~ltgdemo/StockReporter/about.html.
[32] T. Sudkamp, Refinement of temporal constraints in fuzzy associations, Int. J. Approximate Reasoning (submitted).
[33] The Weather Channel, http://www.weather.com/index.html.
[34] H. Toivonen, Discovery of Frequent Patterns in Large Data Collections, PhD Thesis. University of Helsinki, Finland, 1996.
[35] R.R. Yager, On linguistic summaries of data, in: G. Piatetsky-Shapiro and B. Frawley, eds, Knowledge Discovery in Databases, MIT Press, Cambridge, MA, 1991, 347-363.
[36] Yahoo! Weather, http://weather.yahoo.com/.
[37] J. Yu, E. Reiter, J. Hunter, C. Mellish, Choosing the content of textual summaries of large time-series data sets, Natural Language Engineering (2007) (To appear),
[38] L.A. Zadeh, From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions, IEEE Trans. Circuits and Systems - 1: Fundamental Theory and Applications, Vol. 45, 1999, 105-119.
[39] L.A. Zadeh, Toward a perception-based theory of probabilistic reasoning with imprecise probabilities, Journal of Statistical Planning and Inference, vol, 105, 2002, 233 –264.
[40] L.A. Zadeh, Precisiated Natural Language (PNL), AI Magazine, Vol. 25, 2004, 74-91.
[41] L.A. Zadeh, M. Nikravesh, Perception-based intelligent decision systems, in: AINS; ONR Summer 2002 Program Review, July 30 - August 1 (UCLA 2002).
[42] L.A. Zadeh, Generalized theory of uncertainty (GTU)—principal concepts and ideas, Computational Statistics & Data Analysis 51 (2006) 15 – 46.

Data Mining for Fuzzy Relational Data Servers

A.P. Velmisov[1], A.A. Stetsko[2], and N.G. Yarushkina[3]

[1] 32, Severny Venets street, Ulyanovsk, 432001, Russia, UlSTU
sasa@ulstu.ru
[2] 32, Severny Venets street, Ulyanovsk, 432001, Russia, UlSTU
ppnr@ulstu.ru
[3] 32, Severny Venets street, Ulyanovsk, 432001, Russia, UlSTU
jng@ulstu.ru

Abstract. Methods of Perception-based Data Mining (Data Miners DM) for fuzzy relational data servers are considered in the article. The considering problems of DM are limited by problems of clustering and mining of dependences in the form of fuzzy rules because these problems are especially important in practice. The hybrid algorithm of fuzzy clustering and the way of use of a fuzzy inference system as a DM for fuzzy relational data is offered in the article.

Keywords: Perception, Data Mining, fuzzy data server, fuzzy clustering.

1 Introduction

The modern level of automation leads to forming the data bases of large volumes at enterprises and organizations. The laws hidden in data bases permit to ensure optimization of enterprises and organizations activities, support to decision making. But the application of perception-based DM is difficult because for a specific problem of some subject field a sampling is often characterized not by the numeral character of attributes. In particular, in macroeconomic, sociological, marketing, medical data bases the linguistic form of data representation is widely used. To operate such data, it is necessary to use fuzzy data base. At present there is a number of theoretic and practical developments which permit to use clustering procedures for data storages. This is represented in the works [1,2,3] in which perception based intelligent decision systems are proposed; [4] in which perception based time series data mining is investigated; [5] where inquiries in hybrid clustering algorithm using two level of abstraction are investigated; [6,7] in which a fuzzy relational data server is developed and driven to a real embodiment. In spite of considerable results achieved in intelligent data analysis there is a number of undecided problems. So, use of nonnumerical attributes, including fuzzy ones, is not supported by the developed methods of analysis. Fragmentariness of sampling and nonnumerical attributes do not permit to express an object content in the traditional mathematical form of equations or formulas and presuppose use of rules (knowledge) bases. There are no methods of objective adjustment of the membership function . Thus, development of new methods of perception-based DM, modernization of existing methods and working

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 220–229, 2007.
springerlink.com

out the DM for fuzzy relational data servers is an urgent problem. It is natural to build them on the basis of fuzzy hybrid systems.

2 The Formulation of the Problem of Generation of Fuzzy Rules on Base Clustering Procedures

Let us present the method which permits to generate new fuzzy rules do not contradicting existing rules proceeded from the analysis of experimental knowledge about an object.

Let us assume that the object being investigated has n inputs (the vector input x) and has one output. Let us assume that the experiment can be realized at the object which consists in registration of N pairs of values (x_i, y_i), $i = 1,...,N$ the value of N allows modification if necessary. The algorithm of construction of the system can be described in the following way.

Step 1. The initial knowledge base of the model is made of $m, (m < N)$ arbitrary values (x_i, y_i) and this base is being mapped by the matrix $U_{m\times(n+1)}$ with rows of the form $(x_i^T, y_i) = (x_{i1},...,x_{in}, y_i)$.

Obviously, such representation is equivalent to the set of fuzzy rules of the form:

The rule i: if x_1 is A_{i1} AND ... AND if x_n is A_{in} THEN $y=y_i$ i=1,...,m,

where A_{ij} is fuzzy label for x_{ij}.

Step 2. For each new experimental point (x, y) the value being forecasted is calculated by the formula corresponding to the centroid method

$$\hat{y} = \frac{\sum_{i=1}^{m} y_i \varphi(\|x - x_i\|)}{\sum_{i=1}^{m} \varphi(\|x - x_i\|)}$$

, where $\varphi(\ )$ is the function of a bell-shaped or exponential form, for example, $\varphi(\|x - x_i\|) = e^{-\lambda \sum_{j=1}^{n} |x_j - x_{ij}|}$, λ is the parameter of the function.

Step 3. The inequality $|y - \hat{y}| > d$ is checked, where d is approximation error. When the inequality is true, the knowledge base of the system fills up by means of extension of the matrix U by adding the row (x, y) into it. Otherwise the matrix remains without changes.

Step 4. The stop rule is checked. In this variant of the algorithm the construction of the model is considered to be completed if, according to steps 2 and 3, all N experimental points are looked through. If not all the experimental points are looked through, steps 2 and 3 are repeated, else - stop.

During algorithm realization the parameters λ and d are considered to be defined beforehand. When using the model, the matrix U is also considered to be defined.

It is not difficult to see that the described algorithm, in the main, corresponds to the simplified algorithm of a fuzzy logical deduction, but differs from the latter by the fact that the knowledge base does not remain fixed, but modernizes itself as the experimental data are coming. And the consistency of the new rule with respect to the set of rules from the knowledge base is guaranteed by the offered procedure of its filling up.

3 The Model and Realization of the Fuzzy Relational Data Server

Methods of Data Mining for fuzzy relational data servers are considered in this section. Clustering algorithms discussed in Sections 4 and 5 are related to fuzzy data base. Fuzzy relational data model is defined in this section. The fuzzy relational model of data is offered, and the following requirements and limitations to the model are determined. Let the finite set of names of attributes $\{A_1, A_2, ..., A_n\}$ is called the scheme of relation R. Each attribute name A_i is associated with the set D_i, which is called the domain of the attribute A_i, $1 \le i \le n$. Domains are arbitrary nonempty finite or countable sets. And let $D = <D_1, D_2, ..., D_n>$.

Let us call the domain of an attribute of the relation fuzzy if the following is determined for it:

1. the name of the attribute A_i;
2. the universal set X;
3. the terminal set of values T, which are fuzzy labels.

Let us call the finite set of mappings $\{t_1, t_2, ..., t_p\}$ from R into D a fuzzy relation if at least one $ti \in D_i$ and the D_i is a fuzzy domain. The model is intended for representation of fuzzy numbers. Hence, the set of real numbers is the domain of the attribute of a fuzzy number. A fuzzy number is determined on the basis of:

1. the membership function ;
2. assessment of linguistic term.

Let us interpret the linguistic assessment as one of possible values of linguistic variable, which is determined by the corresponding term.

The membership function , which defines a fuzzy number, satisfies the following properties:

1. restriction (by definition $max(\mu(x)) \le 1$, $min(\mu(x)) \ge 0$);
2. continuity (the function limit exists in each point).

If for some argument the value of the membership function is indeterminate, it is assumed that this value is equal to zero.

For representation of exact (precise) values, the degenerate form of the membership function is used, returning 1 for an exact value being represented and 0 for all the rest values.

In the model the table way of definition of the membership function is taken. The quantity of pairs $\mu(x_i)/x_i$ defining the membership function is unlimited. It is assumed that the set of such pairs defines the broken line which is the plot of the membership function . Such approach is well coordinated with data representation in relational model. In the developed model the membership function is defined in general case subjectively depending on the problem being solved. No other restrictions are imposed on the form of the membership function .

In the developed model, besides the representation of fuzzy data, mechanisms of their processing are also established: functions of one argument (defuzzification, normalization, alpha-cut); set operations (complement, intersection, union); arithmetical operations (addition, subtraction, multiplication, division); operations of comparison.

Set operations are realized using the functions of minimum and maximum because the subjective character of membership functions is assumed. Arithmetical operations are based on the Zade generalization principle:

$$C = A \nabla B = [\sup \min(\mu_A(a), \mu_B(b))] / [a \nabla b],$$
$$a \in S_A,\ b \in S_B,$$

where S_A, S_B, μ_A, μ_B are correspondingly carriers and membership functions of fuzzy numbers A and B, ∇ is one of four arithmetical operations.

The operation of comparison is the base in the mechanism of fuzzy data processing because the definitions of relational operators are based on it.

We can use above described fuzzy relational data server to store fuzzy attribute but important question exist. What clustering procedure can be used for DM in fuzzy data base?

4 The Hybrid Algorithm of Fuzzy Clustering

The hybrid algorithm is offered for effective clustering. This algorithm uses the new function of assessment of separability and compactness.

Using the method *Maxmin* the initialization of separations is done by means of separation of objects as much as it is possible. Then, by means of optimization, a local optimum is found. Each center of a cluster is used as the initial point. Then the transition to the next iteration follows. The algorithm is executed till the convergence of the algorithm or while the defined quantity of iterations is not reached.

After the optimum decision for the chosen quantity of clusters c is found, we find the separation c-1 using the algorithm of merging.

Only several algorithms of the great number of existing ones are oriented on the clustering procedure of dense points. These methods assume that the number of clusters and/or some threshold values, change of which has a strong influence on the result, is defined by a user.

The new function of assessment is offered, which is applicable when the quantity of clusters is very big.

The new hybrid algorithm of clustering is offered. In this algorithm the optimized *Maxmin* method in combination with the merging strategy is applied in such a way that it is always possible to form optimum variants for a variable quantity of clusters. Then the optimum variant of clustering is chosen with the help of the function of assessment, which is based on the measures of separability and compactness.

4.1 The Function of Assessment

The function of assessment is based on the measures of separability and compactness of clusters.

Definition 1. The set of clusters $C = \{C_1, C_2, ..., C_c\}$ is given for the set of objects $X = \{x_1, x_2, ..., x_N\}$. Let $C' = \{C_i\}$, where $C_i \in C$ and C_i is not single set. Let k is quantity of objects belonging to C'.

The compactness CP of the set of clusters C' is defined as:

$$CP = \frac{k}{\sum_{i=1}^{k}\left(\sum_{x_j \in C_{pi}, x_j \neq r_i} \mu_i(x_j)^2 d(x_j, r_i)^2 / \sum_{x_j \in C_{pi}, x_j \neq r_i} \mu_i(x_j)^2\right)}, \text{ where}$$

$\mu_i(x_j)$ is the value of the membership function x_j to C_i, r_i is the center of C_i, c is the quantity of clusters, $2 \leq c \leq N$, $d(x_j, r_i)$ is the distance between x_j and r_i.

Definition 2. The separability of the set $C = \{C_1, C_2, ..., C_c\}$ for the set of objects $X = \{x_1, x_2, ..., x_N\}$ is defined as:

$$SP = \left(\frac{\sum_{1 \leq j \leq c, i \neq j}^{c} \sum_{j+1 \leq i \leq c}^{c} \min\{d(r_i, r_j)\}}{c}\right)^2,$$

where c is the quantity of clusters, r_i is the center of the i-th cluster, $d(r_i, r_j)$ is the distance between r_i and r_j.

Definition 3. The set of clusters $C = \{C_1, C_2, ..., C_c\}$ is given for the set of objects $X = \{x_1, x_2, ..., x_N\}$. Let $C' = \{C_{pi}\}$, where $C_{pi} \in C$ and C_{pi} is not a unit set,

$i = 1,2,...,k$, where $k = C'$. The separability/compactness SP of the set of clusters C is defined as $SC = \frac{k}{c} \times SP \times CP$.

The aim of the algorithm to find the set of clusters satisfying $\max_{2 \le c \le N} \left\{ \max_{\Omega_c} \{SC\} \right\}$, where Ω_c defines all the sets which are candidates for the definite quantity of clusters c.

4.2 A Median of the Set and the Membership Function

Let us define the concept of a median instead of a mean value for a subset of the data set. Let C_i is the subset of the data set with the function of the distance d. The point $x_0 \in C_i$ is called the median of the C_i if $\sum_{y \in C_i} d(x_0, y) = \min_{x \in C_i} \left\{ \sum_{y \in C_i} d(x, y) \right\}$.

Let $X = \{x_1, x_2, ... x_N\}$ is the data set, let r_j is the center of j-th cluster, $j = 1,2...,c$. Let us define the membership function μ_{C_j}, $j = 1,2,...,c$, for $\forall x \in X$ as:

$$\mu_{C_j}(x) = \begin{cases} 1 \ \text{if } d(x, r) = 0, \\ 0 \ \text{if } d(x, r_k) = 0, k \neq j, \\ \left(\sum_{v=1}^{c} \frac{d(x, r_j)}{d(x, r_v)} \right)^{-1} \quad else. \end{cases}$$

Thus, $\sum_{j=1}^{c} \mu_{C_j}(x) = 1, \forall x \in X$ and $\sum_{k=1}^{N} \mu_{C_j}(x_k) \le N$, $j = 1,2,...,c$.

The fuzzy separation can be transformed to a precise one in the following way:

$$\mu_{C_j}(x_k)_{hard} = \begin{cases} 1 \ \text{if } \mu_{C_j}(x_k) = \max_{1 \le v \le c} \{\mu_{C_v}(x_k)\} \\ 0 \ else. \end{cases}$$

It is obvious that $\mu_{C_j}(x) = 1$ if r_j is the nearest center to the point x_k.

4.3 The Algorithm of Merging

The algorithm of merging applies the measure of similarity for choosing the most resembling pairs of clusters. In the algorithm of merging the most "bad" cluster is chosen and then it is removed. All the elements which belong to this cluster are moved to the "nearest" cluster, after that redetermination of clusters centers takes place

Definition 4. The set of clusters $C = \{C_1, C_2, ..., C_c\}$ for the set of objects $X = \{x_1, x_2, ..., x_N\}$ is given. For each $C_i \in C$, if C_i is not a unit set, let us denote the compactness C_i by the cp_i and define the cp_i:

$$cp_i = \frac{1}{\sum_{x_j \in C_i, x_j \neq r_i} \mu_i(x_j)^2 d(x_j, r_i)^2 \Big/ \sum_{x_j C_i, x_j \neq r_i} \mu_i(x_j)^2}$$

, where $\mu_i(x_j)$ is the value of the membership function x_j to the cluster C_i, r_i is the center of the cluster C_i, c is the quantity of clusters and $2 \leq c \leq N$.

Definition 5. The set of clusters $C = \{C_1, C_2, ..., C_c\}$ for the set of objects $X = \{x_1, x_2, ..., x_N\}$ is given. For each $C_i \in C$, if C_i is not a single set, let us denote the separability C_i by sp_i and define the sp_i:

$$sp_i = \left(\min_{1 \leq j \leq c, i \neq j} \{d(r_i, r_j)\} \right)^2$$

, where r_i is the center of the cluster C_i, r_j is the center of the cluster C_j, c is the quantity of clusters and $2 \leq c \leq N$.

Definition 6. The set of clusters $C = \{C_1, C_2, ..., C_c\}$ for the set of objects $X = \{x_1, x_2, ..., x_N\}$ is given. For each $C_i \in C$, if C_i is not a single set, let us denote the separability/compactness C_i by sc_i and define the sc_i in the following way:

$sc_i = sp_i \times cp_i$. Thus, the "worst" cluster will have the least value sc_i.

4.4 The Hybrid Algorithm of Clustering

The entry is the data set $X = \{x_1, x_2, ..., x_N\}$, the maxnum is the maximum quantity of clusters.

The exit is the optimum set of clusters $C = \{C_1, C_2, ..., C_c\}$.

Step 1. $c_{opt} = \max num$, $c = \max num$, $i = 1$. At random the object $x \in X$ is chosen as the start point p. The difficult *maxmin* algorithm is executed with parameters X, c, i, p for the search of the optimum set of clusters $C = \{C_1, C_2, ..., C_c\}$ for c.

The function of assessment SC is calculated for C.

Step 2. The algorithm of merging is executed for getting the set of clusters $C' = \{C'_1, C'_2, ..., C_c\}$, the center C'_1 is chosen as the start point p, $c = c - 1$, $i = 2$. The *Maxmin* algorithm with parameters X, c, i, p for search of the optimum set of clusters $C^* = \{C^*_1, C^*_2, ..., C^*_c\}$ for c. The median is calculated for C^*_i as new center r_i. The function of assessment SC is calculated for C^* and it is taken as SC^*. If $SC^* > SC$, then $SC = SC^*$, $C = C^*$, $c_{opt} = c$. Repeat step 2 while $c \geq 2$.

Step 3. The conclusion: the $C = \{C_1, C_2, ..., C_{opt}\}$ is the optimum set of clusters.

This algorithm has a number of advantages over other algorithms of clustering.

Table 1. The comparison of algorithms of clustering

algorithm	High level of grouping for input data	Necessity of indication of clusters quantity	Sensitivity to input parameters	Applicability to nonuniform data
Hybrid algorithm	Yes	No	No	Yes
k-means	Yes	Yes	Yes	Yes
Substractive	Yes	No	Yes	No
Maxmin	Yes	No	Yes	Yes
Fuzzy c-means	No	Yes	Yes	Yes

5 The Result of Clustering of a Sociological Data Base

In 2000 the Institute of Social-economic Problems of Population attached to the Russian Academy of Sciences developed the questionnaire "The social-economic situation of population of Russia". The questionnaire contains such sections as "General data", "Employment", "Social mobility", "Social status", "Political activity", "Social-demography structure of household", "The level of life of household", "Expenses", "Property", "Conditions of living", "Private subsidiary economy (for rural population), a garden plot (for townspeople)". The sampling of data for the analysis is represented by results of the survey in Dimitrovgrad of the Ulyanovsk region.

The aim of this questionnaire was study of a social-economic state of the population of Russia. These data are necessary for correction of economic policy of Russia, development of recommendations for measures aimed at increasing the level of life of Russian citizens. The questionnaire consists of 129 items (438 questions). 1080 people were chosen at random for conduction of the survey.

It is not possible to give exact answers to some of the questions, for example, question 46 "How often do you work at home in the evening and at the weekend?" The variants of answers are "often", "rarely", "never". Clustering of such data is complicated because, when dividing into classes, it is necessary to unite different linguistic labels. The data got on the basis of the survey were dipped into the fuzzy relational data server. As a result the data suitable for automated information processing using the DM were got. The fuzzy labels were associated with the membership functions.

The data were dipped into the fuzzy data server and studied with the help of the developed DM.

When clustering by subjective attributes, we get the following results:

Table 2. Results of clustering

Question	Class 1	Class 2	Class 3
The social status of the family	Middle	Low	Low
The level of material well-being*	4,20	2,73	2,35
The degree of professionalism*	6,09	4,76	3,29
Adaptation to an economic situation	5,34	3,07	1,88
Dependence of well-being on individual efforts	6,06	3,27	8,00
The social status*	4,72	3,17	2,88
The society stratum for your family	Middle stratum	Below the middle	Middle stratum
Will your life change?	Will not change	Will probably grow worse	Will probably grow worse

* according to a 9-point scale

The class of people of the middle status is evidently notable, they are well provided for, can work well in their field, well adapted and believe that the situation in the country will not grow worse. Representatives of the third class suppose that the status of their family and the level of providing is low, the adaptation to the situation is very low, and they know that their status is completely determined by their own efforts. They assess their qualification as low and think that the life will grow worse. The second class looks like the third one with the difference that its representatives consider themselves a little more adapted, they are confident in their qualification, though do not believe in their strength.

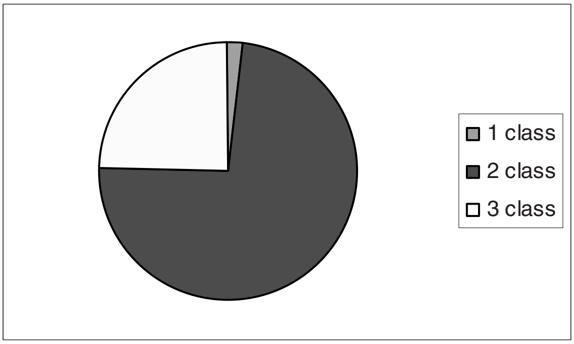


Fig. 1. Separation into classes

6 Conclusion

Using the mathematical apparatus described above, the Perception-based DM for fuzzy relational data bases with fuzzy attributes has been realized. The software product permits to perform the clustering of fuzzy data and disclose dependences in the form of fuzzy rules. The represented algorithm permits to conduct the clustering of nonuniform data, it is not sensitive to input parameters and does not require the indication of clusters quantity. The developed software product can be effectively used for processing of sociological, medical, macroeconomic, marketing and other data bases which assume the storage of data in the form of linguistic labels and fuzzy attributes.

References

1. Zadeh L.A., Nikravesh M. Perseption based intelligent decision systems. In: AINS; ONR Summer 2002 program Review, July 30-August 1(UCLA2002)
2. Hoppner F, Klawonn F., Kruse R., Runkler T. Fuzzy Cluster Analysis, J. Wiley, Chichester, 1999
3. Pedrycz W. Knowledge-based Fuzzy Clustering, , J. Wiley, N. York, 2005
4. Batyrshin I.Z, Sheremetov L.B. Perception based associations in time series data bases. In: NAFIS 2006, Canada, Montreal,2006
5. Ghosh R., Mammadov M., Ghost M., Yearwood J. A Hybrid clustering Algorithm using two level of abstraction. In: IFSA 2005, China, Beijing, V. 1, pp. 586-591
6. Yarushkina N., Semoushin I. A Soft Computing-based Integration Environment for Assessing the Performance of a Complex Enterprise. In: IFSA 2005, China, Beijing, V. III, pp. 1430-1435
7. Yarushkina N. A Soft Computing-based Integration Environment for financial analysis of a complex enterprise. In: BISCSE'2005, University of California, Berkeley, USA,2005, pp. 67-68

Computational Intelligence Models of the Distributed Technological Complexes

Andriy Sadovnychyy[1], Sergiy Sadovnychiy[2], and Volodymyr Ponomaryov[1]

[1] National Polytechnical Institute, ESIME Culhuacán. Av. Santa Ana No. 1000, Col. San Francisco Culhuacan, Mexico, D.F., 04430, Mexico
asadovny@cablevision.net.mx
[2] Mexican Petroleum Institute, Pipeline Research Program. Eje Central Lázaro Cárdenas Nte.152.Col. San Bartolo Atepehuacan, 07730, Mèxico, D.F., México

Abstract. The article deals with simulation of the distributed technological complexes. Visual simulation structure of manufacture is offered. On base of visual model the simulation model is formed automatically. The simulation model is described by means of the mathematical apparatus of Multitask Presence Logic (MAL). The basic concepts of MAL and some of examples of their application for construction of manufacture model are given. The knowledge base structure is designed on the basis of frame models for problems of decision-making support. The application of frame models allows to change structure base of knowledge directly during of simulation. The knowledge base is designed for use of linguistic variables and, accordingly, a linguistic data input/output.

1 Introduction

The dynamics of the market relationship (cycle of a commodities circulation, steady variation of branch structure, diversification, renewal of the products nomenclature), integration of international business and the resource limitations are result of essential speed increment of material and financial flows. One of the important directions of manufacturing logistics area is the creation of system model focused on the multialternative analysis of the new schemes of organisation and control by computerised processes, which one flexibly and effectively provides interaction of basic elements of logistics circuit: « deliveries - manufacturing - warehousing - transportation - selling».

The developing of a new class of models based on system submissions of industrial - technological complex (TC), knowledge-orientation approach and imitating simulation of processes within the framework of industrial logistics of virtual productions is indispensable.

2 Systems for Simulation of Technological Complex Control

The methods of formalisation of submission and simulation of manufacturing systems constantly increase because the existing tasks are diverse and their mathematical description is investigated insufficiently. Therefore, at development of the system simulation model for the analysis of manufacturing systems there is necessity in objects

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 230–239, 2007.
springerlink.com

structurization (partition on subsystems and elements) and determination of the mathematical schemes and model approximation for description of separate elements [2]. For solving the problems it is necessary to use the methods of extraction of knowledge about the system and data structure in the linguistic form that can be further transformed in formal models. The methods of computing with words and perceptions [3] and perception based data mining [4] give one of methods of solving of this problems.

It is expediently to conduct classification of processes for finding-out of formalisation methods. It is possible to classify five basic categories of processes: productions; distributive processes; processes of service; control procedures (acceptance of administrative solutions); control procedures by the project of the distributed industrial complexes creation.

Such classification, certainly, does not mean that all processes precisely fall in one of these categories. For example, the fulfilment of the orders on manufacture can include as operation of an orders reception, as operation of their fulfilment. In this case, the orders reception is process of service, and fulfilment of the orders is manufacture process.

The internal representation of the standard process includes ensemble of: processes providing functioning of object; elementary operations which are forming the processes; events influencing on fulfilment of processes; the executors participating in operations; resources used for fulfilment of operations; objects used by operations. Therefore, the construction of an imitating model of the logistic analysis of TC functioning and control processes consists of two stages: structurization and formalisation of model.

3 Structurization Idea

The modular concept of structurization is used. All blocks have two presentation layer. The user's level is visual and the simulation system level is mathematical. On the visual level the blocks are represented by graphics objects and by linguistic constructions. Graphics mapping and linguistic constructions vary depending on user's categorial set. For each categorial set the program complex offers the library of visual images of standard blocks with customized behaviour. The icons of cross-linking modules connected by graph (graphic links) display TC structure. The capability of simple manipulation by graphics objects representation in a system provides base means for construction of the problem-oriented language of visual programming. The standard items library for simulation of virtual manufacturing systems and processes includes following modules: the supplier; manufacturing subdivision; the consumer; service; conveyer; storage area; transportation; integration; separation; packaging; copying; synchronization etc. Each unit has a graphics image, the appropriate name and a set of parameters in linguistic form which characterize the displayed object. In the model of a logistic circuit the specialized standard module "transportation" is formed. This module simulates all processes occurring with a hauling unit or cargo on the given route segment. In this module the transportation cost, actual time on the way and different kinds of costs are calculated. The transportation process can have periodic stop blocks, transfer of cargo, and change of transport types. For simulation of transport logistics tasks the system is supplemented by such standard modules as: sources of hauling unit; outlet of hauling unit; service stations ; distribution centers, warehouses; terminals. The modules simulate storage, processing, consolidation and transfer of cargo on intermediate points. They also are displayed graphically and have the characteristic parameters represented in the linguistic form. The figure 1 shows the block diagram of a standard module of the model. The composition of functional units depends from unit purpose. The adjustment of the composition of the functional modules is realized by the inheritance from the standard unit of some structural properties, attributes and model of behaviour [5].

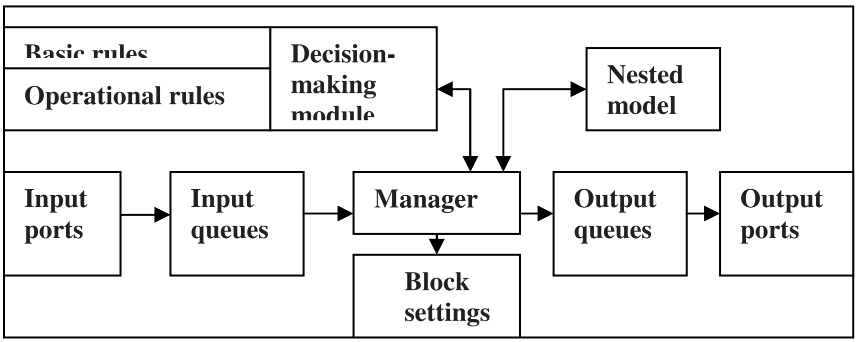


Fig. 1. The model's core module block diagram

4 Formalization Idea

It is necessary to note, that for formation of analysis model of the TC control and functioning processes there is an uncertainty in the description of its individual elements, which one is appeared in following [5]:

- Incompletness and inaccuracy of information about state of a system object and environment;
- Presence of random factors;
- Presence of the factors dependent from of other objects (systems) behavior;
- Change of external and internal environment, which one makes impossible exact prediction of the objects state and therefore, prediction of results of the received solutions;
- Complexity and multiplicity of processes and conditions of system object functioning.

The description uncertainty of the system objects can be reduced by the introducing of limitations and assumptions that is decrease of a formalization degree of real processes. But it can result in violation of adequate description of a physical reality. The declarative form of a knowledge representation in the form of Formalized Axiomatic Theories (FAT), based on the Multivalued Presence Logic (MAL) and the Category Theory was used in the system as an universal method of formalization of the models members. Use of MAL allows to describe of uncertainty and contradictions arising in the process of planning and control of a distributed TK.

The logic models of knowledge representation based on MAL, have the series of advantages, basic of which are following:

1. All "truthes" in logic calculus of “presence” are examined as relative, instead of absolute, as in the classic quantificational logic, because in physical reality the "truth" depends on a position of objects in space - time, current conditions and other parameters.

2. The "truthes" in the “presence” logic are esteemed during their formation and development, while in the classic quantificational logic they are static.

3. The "truthes" concerning a current condition of a reality, have unequivocal pragmatical interpretation, i.e. it is possible beforehand to specify, what operatings should undertake the system in view of its purposes.

This categorical-theoretic approach allows to examine the TC as a complex dynamic system which consist from various nature objects and thus, it can be by alternative of the known theory of sets [6]. At the same time, this approach provides:

1. Presentation by the same way the abstract and particular concepts (properties, states, processes etc.) as the objects of categories of classes.

2. Expression by means of the morphisms of a category of any links among objects (embedding of the one type of objects into others, and also a functional (role) imagings).

3. Presentation of the dynamics of a system operation by means of the categories and by adding of the special morphisms of change (the differential of change of “presence”).

The formalized theory T = {Z, C, A} is named a formal system {L, C} with the set, adjoined to it of the formulas A, called by nonlogical axioms which recorded in language Z. The consequence operation C = { $\bigcap$, $\bigcup$,~} allow to execute a theorem making in T. The peculiarity of this logic are that the formulas designating objects (concept, property, processes etc.), can take on the value S - “present”, or the value N - “absent” (as against the classic logic in which one the objects can take on the value "true", or the value "false"). On the set of properties of X object the Boolean algebra of “presence” is assigned:

$$\{X, \bigcap, \bigcup, \Rightarrow, \sim\} \tag{1}$$

Here, the complement operation ~ is determined by equation: $x \bigcap \sim x = N$, $x \bigcup \sim x = S$, and the operation $\Rightarrow$ of relevant logical consequence by condition $x \Rightarrow y$, in case only the property x include property y, i.e. $x \in y$. Any given object is represented as the formula in language of the Boolean algebra of “presence” (1):

$$A \equiv \bigcup_{i=1}^{N} \bigcap_{m=1}^{n_i} \varphi_{ij} \tag{2}$$

where, φ_{ij} is a some property or its complement by "presence"; N , n_i are number of attributes sets and number of attributes in each set, accordingly. These properties are used for identification of object A.

Generally not all attributes are accessible for the system, therefore they can be sectioned into two classes: properties, accessible to a system - π_{ij} , and inaccessible ξ_{ir} :

$$A = \bigcup_{i=1}^{N} \left(\bigcap_{j=0}^{k_i} \pi_{ij} \right) \bigcap \left(\bigcap_{r=0}^{m_i} \xi_{ir} \right) \tag{3}$$

where $k_i + m_i = n_i$ for all i .

At the fixed values of "presence" of accessible properties and depending on values of inaccessible properties the equation (3) generates the set of equations, each of which can be esteemed as neighborhood (approximating) of an initial equation (2).

On the basis of the ordering relationship $x \geq y$ (by the contents), expressed with procedure $x \Rightarrow y$, in this set can be allocated the greatest lower and least upper bounds, which ones are adopted as "locking" C and interiority I of object A:

$$IA = \begin{cases} \bigcup\limits_{i \in Q} \left(\bigcap\limits_{j=0}^{k_i} \pi_{ij} \right), if\ Q = \{ i | \, m_i = 0 \} \ \neq \varnothing, \\ N, \ \ if\ Q = \varnothing, \end{cases}$$

and $$CA = \begin{cases} \bigcup\limits_{i=1}^{P} \left(\bigcap\limits_{j=0}^{k_i} \pi_{ij} \right), if\ P = \{ i | \, k_i = 0 \} \ = \varnothing, \\ S, \ \ if\ P \neq \varnothing. \end{cases} \tag{4}$$

Thus, in indeterminancy conditions concerning of "presence" values of a part of properties on set of the equations describing objects, the topology can be assign and it allows to pass from equation Boolean algebra to topological Boolean algebra:

$$\{X, \bigcap, \bigcup, \Rightarrow, \sim, I\}, \tag{5}$$

where the interiority I has all properties of similar operation in topology of sets.

This algebra is situated in the four-digit logic or in the multivalued logic of "presence". Taking into account, that $C \sim A \equiv \sim IA$, $I \sim A \equiv \sim CA$, for each object A, it is possible to put in conformity the vector:

$$\vec{A} = \langle IA, C \sim A, CA, I \sim A \rangle . \tag{6}$$

For values of "presence" of vector it is possible to enter following identifications:

$$\vec{A} = \langle S, N, S, N \rangle = \Pr$$ - "presence",

$$\vec{A} = \langle N, S, N, S \rangle = Ab$$ - "deficiency",

$$\vec{A} = \langle N, S, S, N \rangle = Un$$ "indeterminancy",

$$\vec{A} = \langle S, S, S, S \rangle = Cn$$ - "contradiction".

The algebra of such vectors is topological Boolean algebra. The indeterminancy value of "presence" Un corresponds to a case of the information insufficiency about object properties for definition of object presence. The contradiction Cn can serve as attribute of the theory and actuality mismatch .With use such logic it is possible to build not only theory of simple but also complex objects, including dynamic.

The theory of dynamic object can be presented by the introducing of differentiation operator of "presence" $\Delta_{\alpha\beta} x_i$. Differentiation operator $\Delta_{\alpha\beta} x_i$ takes value of presence - "present", if the value of "presence" x_i changes with α on β as a result of the object state change or of the object model change. Algebra of such differentiation operator with operation of a composition $\circ$:

$$\{X, \Delta_{\alpha\beta} \bigcap, \bigcup, \Rightarrow, \sim, \circ\}, \quad (7)$$

supplements the algebra (5).

The system of the axiomatic theories is made by axioms expressed through differentials of "presence". The ratio for operation of a composition are the conclusion rules. The formation of the axiomatic theories is made for compound objects on the basis of the theory of categories. The definition of a category K includes: Class $Ob(K)$ - sets of objects of a category and collection of sets $Hom(X,Y)$, on one for each pair of objects $X, Y \in Ob(K)$. The elements of objects are called morphism (from X in Y) and are designated $\mu : X \to Y$. The arrow means the mapping of object X in object Y.

In this case, each property or concept is esteemed as object of some category. If some concept is submitted by object Y and this concept or the property X_i is essential for definition of "presence" or "deficiency" of object Y, the category include morphism $\{\mu_i : X_i \to Y\}$. The set of such morphisms formes the "covering" of object Y for all essential properties of object Y or for all its components. The object X_i can have essential property V $\{\varphi_j : V_{ij} \to X_i\}$, too.

As far as, the compositions of morphisms also form "covering family" for Y, there is the set of "covering families" C (U), formed by compositions of morphisms of various multiplicity. They are called - covering sieves of topology. The combination of a category, specified for all object sieves forms Grotednic topology, and the category together with such topology is called as a site.

The copies of object X_i can have different values of "presence" and the morphism μ_i can be presented like the quadruple of morphisms which one can be considered as a morphism value of “presence” μ_i..

$$\mu_i : \begin{cases} IX_i \to Y = \Pr \mu_i : X_i \to Y; \\ I \sim X_i \to Y = Ab\mu_i : X_i \to Y; \\ (CX_i \cap C \sim X_i) \to Y = Un\mu_i : X_i \to Y; \\ (IX_i \cap I \sim X_i) \to Y = Cn\mu_i : X_i \to Y, \end{cases}$$

Each from the quadruple of morphisms $\mu_i : X_i \to Y$ associate with the set of axioms, containing the corresponding value of "presence" object X_i. Thus, the mapping of "covering elements" X_i in the theory of object Y is specified. Besides for each formula of the Y object theory is corresponded the set of value of "presence" object X_i, encountered in this equation as a consequence. Thus, the limitation of the Y object theory on the object X_i is specified and the bundle of the theories above a site called topos of Grotednic is formed.

The particularity of topos supplemented by "presence" logic is possibility to build the hierarchical all-level theories corresponding to hierarchical systems of objects, aggregated among themselves. Thus, it is possible to execute embedding of the theory of object of slave level in the theory of the given object and to use the received theory for seeking up of theorem proving. The received global theory has a capability of dynamic interpretation and is characterized by set of levels of complexity of objects and the set of levels of a commonality, on which one their properties are esteemed.

The values of morphisms "presence", included in the local theories, are determined by the data of programming modules and by means of an inference at interpretation of knowledge. Thus, computation of "presence" and the construction methods of the formalized axiomatic theories make the theoretical basis of knowledge representation.

5 Knowledge Base Structure

The knowledge-oriented approach of construction of a simulation model of the logistic analysis of virtual manufacturing systems consists from the three stages. On the first stage the TC is analyzed and, as the result, the formalized descriptions - FAT of main system objects are determined. At the second stage the operation algorithms and the aims of control are described. As result of this operating, the FAT of dynamics of a system behavior (situations description, control policies, changes of objects state etc.) is received. The models of this level allow to take account of processes essence which proceed with the multiple conflict situations. These conflict situations arise during of elements functioning of a manufacturing system. At the third stage the standard elements of internal simulation structure (device, queue etc.) are formed. At last stage as a model of knowledge representation the frames are used [7].

For practical use, all knowledge models are organized in a common multilevel knowledge base. In the same base, the linguistic knowledge models which described the goal, norms and assessment of TC functioning results are put.

The semantic knowledge is mapping in knowledge base headings: INT - intentional; POS - possibility; ST - object state. Intentional of object A includes the set of essential (distinctive) attributes π_i.

$$INT_A : \bigcap_{i=1}^{N} (\mu_i : \pi_i \to A) \tag{8}$$

In a POS headings the possibility of use of object A in some processes (operations) are described:

$$POS_A : \bigcup_{i=1}^{D} ((\rho_i : b_i \to r_i) \cap (\xi_i : A \cap r_i \to d_i)) \tag{9}$$

where b_i is an equation of conditions of fulfilment of a role r_i, and d_i is an attribute of a capability of fulfilment of i - th operating. In a heading ST the laws of state change S_i of the object A are described as a result of any control effect u_i:

$$ST_A : \bigcup_{i=1}^{K} ((\upsilon_i : c_i \cap u_i \to s_i) \cap (\nu_i : s_i \to A)) \tag{10}$$

where c_i is the equation of conditions of transition in a state S_i

The pragmatical knowledge in each level is allocated on specially isolated headings knowledge base. The pragmatical knowledge should provide the construction of the system purpose structures and realization of the algorithm for their achievement. The procedures of knowledge processing for the solution of problems of achievement of required aim can be described as set functorial morphism.

The detailed consideration of structure and processes occurring in TK has shown that during the simulation a set of situations under the analysis of a current state of the model elements and decision making about a further direction of simulation are arisen. For these purposes the subsystem of support of decision making was designed. This subsystem, as well as all modelling complex, can operate independently for problem solving in other application area. The subsystem of decision marking support serves for the solution of the following tasks:

- Formation of the initial data for internal structure of simulation;
- Detection of the contradictions and correction of the initial data on a course of simulation;

- Identification of a current condition and decision marking about selection of a simulation direction;
- Observation and detection of unnominal situations;
- Issue of the reference information on a course of simulation;
- Interpretation of the simulation results in the linguistic form.
- The problem-oriented simulation system has built-in library of standard cross-linking elements, broad complex of dialogue control simulation tools, a graphic, animation and linguistic means of the simulation results representation etc. The system provides following functionalities:
- Visual representation of the allocated all-level structure of a virtual manufacturing system by means of the graphic editor with any degree of detailed elaboration;
- Fixation of production structure on a digital map;
- The automated construction of a simulation model of the manufacturing analysis which exclude an intermediate phase of generation of a code and compilation.

6 Simulation Results

The simulation results of are formed during simulation. These results can be esteemed both for all system as a whole, and for each structural element. The results include: percent of fulfilment of the production schedule (for all commodity nomenclature); the sizes of probable deficits of specified nomenclature of components and commodity; failures of deliveries with the indication of the reasons of their originating; volumes of the uncompleted production; actual times of work cycles of manufacturing of each product; a time-table (schedules) of material resources movements among the logistic network nodes of a manufacturing system; a time-table of the loading of elements of manufacturing structure; costs of technological operations fulfilment; logistic costs; quantity of failures and time of idle hourses of manufacturing structure elements etc.

The decision making for production management on the base of the research results can be implemented by the following way:

- Change of nature of technological processes with simultaneous modification of a manufacturing structure;
- Change of control policies of a transport and warehouse system;
- Organization of additional actions for maintenance of indispensable productivity and throughput capacity of a system;
- Change of material flows structure, including modification of the deliveries structure both inside the company, and at interplay with the external suppliers;
- Changes of organizational structure of company management (horizontal and vertical "compression", distribution of the administrative and executor functions);
- Determination of desirable market and manufacturing policy of the company (definition of policy of the order allocation at the company, policies of purchases and selling).

The last operation is executed by a assignment of the simulation scenario with various versions of control policies as for individual elements, as for all system. The received simulation results of all scenarios can be estimated by means of a system of support of decision making by one or several criteria. The system provides the support of a base set of control policies with a possibility of definition and assigning of particular policies, use which one is required in this or that situation.

7 Conclusion

The offered simulation system allows to investigate the various technological process modes of the distributed network of a technological complex of preset architecture with accumulation of the statistical information. This system provides the data preparation for decision making in the field of control of a manufacturing system on following stages of functioning: TC creation; TC re-structuring; diversification; automation and modernization; booking replacement.

References

[1] Kusiak A. : Computational intelligence in design and manufacturing. John Wiley, (2000).
[2] Technology of system simulation. Edited by Emelianov S.V. Moscow, Mashinostroenie, (1988).
[3] L.A. Zadeh, From computing with numbers to computing with words - from manipulation of measurements to manipulation of perceptions, IEEE Trans. on Circuits and Systems - 1: Fundamental Theory and Applications, 45105-119. (1999)
[4] Batyrshin I., Kacprzyk J., Sheremetov L., Zadeh L.A. (Eds.). Perception-based Data Mining and Decision Making in Economics and Finance. Series: Studies in Computational Intelligence , Vol. 36. , Springer, Berlin / Heidelberg. (2007).
[5] Fedorovich O., Sadovnychiy S., Prohorov A. and Sadovnychyy A.: System Models of Logistics Analisis of the Virtual Distributed Technological Complexes. Journal "Electromagnetic Waves and Electronic System". vol. 8, No. 10 59-65. (2003)
[6] Bukur I., Delianu A.: The introducing in the theory of categories and functors. Moscow, Mir, (1972).
[7] Kess Y.Y. The analysis and synthesizing of the frame-based models of automatic control systems. Moscow .: Energoatom, (1986).

Part VI

Soft Computing in Medical Sciences

Similarity and Distance–Their Paths from Crisp to Fuzzy Concepts and an Application in Medical Philosophy

Rudolf Seising and Julia Limberg

Medical University of Vienna, Core Unit for Medical Statistics and Informatics, Spitalgasse 23, A-1090 Vienna, Austria
Rudolf.Seising@meduniwien.ac.at

Abstract. Similarity and distance are matters of degree for which we can find crisp and fuzzy concepts in the history of mathematics. In this paper we present Lotfi Zadeh's and Karl Menger's crisp concepts before the time of fuzzy sets. Finally we show an application in medical philosophy by Kazem Sadegh-Zadeh that extends into new theoretical research in the field of modern genetics.

Keywords: Similarity relation, distance, metrics, probabilistic metrics, t-norms, ensembles flous, fuzzy sets medical philosophy.

1 Introduction: Similarity and Distance in Signal Transmission

"Let $X=\{x(t)\}$ be a set of signals. An arbitrarily selected member of this set, say $x(t)$, is transmitted through a noisy channel Γ and is received as $y(t)$. As a result of the noise and distortion introduced by Γ, the received signal $y(t)$ is, in general, quite different from $x(t)$. Nevertheless, under certain conditions it is possible to recover $x(t)$ – or rather a time-delayed replica of it – from the received signal $y(t)$." In these words Lotfi A. Zadeh presented in 1952 a basic problem "in communication of information" to the New York Academy of Sciences. ([1], p. 201.)

Zadeh represented the formula $x = \Gamma^{-1}y$, where $\Gamma^{-1}y$ is the inverse of Γ, if it exists, over $\{y(t)\}$ corresponding to the relation $y = \Gamma x$ between $x(t)$ and $y(t)$. He represented signals as ordered pairs of points in a signal space Σ, which is imbedded in a function space with a delta-function basis, and to measure the disparity between $x(t)$ and $y(t)$ he attached a *distance function* $d(x, y)$ with the usual properties of a metric. Then he considered two special cases:

1) To achieve a perfect recovery of the transmitted signal $x(t)$ from the received signal $y(t)$. Zadeh supposed that "$X = \{x(t)\}$ consist of a finite number of discrete signals $x_1(t)$, $x_2(t)$, …, $x_n(t)$, which play the roles of symbols or sequences of symbols. The replicas of all these signals are assumed to be available at the receiving end of the system. Suppose that a transmitted signal x_k is received as y. To recover the transmitted signal from y, the receiver evaluates the 'distance' between y and all possible transmitted signals x_1, x_2, …, x_n, by the use of a suitable distance function $d(x, y)$, and

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 243–252, 2007.
springerlink.com

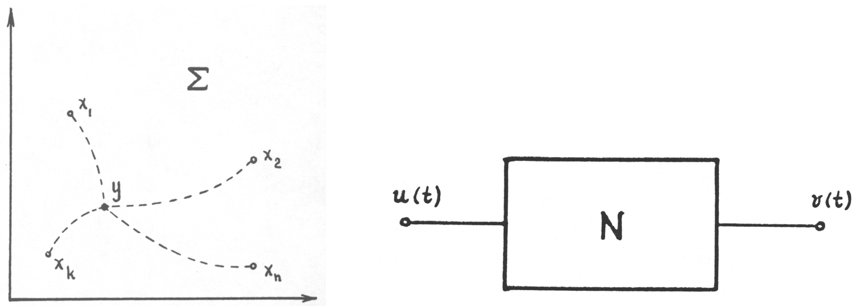


Fig. 1. Left: "Recovery of the input signal by means of a comparison of the distances between the received signal y and all possible transmitted signals." ([1], p. 202). Right: Schematic representation of a signal transmission system. ([2], p. 294).

then selects that signal which is 'nearest' to y in terms of this distance function (Figure 1). In other words, the transmitted signal is taken to be the one that results in the smallest value of $d(x, y)$." ([1], p. 201) By this process the received signal x_k is always 'nearer' – in terms of the distance functional $d(x, y)$ – to the transmitted signal $y(t)$ than to any other possible signal x_i, i.e.

$$d(x_k, y) < d(x_i, y), \quad i \neq k, \quad \text{for all } k \text{ and } i.$$

Zadeh conceded that "in many practical situations it is inconvenient, or even impossible, to define a quantitative measure, such as a distance function, of the disparity between two signals. In such cases we may use instead the concept of neighborhood which is basic to the theory of topological spaces" ([1], p. 202).

2) *Multiplex transmission of two or more signals*: A system has two channels and the sets of signals assigned to their respective channels are $X = \{x(t)\}$ and $Y = \{y(t)\}$. If we are given the sum signal $u(t) = x(t) + y(t)$ at the receiving end, how can we extract $x(t)$ and $y(t)$ from $u(t)$? – We have to find two filters N_1 and N_2 such, that, for any x in X and any y in Y,

$$N_1\,(x + y) = x \quad \text{and} \quad N_2\,(x + y) = y.$$

This analogy between the projection in a function space and the filtration by an ideal filter led Zadeh in the early 1950s to a functional symbolism of filters [3].

When the electrical engineer Lotfi A. Zadeh was in his scientific start-up period, he was especially interested in the relationships of real objects to their theoretical counterparts. This interest is expressed in the work he produced during the 1950s: He made extensive use of modern mathematical methods for his general theory of linear signal transmission systems [2]; here he employed Fourier analysis as well as Hilbert space and operator calculus, which had been developed almost 20 years earlier for use, for example, in the quantum mechanics of John von Neumann. In the tradition of Norbert Wiener, Zadeh applied these mathematical methods to problems in communication engineering in order to represent general principles in the transformation of signals and in the characterization of input-output relationships in non-linear systems.

Nevertheless, his theory also paid tribute to the distinctive features of the technical components of real communication systems: “Many of the mathematical techniques used in the theory are commonly employed in quantum mechanics, though in a different form and for different purposes. The dissimilarities are due largely to the special character of signal transmission systems and the nature of the problems associated with the transmission and reception of signals.” ([2], p. 293f) Here Zadeh considered a signal transmission system to be the system *N* shown in the figure 1 (right side), without any further restrictions, into which an input signal $u(t)$ enters and which emits an output signal $v(t)$. *N* is linear if it satisfies the following additivity properties:

$$u(t) = c_1u_1(t) + c_2u_2(t) \quad \text{and} \quad v(t) = c_1v_1(t) + c_2v_2(t),$$

where c_1 and c_2 are any constants and $v_1(t)$ and $v_2(t)$ are the system responses to $u_1(t)$ and $u_2(t)$.

In his theory, Zadeh represented the signals as vectors in an infinite-dimensional signal space Σ.[1] Vectors of such a space Σ can be projected onto a subspace *M* (and along its complement *M’*), and Zadeh interpreted such a projection in his mathematical theory “in physical terms as the filtering of the class of signals representing Σ with an *ideal filter N* which passes without distortion (or delay) all signals belonging to *M* and rejects all those belonging to *M’* ”. ([3], p. 305.)

When considering similar diagrams characterizing electrical filters, in which the current is applied against the frequency, it is readily apparent that there aren’t any ideal filters in reality at all. One always observes *tolerance ranges*, there is always a surrounding area of cut-off frequencies that form the transition between the passband and the stopband; these are not sharp boundaries; in reality they are always fuzzy!

“From the mathematical point of view, the theory of filtering is essentially a study of certain types of mappings of function and sequence spaces.” Zadeh wrote this in one of his last texts on the filter theory ([4], p. 35.), but the mathematically/ analytically sophisticated results of his publications on this subject do not need to be discussed any further here. The obvious discrepancy between mathematical theory and electrotechnical communication systems eventually led him to completely new ideas on how to expand the theory in order to apply it. By forming new concepts and using the subsequent results, Zadeh tied directly into Wiener’s statistical basis of filter theory but then immediately surpassed it.

If filters that were realized as electrical circuits did not operate according to the mathematical theory, the one must be content with less – *optimum* filters: “A distinction is made between ideal and optimum filters, the former being defined as filters which achieve a perfect separation of signal and noise.” ([4], p. 35.) If ideal filtration is not possible, though, which is often the case when the signal is mixed with noise, then one must accept that the filtration can only be incomplete. In such cases, a filter that delivers the best possible approximation of the desired signal – and a particular meaning of “best approximation” is used here – is called an optimum filter. Zadeh stressed that an optimum filter depends upon a reasonable criterion: “Because of its complex nature, the question of what constitutes a “reasonable” criterion admits of much argument. Since $x(t)$ and $y(t)$ are, in general, random functions, it is natural to formulate such a criterion in statistical terms.” ([4], p. 47.)

---

[1] In particular, Zadeh included the Hilbert space L_2 of all quadratically integratable functions.

However, there were usually not enough statistical data to determine an optimum filter by means of statistical calculations. For this reason, Zadeh now proposed an alternative and thus a practicable approach to evaluating the performance of a filter *F* that operated not by statistical methods, but rather once again used a distance in the function space.

Without going any deeper into mathematical details, it should be stated very clearly here that Zadeh had turned away from statistical methods during the course of his work on optimum filters in the mid-50s, and he recognized that a more promising approach was that of finding an optimum filter relative to a distance to be minimized in the in the function space of the signals. Today we know that this renunciation of statistics was a great deal more complete than this.

2 Menger's Microgeometry

Karl Menger was born in Vienna and he entered the University of Vienna in 1920. 1924, after he had got his Ph. D., he changed to Amsterdam because of his interests in topology, he became assistant of the famous mathematician and expert in topology L. E. J. Brouwer's assistant. In 1927 he returned to Vienna as a professor of geometry at the University and he became a member of the Vienna Circle. In 1937, one year before the "Anschluss", Menger immigrated to the USA where he got a professorship at the University of Notre Dame. In 1948 Menger came to the department of mathematics of the Illinois Institute of Technology and he remained in Chicago the rest of his life.

Menger's publications on the theory of metric and convex spaces in the 1920s and 1930s in his Vienna times and in the 1940s and 1950s on *Statistical Metrics* [5] where he also introduced the term "triangular norm" or "t-norm", and on "ensembles flous" [6] are well-known and I addressed these subjects in other papers (e.g. [7-9]. Menger continued this work on similarities and distances in 1966 to contribute to a symposium of the American Association for the Advancement of Science to commemorate the 50th anniversary of Ernst Mach's death. Here, Menger spoke about *Positivistic Geometry* [10]. Ernst Mach's thinking had made a lasting impression on the Vienna Circle and Menger tied his comments into Mach's statements in the *Principles of Thermodynamics* [11] In the chapter entitled *The Continuum*, Mach had begun by characterizing the continuum as a system of members that possessed one or more properties *A* in varying measures and such that an infinite number of members can fit between each pair of members that differ from each other in a finite way with respect to *A*. These consecutive members then display just one infinitely small difference from each other. Mach determined that "there was no reason to object to this fiction or the random conceptual construction of such a system". Yet the natural scientist does not deal only in pure mathematics. So he has to ask himself whether there is anything in nature that corresponds to such a fiction! At the end of his chapter on the continuum, Menger quotes Ernst Mach: "Everything that *appears* to be a continuum could probably consist of *discrete* elements if these were only small enough and numerous enough with respect to our smallest practically applied measures. Anywhere we think we will find a continuum, the fact is that we are still employing analogous considerations to the smallest perceivable parts of the system in question and can observe a behavior similar to those of larger ones. Insofar as *experience* has raised no objections, we can maintain the fiction of continuum, which is by no means harmful but is merely *comfortable*. In *this sense,* we also refer to the system of *thermal conditions* as a *continuum*." ([11], p. 76)[2]

Menger then transitioned to Henri Poincaré, the other great scientist who had substantially influenced the thinking of the Vienna Circle and who had analyzed more aggressively than anyone else the problems with the concept of the physical continuum. Poincaré had characterized the physical continuum as follows [12]:

$$A = B, \qquad B = C, \qquad A \neq C.$$

In doing so, he was trying to symbolize the fact that when an element from the physical continuum can be differentiated from two others, these two other elements can nevertheless easily differ from each other. In Poincaré's view, the mathematical continuum was constructed with the primary intention of overcoming the above formula, which he called "contradictory or at least inconsistent". Only in the mathematical continuum did the equations $A = B$ and $B = C$ imply equation $A = C$. In the observable physical continuum, on the other hand, where "equal" meant "indistinguishable", the equations $A = B$ and $B = C$ did not imply the equation $A = C$ at all. The raw result of this finding must rather be expressed by the following relation and must be considered a formula for the physical continuum:

$$A = B, \qquad B = C, \qquad A < C.$$

In the 1940s and '50s, Menger had expressed a number of thoughts on the concept of physical equality based on this distinction between the mathematical and physical continuum. With regard to Poincaré's above-mentioned argument, he described the difference as a non-transitive relation, and his other examinations led him to an even more radical conclusion: "A closer examination of the physical continuum suggests that in describing our observations we should sacrifice more than the transitivity of equality. We should give up the assumption that equality is a relation." ([13], p. 178.) A simple experiment showed, for example, that irritating two points *A* and *B* on the skin at the same time sometimes resulted in just *one* sensation, but sometimes in *two* sensations, as well. The fact that two different sensations could occasionally be artificially equated with one another occurred merely because we have faith in the majority of the impressions we experience and that we perform averaging processes.

We can obtain a realistic description of the equality of two elements when we associate a number with *A* and *B*, namely the *probability* that *A* and *B* are indistinguishable. In applications, this number can be represented by relative frequencies of the instances in which *A* and *B* cannot be distinguished from one another. This idea, Menger further argued, canceled out Poincaré's paradox:

For if it is only very likely that *A* and *B* are equal, and very likely that *B* and *C* are equal, why should it not be less likely that *A* and *C* are equal? In fact, why should the equality of A and C not be less likely than the inequality of *A* and *C*? ([13], p. 178)

[2] [343] Menger cited this work in English without providing any further source reference.

For a probability $E(a, b)$ that a and b are equal, Menger then postulated the following properties:

(1)	$E(a, a) = 1$	for all a;	(reflexivity)
(2)	$E(a, b) = E(b, a)$	for all a and b;	(symmetry)
(3)	$E(a, b) \bullet E(b, c) \leq E(a, c)$	for all a, b, c.	("a minimum of transitivity")

Menger proposed calling a and b *certainly-equal* if $E(a, b) = 1$. This led to an equality relation, since all elements that are certainly-equal to a can be combined into an *equality set* A and each pair of these sets is either disjoint or identical. He defined $E(A, B)$ as the probability that every element of A is equal to every element of B. This number is not dependent upon the selection of the two elements in each case. Menger thus used probabilities to define a distance function of this type, as is seen in this following equation: $- \log E(A, b) = d(A, B),$ which fulfills the following properties, i.e. Maurice Fréchet's postulates for the distance in a metric space:

(D1) $d(A, A) = 0$;
(D2) $d(A, B) \geq 0$;
(D3) $d(A, B) \neq 0$ if $A \neq B$;
(D4) $d(A, B) = d(B, A)$ for all a, b;
(D5) $d(A, B) + d(B, C) \neq d(A, C)$.

Menger proposed applying his findings on statistical metrics to positivistic geometry, in particular for the problem of the "physical continuum". In his *Dernières Pensées*, Poincaré had suggested defining the physical continuum as a system S of elements, each pair of said elements being linked together by a chain of elements from S such that every element of the chain is indistinguishable from the following element.

Menger saw the most difficult problem for microgeometry in the individual identification of elements of the space. He published a suggestion for it in a French text in 1951, which he penned during his guest residency at the Sorbonne. In addition to studies of well-defined sets, he called for a theory to be developed in which the relationship between elements and sets is replaced by *the probability that an element belongs to a set*: »Une relation monaire au sens classique est un sous-ensemble F de l'univers. Au sens probabiliste, c'est une function Π_F définie pour tout $x \in U$. Nous appellerons cette fonction même un ensemble flou et nous interprétons $\Pi_{F(x)}$ comme la probabilité que x appartienne à cet ensemble.« ([15], p. 2002) In English, Menger later replaced the term *ensemble flou* with the expression *hazy set* and to elucidate the contrast he referred to conventional sets as *rigid sets*.

Menger clearly recognized that the difficulty of using his term *ensemble flou* in microgeometry, a term which was after all defined by means of probabilities and thus represented a *probabilistic set*, lay in the fact that the individual elements had to be identified. Since this was simply not possible, though, he suggested combining the *ensemble flou* with a geometry of "lumps", for lumps were easier to identify and to differentiate from one another than points. Lumps could assume a position between *indistinguishability* and *apartness*, which would be the condition of overlapping. It was irrelevant whether the primitive (i.e. undefined) concepts of a theory were characterized as *points* and *probably indistinguishable* or as *lumps* and *probably overlapping*. Of course, all of this depended on the conditions that these simple concepts had to fulfill, but the properties stipulated in the two cases could not be identical. "I believe that the ultimate solution of problems of microgeometry may well lie in a probabilistic theory of hazy lumps. The essential feature of this theory would be that lumps would not be point sets; nor would they reflect circumscribed figures such as ellipsoids. They would rather be in mutual probabilistic relations of overlapping and apartness, from which a metric would have to be developed.” ([13])

After Menger had also mentioned the term *ensemble flou* in his speech for the symposium in honor of the 50th anniversary of the death of Ernst Mach, he established a link in a subordinate clause to – as he saw it – a very similar development: “In a slightly different terminology, this idea was recently expressed by Bellman, Kalaba and Zadeh under the name fuzzy set. (These authors speak of the degree rather than the probability of an element belonging to a set).” ([10], p. 232)

It is obvious that Menger was mentioning the commonalities and differences between his *ensembles flous* and Zadeh’s *fuzzy sets* by name. Both scientists had been confronted with the problem of representing distances having points in abstract spaces to each other. Menger sought generalizations of the existing theory of metric spaces that could be applied in the sciences. As a mathematician, he started with concepts of probability theory to represent unknown quantities. In the 1950s, Lotfi Zadeh was considering an application problem: How can signals received by a communication system be identified and separated? To do so, he represented these signals as abstract vectors in a vector space and then likewise tried to find a definition of the unknown distance in that kind of space. Zadeh then also considered various metrics. Menger developed a theory for the probability of equality Zadeh had the fundamental idea in the development of fuzzy theory when he attempted to characterize the relationship between elements and sets without using concepts from probability theory.

3 Medical Philosophy

From the 1980’s, the Iranian-German physician and philosopher of medicine Kazem Sadegh-Zadeh discussed in the nature of health, illness, and he adopted a fuzzy-theory approach to postulate a novel theory of these concepts: “health is a matter of degree, illness is a matter of degree, and disease is a matter of degree” [15].

He introduced the concept of *patienthood* (“being afflicted by a malady” [14]) in the discussion “of which the notion of *health* will be the additive inverse in the following sense: Health = 1 – patienthood.” In 2000 Sadegh-Zadeh presented a whole framework of the novel theory he aims for: “With Ω being the set of human beings at a particular time we have a fuzzy subset P (the set of patients) of Ω “whose members are to various extents characterized by discomfort, pain, endogenously threatened life, loss of autonomy, loss of vitality, and loss of pleasure. The extent to which an individual is a member of this fuzzy set P is called the degree of patienthood.” [15] The membership function $\mu_p(x) \in [0,1]$ indicates the degree of patienthood of the individual $x \in \Omega$. H (the set of healthy people) is the complement of P. He continued:

$$\mu_p(x) = \text{degree of patienthood of } x, \qquad \mu_h(x) = \text{degree of health of } x,$$

by definition $\mu_h(x) = 1 - \mu_p(x)$, $H = \{(x, \mu_h(x)) | \, x \in \Omega\}$ is the fuzzy set *health*.

In Sadegh-Zadeh’s philosophy of medicine, disease entities “may be conceptualized as fuzzy sets” and “symptoms, and signs would then belong to an individual disease to particular extents. Thus, an individual disease would appear as a multidimensional cloud rather than a clear-cut phenomenon". [15] He introduced the term "disease" not with a linguistic but a social definition: There are complex human conditions "that in a human society are termed *diseases*", and he specifies potential candidates "like heart attack, stroke, breast cancer, etc. Such complex human conditions are not, and should not be, merely confined to biological states of the organism. They may be viewed and represented as large fuzzy sets which also contain parts that refer to the subjective, religious, transcendental and social world of the ill, such as pain, distress, feelings of loneliness, beliefs, behavioral disorders, etc. [15] Thus, Sadegh-Zadeh's fuzzy approach to philosophy of medicine is oriented to the actual lives, needs and interests of people in their communities.

Let's take a set of human conditions $\{D_1, D_2, \ldots., D_n\}$ with its corresponding criteria $\{C_1, C_2, \ldots, C_n\}$ into account. Sadegh-Zadeh established the following rules:

Every element $D_i \in \{D_1, D_2, \ldots., D_n\}$ is a disease and every element that is *similar* to a disease with respect to the criteria $\{C_1, C_2, \ldots, C_n\}$ is a disease.

What is meant by *similarity*? – Sadegh-Zadeh introduced the fuzzy set *differ(A,B)*, i. e. the difference of two fuzzy sets A and B, which has to be calculated as follows:

$$differ(A,B) = \frac{\sum_i \max(0, \mu_A(x_i) - \mu_B(x_i)) + \sum_i \max(0, \mu_B(x_i) - \mu_A(x_i))}{c(A \cup B)}$$

c, given in the denominator, is the sum of the membership values of the corresponding fuzzy set (*fuzzy set count*).

Similarity of two fuzzy sets is resulted from the inversion of their fuzzy difference. For example, one raises the question how similar are the two diseases D_i and D_j considering a few criteria $\{C_1, C_2, \ldots. C_m\}$.

Let's assume A to be a fuzzy set of arbitrary dimension and X as a part of this set; so $A \backslash X$. Human conditions, like heart attack and stomach ulcer, can be arranged according to their assimilable criteria $\{C_1, C_2, \ldots. C_m\}$:

- heart_ attack\{$(C_1, a_1), (C_2, a_2), \ldots., (C_m, a_m)$}
- stomach_ulcer\{$(C_1, b_1), (C_2, b_2), \ldots, (C_m, b_m)$}
- heart_ attack\{bodily_lesion, 1), (pain, 0.7), (distress, 0.8)}
- stomach_ulcer\{(bodily_lesion, 1), (pain, 0.3), (distress, 0.5)}

To calculate similarities between fuzzy sets, the following theorem is used:

$$Theorem: \; similar(A,B) = \frac{c(A \cap B)}{c(A \cup B)}$$

Similar comparisons include several degrees of partial (p) similarity, symbolized as p-similar($A \backslash X$, $B \backslash Y$), under the terms of the following definition:

$$p\text{-similar}\ (A \backslash X, B \backslash Y) = r, \ \text{if similar}\ (X, Y) = r.$$

Assuming that $\{D_1, ..., D_n\}$ would be a small set of human conditions, because of a set of criteria $\{C_1, \ldots, C_n\}$ which these conditions have in common. Each of these conditions is interpreted in a certain society as a disease. For this society there is an agreement of degree ε of partial similarity. This degree is a pillar of this society's notion of disease:

- Every element in the basic set $\{D_1, \ldots, D_n\}$ is a disease.
- A human condition H\X is a disease, if there is a disease $D_i\backslash Y \in \{D_1, \ldots, D_n\}$ and there is a $\varepsilon > 0$, so that p-similar $(H\backslash X, D_i\backslash Y) \geq \varepsilon$

According to this definition a proper choice of ε is essential: The smaller the ε that is chosen, the more diseases will exist and vice versa. However, the value of ε is not chosen by the doctor, but by society. Anyway, this notion of disease is a notion that can be comprised in binary logic, because there is made an explicit difference between states that are consistent with a disease and states that are not. Therefore, Sadegh-Zadeh expands this notion of disease to a notion of "Disease to a certain degree": Let's assume H to be a small set of human conditions. A fuzzy set Δ over H is considered as a set of diseases only if there is a subset $\{D_1, \ldots, D_n\}$ of H and there is a function μ_Δ, so that: μ_Δ: H → [0,1] with

$$\mu_\Delta(H_i\backslash X) = \begin{cases} 1, \text{ if } H_i\backslash X \in \{D_1, \ldots, D_n\}, \text{ called prototype disease.} \\ \varepsilon, \text{ if there is a prototype disease } H_j\backslash Y \text{ with } p\text{-similar } (H_i\backslash X, H_j\backslash Y) = \varepsilon \\ \text{and no prototype disease } H_k\backslash Z \text{ with } p\text{-similar } (H_i\backslash X, H_k\backslash Z) > \varepsilon \\ \text{and } \Delta = \{(H_i, \mu_\Delta(H_i)) \mid H_i \in H\}. \end{cases}$$

In this expanded definition a fuzzy set of following kind is created:

$$\Delta = \{(D_1, \mu_\Delta(D_1)), \ldots, (D_q, \mu_\Delta(D_q))\},$$

which consists of individual archetypes of diseases, which are all members of the set Δ to different degrees. The membership-degree $\mu_\Delta(D_i)$ is of interval [0,1].

From this, we conclude that a person may have a disease to a certain degree and that this person may have no disease to a certain degree at the same time.

4 Conclusion: Similarity and Distance in the Genetic Material

Diseases have their origin in the genetic material of the living organisms. In the year 2000, Sadegh-Zadeh continued his medical-philosophical examinations into theoretical research on genomes [16]. Essentially, the human's germplasm is determined by the DNA, which is a double-stranded molecule consisting of different organic bases adenine (A), guanine (G), cytosine (C), and thymine (T) and which is coded by RNA that has the alphabet " <U, C, A, G> with uracil (U) instead of T.

Can living beings have a base in a unique gene to a certain degree? And can we find similarities and distances of polynucleotides? Can we compare very long base sequences with the use of fuzzy set theory? These questions yield to further work by Sadegh-Zadeh and also Torres and Nieto [17] that we describe in another paper [18].

Acknowledgments. We would like to thank Erich-Peter Klement (Linz) and Kazem Sadegh-Zadeh (Münster) for special advices. This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

1. Zadeh, Lotfi A.: Some Basic Problems in Communication of Information. *The New York Academy of Sciences*, March 1952, Ser. II, Vol.14, No. 5, pp. 201-204.
2. Zadeh, Lotfi A.: A General Theory of Linear Signal Transmission Systems. *Journal of The Franklin Institute*, April 1952, pp. 293-312.
3. Zadeh, Lotfi A. / Miller, Kenneth S.: Generalized Ideal Filters. *Journal of Applied Physics*, February 1952, Vol. 23, No. 2, pp. 223-228.
4. Zadeh, Lotfi A.: Theory of Filtering. *Journal of the Society for Industrial and Applied Mathematics* 1.1 (1953), pp. 35-51.
5. Menger, Karl: Statistical Metrics. *Proceedings of the National Academy of Sciences*, Vol. 28 (1942), pp. 535-537.
6. Menger, Karl: Ensembles flous et fonctions aléatoires. *Comptes Rendus Académie des Sciences* 37 (1951), S. 226-229.
7. Seising, Rudolf: Pioneers of Vagueness, Haziness, and Fuzziness in the 20th Century. In: Nikraves, Masoud, Janusz Kacprzyk and Lotfi A. Zadeh (eds.): *Forging the Frontiers I. Studies in Fuzziness and Soft Computing*. New York [u.a.]: Springer (Series in Fuzziness and Soft Computing) (in print 2007), p. 61-88.
8. Seising, Rudolf: On the Absence of Strict Boundaries − Vagueness, Haziness, and Fuzziness in Philosophy, Medicine, and Science, *Applied Soft Computing*, Special Issue: Forging the Frontiers (in print 2007).
9. Seising, Rudolf: *The Fuzzification of Systems. The Genesis of Fuzzy Set Theory and Its Initial Applications − Its Development to the 1970s*. Berlin [u.a.]: Springer (in print 2007.
10. Menger, Karl: Geometry and Positivism. A Probabilistic Microgeometry, in: Karl Menger: Selected papers in logic and foundations, didactics, economics. Vienna Circle Collection Band 10, Dordrecht, Holland: D. Reidel Publ. Comp., 1979. pp. 225-234.
11. Mach, Ernst: *Die Principien der Wärmelehre*. (4th ed.). Leipzig: Verlag von Johann Ambrosius Barth 1923.
12. Poincaré, Henri: *La science et l'hypothèse.* La valeur de la sciennce; Des fondements de la géométrie. Paris: Flammarion 1902.
13. Menger, Karl: Probabilistic Theories of Relations, *Proceedings of the National Academy of Sciences*, Vol. 37 (1951) pp. 178-180.
14. Sadegh-Zadeh, Kazem: Organism and disease as fuzzy categories, presented at the conference on Medicine and Philosophy, Humboldt University of Berlin, July 2, 1982
15. Sadegh-Zadeh, Kazem: Fuzzy Health, Illness, and Disease. *Journal of Medicine and Philosophy*, Vol. 25, No 5, 2000, pp. 605-638: 612.
16. Sadegh-Zadeh, Kazem: Fuzzy Genomes. *Artificial Intelligence in* Medicine, 18 (2000) pp. 1-28.
17. Torres, Angela and Juan J. Nieto: The fuzzy polynucleotide space: basic properties. *Bioinformatics*, 19 (5), 2003, pp. 582-592.
18. Limberg, Julia: Fuzzy Polynucleotides – Two Approaches and Implementations to a Theory of Fuzzy Genomes. NAFIPS'2007 San Diego, CA June 24-27, 2007 (in print).

The Choquet and Sugeno Integrals as Measures of Total Effectiveness of Medicines

Elisabeth Rakus-Andersson and Claes Jogreus

Blekinge Institute of Technology, School of Engineering,
Department of Mathematics and Science, S-37179 Karlskrona, Sweden
{Elisabeth.Andersson, Claes.Jogreus}@bth.se

Abstract. The concepts of the Choquet and Sugeno integrals, based on a fuzzy measure, can be adopted as useful tools in estimation of the total effectiveness of a drug when appreciating its positive influence on a collection of symptoms typical of a considered diagnosis. The expected effectiveness of the medicine is evaluated by a physician as a verbal expression for each distinct symptom. By converting the words at first to fuzzy sets and then numbers we can regard the effectiveness structures as measures in the Choquet and Sugeno problem formulations. After comparing the quantities of total effectiveness among medicines, expressed as the values of the Choquet or Sugeno integrals, we accomplish the selection of the most efficacious drug.

Keywords: Fuzzy utilities, weights of importance, fuzzy decision-making model, Choquet integral, Sugeno integral.

1 Introduction

Theoretical fuzzy decision-making models, mostly described in [8, 9], give rise to successfully accomplished technical applications. However, there are not so many medical applications to decision-making proposals, especially they are lacking in the domain of pharmacy matters. In own research works [5, 6] we have already made attempts of appreciation of drug efficiency. We thus begin the paper with recalling and modernizing the mentioned model of the drug selection in Section 2. Since we need to develop the conception of effectiveness as a fuzzy measure then we will thoroughly discuss the connection between a symptom and a medicine in Section 3. In Section 4 we introduce the item of the Choquet integral seen from the point of view concentrating attention on the utility measure of drugs. After obtaining satisfactory results we continue discussing the effective role of integrals by extending – in Section 5 – the measurable effect of effectiveness on the subject of the Sugeno integral. The authors should mention that the approach to the use of integrals as computation tools of utility constitutes their original and genuine insertion in fuzzy decision-making.

2 The General Outline of a Drug Decision-Making Model

We introduce the notions of a space of states $X = \{x_1,...,x_m\}$ and a decision space (a space of alternatives) $A = \{a_1,...,a_n\}$. We consider a decision model in which n

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 253–262, 2007.
springerlink.com

alternatives $a_1,...,a_n \in A$ act as drugs used to treat patients who suffer from a disease. The medicines should influence m states $x_1,...,x_m \in X$, which are identified with m symptoms typical of the morbid unit considered.

If a rational decision maker makes a decision $a_i \in A$, $i = 1, 2, \ldots, n$, concerning states-results $x_j \in X$, $j = 1, 2, ..., m$, then the problem is reduced to the consideration of the ordered triplet (X, A, U), where X is a set of states-results, A – a set of decisions and U – the utility matrix [5, 6, 8, 9]

$$U = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & \cdots & x_m \end{array} \\ \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \vdots & & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nm} \end{bmatrix} \end{array} \tag{1}$$

in which each element u_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, is a representative value belonging to [0, 1] for the fuzzy utility following from the decision a_i with the result x_j.

The theoretical model with the triplet (X, A, U) can find its practical application in the processes of choosing an optimal drug from a sample of tested medicines [5, 6].

Let us further associate with each state-symptom x_j, $j = 1, \ldots, m$, a non negative number that indicates its power or importance in decision making in accordance with the rule: the higher the number is, the greater significance of symptom x_j will be expected, when regarding its harmful impact on the patient's condition. If we assign w_1, $w_2, \ldots, w_m$ as powers-weights to $x_1, x_2, \ldots, x_m$, $w_j \in W$, $j = 1, 2, \ldots, m$, where W is a space of weights, then we will modify (1) as the weighted matrix

$$U_W = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & \cdots & x_m \end{array} \\ \begin{bmatrix} w_1 \cdot u_{11} & w_2 \cdot u_{12} & \cdots & w_m \cdot u_{1m} \\ w_1 \cdot u_{21} & w_2 \cdot u_{22} & \cdots & w_m \cdot u_{2m} \\ \vdots & & \ddots & \vdots \\ w_1 \cdot u_{n1} & w_2 \cdot u_{n2} & \cdots & w_m \cdot u_{nm} \end{bmatrix} \end{array}. \tag{2}$$

In compliance with data entries determined in (2), the common decisive power of a_i is approximated by the quantity $U_W(a_i)$ defined as an OWA operation [8, 9]

$$U_W(a_i) = \sum_{j=1}^{m} w_j \cdot u_{ij} \,. \tag{3}$$

As a final optimal decision a^* we select this a_i that satisfies

$$U_W(a^*) = \max_{1 \le i \le n} U_W(a_i), \tag{4}$$

i.e., we pick out the decision-drug possessing the highest utility grade with respect to symptoms cured. The distinct utility u_{ij} is comprehended to be the ability of the symptom retreat after medication. In other words, we define utility u_{ij} of a_i taken to x_j as effectiveness of drug a_i observed in the case of x_j.

3 The Decisive Role of Effectiveness in the Final Decision

Let us find a way of determining effectiveness of drugs as mathematical expressions that should take place in the matrix U_W. On the basis of earlier experiments, the physician defines in words the curative drug efficiency with respect to considered symptoms. He suggests a list of terms that introduce a linguistic variable named "drug effectiveness concerning symptom" = {R_1 = *none,* R_2 = *almost none,* R_3 = *very little,* R_4 = *little,* R_5 = *rather little*, R_6 = *medium,* R_7 = *rather large,* R_8 = *large*, R_9 = *very large,* R_{10} = *almost complete,* R_{11} = *complete*} [5, 6]. Each notion from the list of terms is the name of a fuzzy set. Assume that all sets are defined in the space $Z = [0,100]$, suitable as a reference set for supports of R_1-R_{11}.

We propose constrains for the fuzzy sets R_1-R_{11} by applying linear functions [5, 6]

$$L(z,\alpha,\beta) = \begin{cases} 0 & \text{for} \quad z \le \alpha \\ \dfrac{z-\alpha}{\beta-\alpha} & \text{for} \quad \alpha < z \le \beta \\ 1 & \text{for} \quad z > \beta \end{cases} \tag{5}$$

and

$$\pi(z,\alpha,\gamma,\beta) = \begin{cases} 0 & \text{for} \quad z \le \alpha \\ L(z,\alpha,\gamma) & \text{for} \quad \alpha < z \le \gamma \\ 1 - L(z,\gamma,\beta) & \text{for} \quad \gamma < z \le \beta \\ 0 & \text{for} \quad z > \beta \end{cases} \tag{6}$$

where z is an independent variable from [0, 100], whereas α, β, γ are parameters.

Let us now define

$$\mu_{R_t}(z) = \begin{cases} 1 - L(z,\alpha_t,\beta_t) & \text{for} \quad t = 1,2,3,4,5 \\ L(z,\alpha_t,\beta_t) & \text{for} \quad t = 7,8,9,10,11 \end{cases} \tag{7}$$

and

$$\mu_{R_6}(z) = \pi(z,\alpha_6,\gamma,\beta_6) \tag{8}$$

in which α_t, β_t, γ are the borders for supports of the fuzzy sets R_1-R_{11}.

We decide the values of the boundary parameters α_t, β_t, γ in Ex. 1 below.

Example 1

Figure 1 collects the graphs of fuzzy sets R_1-R_{11} that can be approved as the terms composing the contents of the effectiveness list.

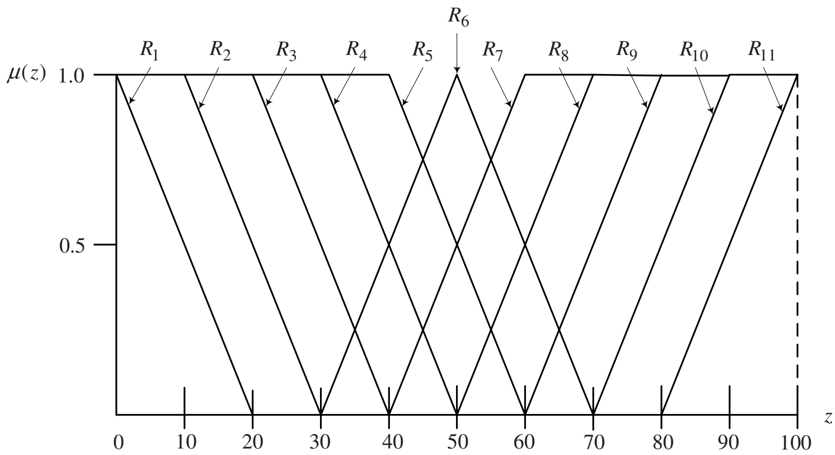


Fig. 1. The fuzzy sets R_1-R_{11}

To each effectiveness, expanded as a continuous fuzzy set, we would like to assign only one value.

Example 2

To find the adequate $z \in [0, 100]$ representing the effectiveness terms R_1-R_{11} we adopt as z the values α_t for $t = 1, 2, 3, 4, 5$, and β_t for $t = 7, 8, 9, 10, 11$ in compliance with (7), respectively γ due to (8). We simply read off the values of α_t, β_t and γ from Fig. 1 in order not to introduce evident calculations. These z-values are elements of the support of a new fuzzy set "*effectiveness*" whose membership function is expressed over the interval [0, 100] by $\mu_{\text{"effectiveness"}}(z) = L(z, 0, 100)$. For the z-representatives of R_1-R_{11}, we finally compute membership values $\mu_{\text{"effectiveness"}}(z)$, which replace the terms of effectiveness-utility as quantities u_{ij}. We summarize the obtained results in Table 1.

Table 1. The representatives of effectiveness

Effectiveness	Representing z-value for effectiveness	$\mu(z) = u_{ij}$
none	0	0
almost none	10	0.1
very little	20	0.2
little	30	0.3
rather little	40	0.4
medium	50	0.5
rather large	60	0.6
large	70	0.7
very large	80	0.8
almost complete	90	0.9
Complete	100	1

The next problem to put into discussion is a procedure of obtaining an importance ratio scale for a group of m symptoms [7]. Assume that for m states-symptoms we wish to construct a scale, rating them as to their importance for making the decision. We compare the symptoms in pairs with respect to their harmful impact on the patient's health. If we confront symptom j with symptom l, then we can assign the values b_{jl} and b_{lj} to the pair (x_j, x_l) as follows, $j, l = 1, 2, \ldots, m$:

(1) $b_{lj} = \dfrac{1}{b_{jl}}$,

(2) If symptom j is more important than symptom l then b_{jl} gets assigned one of the numbers 1, 3, 5, 7 or 9 due to the difference of importance being *equal*, *week*, *strong*, *demonstrated* or *absolute*, respectively. If symptom l is more important than symptom j, we will assign the value of b_{lj}.

Having obtained the above judgments an $m \times m$ matrix $B = (b_{jl})_{j,l=1}^{m}$ is constructed. The weights $w_1, w_2, \ldots, w_m \in W$ are decided as components of the eigen vector corresponding to the largest in magnitude eigen value of the matrix B. We normalize the weights w_j by dividing them all by the largest weight $w_{largest}$. We suggest this simple operation to keep all w_j within interval [0, 1] that now replaces W.

Let us denote the normalized weights by $\hat{w}_j = \dfrac{w_j}{w_{largest}}$. Afterwards we reorder $\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_m$ to generate an arrangement of the normalized weights as the ascending sequence $\hat{w}_1^a, \hat{w}_2^a, \ldots, \hat{w}_m^a$ satisfying the condition $0 \le \hat{w}_1^a \le \hat{w}_2^a \le \ldots \le \hat{w}_m^a = 1$. The symptoms x_j follow the new replacement of associated weights. In order to avoid too many designation signs let us name the ordered and normalized weights $\omega_j = \hat{w}_j^a$ and attached to them symptoms $\chi_j \in X$, $j = 1, 2, \ldots, m$. The matrix U_W is accommodated to new assumptions as $U_{[0,1]}$ given by

$$U_{[0,1]} = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{array} \begin{array}{c} \begin{array}{cccc} \chi_1 & \chi_2 & \cdots & \chi_m \end{array} \\ \begin{bmatrix} \omega_1 \cdot u_{11} & \omega_2 \cdot u_{12} & \cdots & \omega_m \cdot u_{1m} \\ \omega_1 \cdot u_{21} & \omega_2 \cdot u_{22} & \cdots & \omega_m \cdot u_{2m} \\ \vdots & & \ddots & \vdots \\ \omega_1 \cdot u_{n1} & \omega_2 \cdot u_{n2} & \cdots & \omega_m \cdot u_{nm} \end{bmatrix} \end{array}. \tag{9}$$

The formula (3) has been replaced by

$$U_{[0,1]}(a_i) = \sum_{j=1}^{m} \omega_j \cdot u_{ij} \tag{10}$$

in regard to the new order of weights.

After the theoretical accomplishments of proofs yielding methods of stating crucial decision elements u_{ij} (effectiveness of assimilating a_i with x_j or, rather, with χ_j) and ω_j, let us show the results of the procedure that selects an optimal medicine.

Example 3
The following clinical data concerns the diagnosis "*coronary heart disease*". We consider the most substantial symptoms x_1 = "*pain in chest*", x_2 = "*changes in ECG*" and x_3 = "*increased level of LDL-cholesterol*". The medicines improving the patient's state are recommended as a_1 = *nitroglycerin*, a_2 = *beta-adrenergic blockade* and a_3 = *statine LDL-reductor*.

The physician has judged the relationship among efficiency of the drugs and retreat of the symptoms. We express the connections in Table 2.

Table 2. The relationship between medicine action and retreat of symptom

$a_i \backslash x_j$	x_1	x_2	x_3
a_1	complete, $u_{11} = 1$	very large, $u_{12} = 0.8$	almost none, $u_{13} = 0.1$
a_2	medium, $u_{21} = 0.5$	rather large, $u_{22} = 0.6$	little, $u_{23} = 0.3$
a_3	little, $u_{31} = 0.3$	little, $u_{32} = 0.3$	very large, $u_{33} = 0.8$

Further, we conclude that the physical status of a patient is subjectively better if the symptom x_1 = "*pain in chest*" disappears. The next priority is assigned to x_2 = "*changes in ECG*" and lastly, we concentrate our attention on getting rid of x_3 = "*increased level of LDL cholesterol*". We thus construct B as

$$B = \begin{matrix} & x_1 & x_2 & x_3 \\ x_1 & 1 & 3 & 5 \\ x_2 & \frac{1}{3} & 1 & 3 \\ x_3 & \frac{1}{5} & \frac{1}{3} & 1 \end{matrix} .$$

The largest eigen value of B has the associated eigen vector $V = (0.93295, 0.30787, 0.18659)$. V is composed of coordinates that are interpreted as the weights w_1, w_2, w_3 sought for x_1, x_2, x_3. We normalize and rearrange the weights to obtain $\omega_1 = 0.2$ attached to $\chi_1 = x_3$, $\omega_2 = 0.33$ connected to $\chi_2 = x_2$ and $\omega_3 = 1$ as the power of $\chi_3 = x_1$.

Due to (10) we approximate the utilities $U_{[0,1]}(a_i)$ of medicines a_i, $i = 1, 2, 3$, as

$U_{[0,1]}(a_1) = 0.2\ 6\ 0.1 + 0.33\ 6\ 0.8 + 1\ 6\ 1 = 1.284,$
$U_{[0,1]}(a_2) = 0.2\ 6\ 0.3 + 0.33\ 6\ 0.6 + 1\ 6\ 0.5 = 0.758,$
$U_{[0,1]}(a_3) = 0.2\ 6\ 0.8 + 0.33\ 6\ 0.3 + 1\ 6\ 0.3 = 0.559.$

After placing the utilities of drugs in the decreasing order (see (4)) we establish the hierarchy of medicines as $a_1 \succ a_2 \succ a_3$, when supposing that the notion $a_i \succ a_k$ stands for "a_i acts better than a_k" with respect to all involved symptoms, $i, k = 1, 2, 3$.

4 The Choquet Integral as Total Effectiveness

The normalization and the rearrangement of weights have been made in the intention of proving that formula (10) can be interpreted as a rule corresponding to the Choquet integral calculation [1, 2, 3, 4].

We know that the symptoms $\chi_1,...,\chi_m \in X$ act as objects in *X*. To them let us assign the measures $m(\{\chi_j|a_i\})$ = u_{ij}, where the symbols $\chi_j|a_i$ reflect the association between symptom χ_j and medicine a_i, j = 1, 2, …, m, i = 1, 2, …, n. The values $m(\{\chi_j|a_i\})$ are listed in the last column of Table 1.

The weights ω_j are set as the range values $f(\chi_j)$ of a function $f: X \to W = [0, 1]$.

By considering the latest suggestions we define the total utility of a_i gathered for all symptoms $\chi_1, \chi_2, \ldots, \chi_m$ as the Choquet integral

$$U^{Ch}_{[0,1]}(a_i) = \int_{X=\{\chi_1,\chi_2,\ldots,\chi_m\}} f(\chi_j)dm(\chi_j|a_i) \tag{11}$$

with respect to the measures $m(\{\chi_j|a_i\})$.

To find a precise calculus formula of integral (11) we study Fig. 2, made for three symptoms χ_1, χ_2, χ_3. This associates to (11) the following equation

$$\begin{aligned} U^{Ch}_{[0,1]}(a_i) = \int_{X=\{\chi_1,\chi_2,\chi_3\}} f(\chi_j)dm(\chi_j|a_i) = (\omega_1 - \omega_0)\cdot m\{\chi_j|a_i : f(\chi_j) \geq \omega_1\} \\ + (\omega_2 - \omega_1)\cdot m\{\chi_j|a_i : f(\chi_j) \geq \omega_2\} + (\omega_3 - \omega_2)\cdot m\{\chi_j|a_i : f(\chi_j) \geq \omega_3\}, \end{aligned} \tag{12}$$

that practically explains how to understand the Choquet integral arithmetic. The measures of sets consisting of elements $\chi_j|a_i$, defined by properties $f(\chi_j) \geq \omega_1$, $f(\chi_j) \geq \omega_2$ and $f(\chi_j) \geq \omega_3$, are estimated as sums of utilities corresponding to respective $\chi_j|a_i$ fulfilling conditions above.

The general formula of the Choquet integral is revealed in the form

$$U^{Ch}_{[0,1]}(a_i) = \int_{X=\{\chi_1,\chi_2,\ldots,\chi_m\}} f(\chi_j)dm(\chi_j|a_i) = \sum_{j=1}^{m}(\omega_j - \omega_{j-1})\cdot m\{\chi_l|a_i : f(\chi_l) \geq \omega_j\} \tag{13}$$

for $\omega_0 = 0$, l = 1, 2, …, m, i = 1, 2, …, n.

Let us recall that *m* is a utility measure with values in {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1} (the set of u_{ij}-values standing for effectiveness) defined for symptoms after treating them by medicines. If the utility is *none* then its measure will be equal to zero. For the total utility *complete* we reserve the measuring quantity of one. The physician can decide the common utility of medicine for two symptoms being less then the sum of utilities for distinct symptoms, e.g., the effectiveness of a_2 for "*pain in chest* and *changes in ECG*" together is judged as 0.5 while the separate measures of effectiveness emerge 0.6 and 0.5 (see Table 2). The last remark reveals the non-additive property of the effectiveness measure. Without any formal proofs made for confirmation of effectiveness as a fuzzy measure, we intend to use it in Choquet integrals constructed for the sample of medicines to approximate their curative effects.

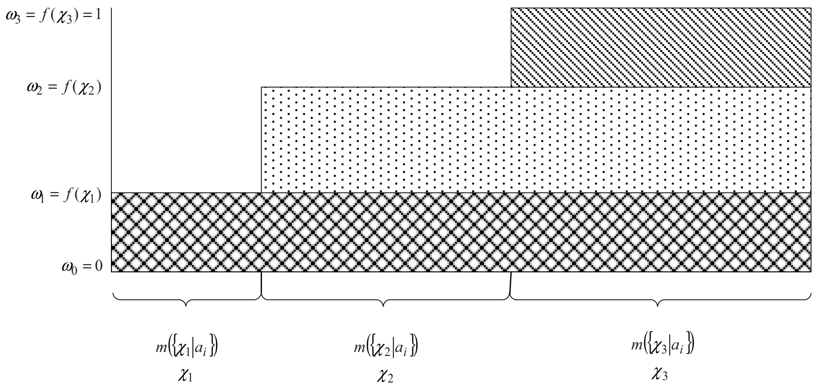


Fig. 2. The Choquet integral in evaluation of a_i's total curative effect

In the next example we compute the entire effectiveness of medicines from Ex. 3 to compare the results obtained there.

Example 4
Let us involve formula (12) together with Fig. 2 to estimate

$$U_{[0,1]}^{Ch}(a_1) = \int_{X=\{\chi_1,\chi_2,\chi_3\}} f(\chi_j)dm(\chi_j|a_1) = (0.2-0)\cdot m\{\chi_j|a_1 : f(\chi_j)\geq 0.2\}$$
$$+(0.33-0.2)\cdot m\{\chi_j|a_1 : f(\chi_j)\geq 0.33\} + (1-0.33)\cdot m\{\chi_j|a_1 : f(\chi_j)\geq 1\}$$
$$= 0.2\cdot m\{\chi_1|a_1, \chi_2|a_1, \chi_3|a_1\} + 0.13\cdot m\{\chi_2|a_1, \chi_3|a_1\} + 0.67\cdot m\{\chi_3|a_1\}$$
$$= 0.2\cdot(0.1+0.8+1) + 0.13\cdot(0.8+1) + 0.67\cdot 1 = 1.284,$$

$$U_{[0,1]}^{Ch}(a_2) = 0.2\cdot(0.3+0.6+0.5) + 0.13\cdot(0.6+0.5) + 0.67\cdot 0.5 = 0.758$$

and

$$U_{[0,1]}^{Ch}(a_3) = 0.2\cdot(0.8+0.3+0.3) + 0.13\cdot(0.3+0.3) + 0.67\cdot 0.3 = 0.559.$$

The results are identical with calculations obtained in Ex. 3, which confirms the proper interpretation of the Choquet integral in the drug ranking $a_1 \succ a_2 \succ a_3$.

5 The Sugeno Integral in Hierarchical Drug Order

To be able to introduce the Sugeno-like integral in the calculations leading to the choice of an optimal medicine, we normalize the measures $m\{\chi_l|a_i : f(\chi_l)\geq \omega_j\}$ from (13), j, l = 1, 2, …, m, when dividing them all by the largest value in the sequence. This operation provides us with the quantities $\hat{m}\{\chi_l|a_i : f(\chi_l)\geq \omega_j\}$ belonging to [0, 1].

As the next estimate of a_i's entire utility we propose a formula [1, 2, 3, 4]

$$U^S_{[0,1]}(a_i) = \int_{X=\{\chi_1,\chi_2,\dots,\chi_m\}} f(\chi_j)dm(\chi_j|a_i) = \max_{1\le j\le m}\left(\min\left(\omega_j, \hat{m}\{\chi_l|a_i : f(\chi_l)\ge\omega_j\}\right)\right) \quad (14)$$

for $j, l = 1, 2, \dots, m$, $i = 1, 2, \dots, n$.

Example 5

The measures $m\{\chi_l|a_1 : f(\chi_l)\ge 0.2\}=1.9$, $m\{\chi_l|a_1 : f(\chi_l)\ge 0.33\}=1.8$ and $m\{\chi_l|a_1 : f(\chi_l)\ge 1\}=1$ found for a_1 in Ex. 4 are now divided by the largest value of m equal to 1.9 to generate their normalized versions $\hat{m}\{\chi_l|a_1 : f(\chi_l)\ge 0.2\}=1$, $\hat{m}\{\chi_l|a_1 : f(\chi_l)\ge 0.33\}=0.947$ and $\hat{m}\{\chi_l|a_1 : f(\chi_l)\ge 1\}=0.526$. In the scenario of (14) we estimate the utility of a_1 as

$$U^S_{[0,1]}(a_1) = \int_{X=\{\chi_1,\chi_2,\chi_3\}} f(\chi_j)dm(\chi_j|a_1) = \max(\min(0.2,1), \min(0.33,0.947), \min(1,0.526))$$
$$= 0.526.$$

For a_2 we get the utility value

$$U^S_{[0,1]}(a_2) = \max(\min(0.2,1), \min(0.33,0.786), \min(1,0.357)) = 0.357,$$

while a_3 possesses the affection grade on symptoms from X approximated as

$$U^S_{[0,1]}(a_3) = \max(\min(0.2,1), \min(0.33,0.428), \min(1,0.214)) = 0.33.$$

Even the application of the Sugeno integral provides us with the same hierarchy ladder of medicines upgraded in the order $a_1 \succ a_2 \succ a_3$. We should mention that the utility values in the last computations are comparable to the "ideal" utility equal to one that can be reached in the state of absolute absence of all symptoms.

6 Conclusions

As a primary method of fuzzy decision-making we have adopted Yager's model in the process of extraction of the best medicine from the collection of proposed remedies. The basis of investigations has been mostly restricted to a judgment of medicine influence on clinical symptoms accompanying the disease. We have also employed the indices of the symptoms' importance to emphasize the essence of additional factors in the final decision. By interpreting the utilities of drugs as measures the authors have furnished such tools as the Choquet and Sugeno integrals to rearrange the conception of the classical fuzzy decision-making, which constitutes an original contribution in decision model. The adaptation of the Choquet integral has provided us with the same results like these ones obtained by applying of the classical method. Even the adjustment of the Sugeno integral to the medication problem has brought the effects totally confirming the medicine order previously determined.

References

1. Grabisch, M., Murofushi, T., Sugeno, M., Kacprzyk, J.: Fuzzy Measures and Integrals. Theory and Applications. Physica Verlag, Berlin Heidelberg (2000)
2. Mirofushi, T., Sugeno, M.: An Interpretation of Fuzzy Measures and the Choquet Integral as an Integral with Respect to a Fuzzy Measure. Fuzzy Sets and Systems 29 (1989) 201-227
3. Mirofushi, T., Sugeno, M.: Fuzzy Measures and Fuzzy Integrals. In: Grabisch, M., Murofushi, T., Sugeno, M., Kacprzyk, J. (eds.): Fuzzy Measures and Integrals. Theory and Applications. Physica Verlag, Berlin Heidelberg (2000) 3-34
4. Sugeno, M.: Fuzzy Measures and Fuzzy Integrals - a Survey. In: Gupta, M. M., Saridis, G. N., Gaines, B. R. (eds.): Fuzzy Automata and Decision Processes, New York North-Holland (1977) 89-102
5. Rakus-Andersson, E.: A Choice of Optimal Medicines by Fuzzy Decision Making Including Unequal Objectives. Issues in the Representation and Processing of Uncertain and Imprecise Information, EXIT – The publishing House of The Polish Academy of Sciences (2005) 307-321
6. Rakus-Andersson, E.: Minimization of Regret versus Unequal Multi-objective Fuzzy Decision Process in a Choice of Optimal Medicines. Proceedings of the XIth International Conference IPMU 2006 – Information Processing and Management of Uncertainty in Knowledge-based Systems, vol. 2, Edition EDK, Paris-France (2006) 1181-1189
7. Saaty, T. L.: Exploring the Interface between Hierarchies, Multiplied Objectives and Fuzzy Sets. Fuzzy Sets and Systems 1 (1978) 57-68
8. Yager, R. R.: Fuzzy Decision-making Including Unequal Objectives. Fuzzy Sets and Systems 1 (1978) 87-95
9. Yager, R. R.: Decision-making Using Minimization of Regret. International Journal of Approximate Reasoning 36 (2004) 109-128

Managing Uncertainty with Fuzzy-Automata and Control in an Intensive Care Environment

Christian Schuh

Core Unit for Medical Statistics and Informatics,
Medical Expert-and Knowledge-Based Systems, Medical University of Vienna,
Spitalgasse 23, A-1090 Vienna, Austria
christian.schuh@meduniwien.ac.at

Abstract. Medical informatics has changed tremendously over the past few decades, and changes in the approach to uncertainty are probably the most important advances in this field. The envisioned role of computer programs in health care is perhaps the most important. Uncertainty is the central, critical fact about medical reasoning. Particularly in an intensive care environment, where decisions must often be made quickly, or that physicians will follow it rather than openly or surreptitiously limiting care on their own.

This paper surveys the utilization of fuzzy logic on the basis of two medical applications. The first, an intelligent on-line monitoring program for the intensive care data of patients with Acute Respiratory Distress Syndrome (ARDS), so called FuzzyARDS which is using the concept of fuzzy automata, and the second is a fuzzy knowledge-based control system, FuzzyKBWean, which was established as a real-time application based on the use of a Patient Data Management System (PDMS) in an intensive care unit (ICU). These complex systems confirm that fuzzy logic is quite suitable for medical application in a per definition uncertainty environment as an ICU, because of its tolerance to some imprecision.

Keywords: Uncertainty, fuzzy logic, fuzzy automation, medical fuzzy applications, decision making.

1 Introduction

Medicine is one field in which the applicability of fuzzy set theory was recognized in the end-1970s. Within this field it is the uncertainty found in the process of diagnosis of disease that has most frequently been the focus of applications of fuzzy set theory. In other words real world knowledge is characterized by uncertainty, incompleteness and inconsistency. Fuzzy set theory, which was developed by Zadeh [1], makes it possible to define inexact medical entities as fuzzy sets. It provides an excellent approach for approximating medical text. Furthermore, fuzzy logic provides reasoning methods for approximate inference.

2 Intensive Care Environment Applications

Two medical applications, FuzzyARDS and FuzzyKBWean are presented in this paper, representing these concepts. FuzzyARDS is an intelligent on-line monitoring

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 263–271, 2007.
springerlink.com

program of data from patients with acute respiratory distress syndrome (ARDS) at an intensive care unit (ICU). It employs fuzzy trend detection and fuzzy automata. FuzzyKBWean is an open-loop fuzzy control program for optimization and quality control of the ventilation and weaning process of patients after cardiac surgery at the ICU. The above-mentioned computer systems have reached the state of extensive clinical integration and testing at the Medical University of Vienna's General Hospital. The obtained results show the applicability and usefulness of these systems.

2.1 FuzzyARDS

FuzzyARDS is an intelligent on-line monitoring program for the intensive care data of patients with Acute Respiratory Distress Syndrome (ARDS) [4].

ARDS is a vaguely defined nosological entity therefore a crisp definition of ARDS seems therefore not adequate. In particular, the commitment to crisp limits of findings is problematic. Fig. 1 shows the thresholds of PaO2 and FiO2 for patients suffering from ARDS.

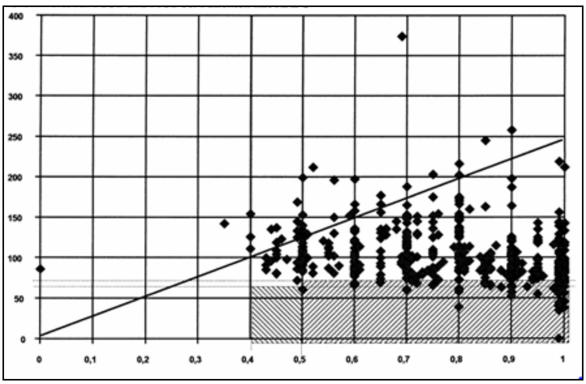


Fig. 1. ARDS threshold of PaO2 and FiO2

Its clinical aim is to detect ARDS in patients as early as possible and to give appropriate therapy advice. ARDS is an ill-defined medical entity and is modeled using the concept of fuzzy automata. States in these automata are considered to be a patient's pathophysiological state or entry criteria for different forms of ARDS therapies (Fig. 2). Patients may be partially assigned to one or several states in such an automaton at the same point in time. Transitions in the automata carry fuzzy conditions that have to be true or partially true to transit from one state to another. Fuzzy conditions are usually high level medical concepts such as low, normal, or high F_iO_2, hypoxemia, or linguistically expressed trend information, e.g., rapidly improving oxygenation. These high-level concepts are permanently evaluated in a data-to-symbol conversion step according to an adjustable time granularity. An extended description of these formal concepts can be found in [5, 8, 29].

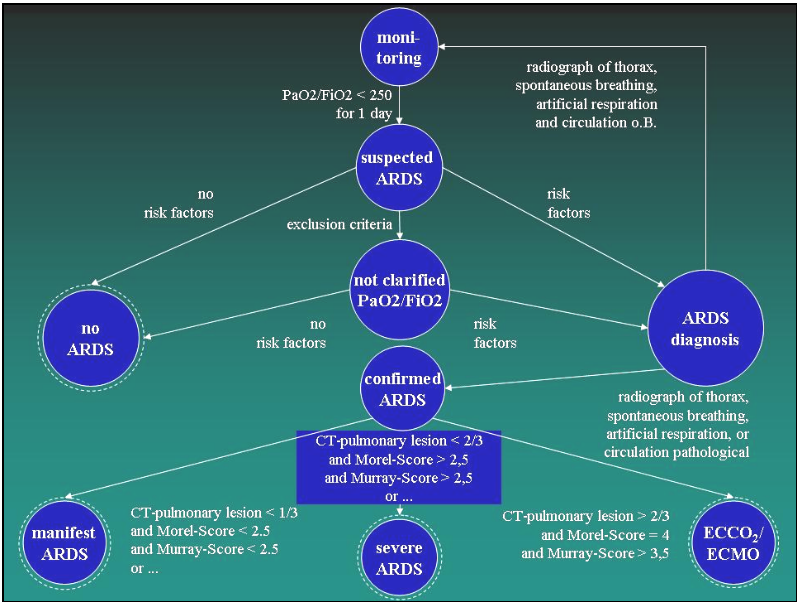


Fig. 2. Fuzzy automata concept for ARDS therapies

In the present phase of development, an international study has been conducted to compare and finally improve the various forms of ARDS definitions used at the medical study centers and compare their respective entry criteria used for therapy decision. A web-based system called FuzzyARDS was programmed allowing patient data entry at the study centers, the definition of fuzzy criteria, the calculation of fuzzy scores with respect to fuzzy criteria in the various stages of illness, and patient data evaluations based on interval techniques to consider missing variables in the given patient data sets. Most of the work done so far is described in [6, 7].

2.2 FuzzyKBWean

2.2.1 FuzzyKBWean Background

Patients require mechanical ventilation during surgery, when they are anaesthetized, and must be slowly weaned from mechanical ventilation after major surgery to a point when they can breathe spontaneously. At this point, the patients can be extubated. In other words, the tube that is placed in the trachea to ensure proper ventilation is removed [20, 24, 29]. The aim of an improved weaning process would be to make the transition from controlled ventilation to total independence (extubation) as smooth and brief as possible.

Using the expertise for computer-assisted weaning in an appropriate manner is a common problem for such applications [9, 10, 13, 14 16]. In order to formalize the knowledge of the system in an easier way, a knowledge acquisition tool, the so-called knowledge-based weaning editor FuzzyKBWEdit has been developed, which helps intensive care specialists to generate a fuzzy knowledge base.

When a know-ledge base has been set up, the editor generates a compiled (scanned and parsed) version of it. This executable version of the knowledge base's 'source code' is used as an interface for the FuzzyKBWean [17].

2.2.2 FuzzyKBWean Application

The computer system FuzzyKBWean is a real-time, open-loop knowledge-based control system that contains the knowledge and expertise of experienced intensive care physicians in computerized form. It offers proposals for ventilator control during the weaning process of patients after cardiac surgery. The respirator changes effected by the physician have to be entered into FuzzyKBWean as a feedback for this open-loop system.

The ventilatory mode used for weaning must allow spontaneous breathing and a gradual reduction of the amount of ventilatory support [11, 19]. In this expert system the APRV mode is implemented because it allows controlled and spontaneous ventilation with one unique mode. The adaptation to the specific needs of a patient necessitates only changes in three variables without any need for mode switching on the ventilator. The BIPAP (Biphasic Positive Airway Pressure) mode is an APRV mode equipped with a standard ventilator [23, 24]. This mode allows spontaneous inspiration during the entire respiratory cycle and, consequently, a very smooth and gradual change from controlled to spontaneous breathing.

The fuzzy knowledge bases established by using FuzzyKBWEdit consist of variables, values, and rules. The variables represent the physiological parameters of the patient and the respirator settings. The values are described in linguistic terms that are formalized by fuzzy sets. FuzzyKBWean uses the Sugeno Defuzzification method [11, 15] for defuzzification.

2.2.3 Data Input

The respirator settings and physiological parameters are taken as input at one-minute intervals from the Patient Data Management System (PDMS) Picis®. The PDMS Picis (Caresuite 97 Chart +, Paris, Barcelona, 1998) is in routine clinical use in the cardiothoracic ICU and collects data from all available monitoring devices. FuzzyKBWean analyzes these data and makes suggestions for appropriate respirator setting adjustments. The attending intensive care specialist is free to decide whether he will follow the advice (open-loop system).

3 Results and Discussion

3.1 FuzzyARDS System

The first system, a web application of FuzzyARDS is shown in Fig. 3. Based on the available FuzzyARDS study system, patient data sets are entered at the various study centers and evaluated in ARDS consensus meetings.

Patient data, that means 25 different parameters are measured, evaluated, and observed at four point of time. Another important facility of the system is the storage of the definition of entry criteria. The results yielded to better understanding of ARDS as a life threatening disease [2, 3, 4, 7]. Based on these results, FuzzyARDS is continuously adjusted to new derived medical knowledge [30].

The system is available for registered users (username and password are required) at: http://medexpert.msi.meduniwien.ac.at/fuzzyards .

The web-application is programmed in Java and uses a mySQL database for data storage.

The present patient's database consists of 280 patients from two different centers a set of 30 patients for testing. Two different sets of therapy entry criteria are used, in the first one eight, the second one consists six criteria. This now pre-evaluated minimal data set to describe severe ARDS patients based on the fuzzy set theory may be useful to evaluate patients for ECLA therapy or for another controlled ARDS-therapy. Some problems have still to be solved, e.g., the definition and incorporation of idle and delay functions in the on-line monitor to avoid oscillations in the patient states. The results of the various evaluations may be used for the design of further infinite deterministic fuzzy-automation. This automation should describe the state as well as the history of a single patient by combining different parameters over time.

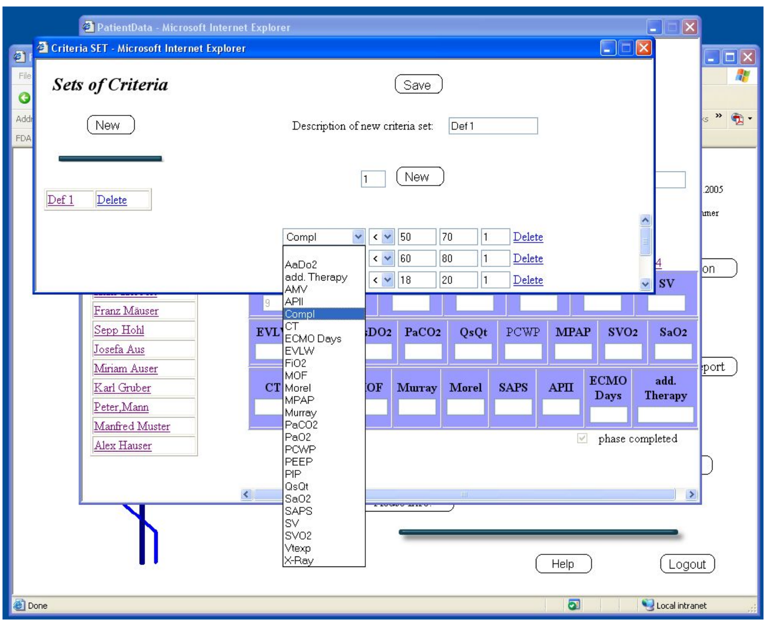


Fig. 3. FuzzyARDS web application (data input)

3.2 FuzzyKBWean System

The fuzzy control and advisory program FuzzyKBWean has been developed with Delphi® 6.0 running on Windows® platforms.

The bedside real time application of Fuzzy-KBWean is shown in Fig. 4.

The user interface has two main- units. The online (real time-data) unit, and a so called history (data base related) unit. It is possible to toggle between these units, so that always one or both of them have the focus. The top panel displays actual values and proposals, middle panel allows data review from any previous time point and, bottom panel displays key variables of the ventilation process together with the proposed new settings.

Fig. 4. FuzzyKBWean frame application

With the integrated knowledge acquisition tool FuzzyKBWEdit knowledge bases can be formalized with very little restriction. Nevertheless, knowledge bases have to be formalized syntax-guided in order to make them usable for the expert system, in our special case for the expert system FuzzyKBWean. When a knowledge base has been formed, the editor generates a compiled (scanned and parsed) version of it. This executable version of the knowledge base's "source code" is used as an interface for the computer-assisted expert system FuzzyKBWean. This entire concept permits the creation of various experimental versions of the knowledge bases. Furthermore, the interface can be easily modified for using other computer-assisted applications in future. The system is used for postoperative cardiac patients in an ICU at the Vienna General Hospital. The advantages of the system are its easy application and the generation of more specific knowledge bases, which allow smoother treatment of weaning patients. FuzzyKBWean is currently being tested with a pilot sample of 28 prospective randomized cases currently undergoing treatment. It can be found that the clinical staffs react with a longer delay to hyper- or hypoventilation then the program does. The mean delay in case of hyper- ventilation was 127 minutes, Standard Error of Mean (SEM) 34; the corresponding value for hypoventilation was 50 minutes (SEM 21). A large body of implicit medical knowledge was transferred to the fuzzy control system. The obtained results confirm the applicability of FuzzyKBWean to represent medical knowledge, thus rendering the weaning process transparent and comprehensible. Periods of deviation from the target are shorter with FuzzyKBWean. Nevertheless, the use of fuzzy sets provides a basis for a smoother adaptation of mechanical ventilation to the patients` advantage, since small changes in the ventilator settings are made continually. Manual settings cannot be very precise, because the minimal step to be set in that environment is one mbar. Only a closed-loop application, i.e. a direct connection between the FuzzyKBWean and the ventilator, would allow smooth adaptation continuously. When a robust performance is achieved in the current randomized trial, a transfer to other clinical settings is planned in order to fully validate the experimental concept.

4 Conclusion

In medicine, two fields of fuzzy applications were developed since the nineteen seventies: computer assisted diagnostic systems and intelligent patient monitoring systems. Both developments of Zadeh's rule of max-min composition, namely fuzzy relations and fuzzy control, have been applied in these areas. For obvious reasons, the available body of medical data (on patients, laboratory test results, symptoms, and diagnoses) will expand in the future. As mentioned above, computer-assisted systems using fuzzy methods will be better able to manage the complex control tasks of physicians than common tools. Most control applications in the hospital setting have to be performed within critical deadlines.

Fuzzy Logic in medicine is still a largely untapped area that holds great promise for increasing the efficiency and reliability of health care delivery. These medical applications described above, and others [19, 22, 23, 27, 28] showing generally promising results, the literature on fuzzy logic applications in medicine remains modest. Fuzzy logic provides a means for encapsulating the subjective decision making process in an algorithm suitable for computer implementation.

Furthermore, the principles behind fuzzy logic are straightforward and its implementation in software is relatively easy. Nevertheless, the applications of fuzzy logic in medicine are few. In addition, fuzzy logic may support the automation of some types of devices used in the delivery of health care services.

Acknowledgment. This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

1. Zadeh, L. A.: Fuzzy Sets. *Information and Control*, 8, 1965, pp. 338-353
2. Rensing H. and Bauer M.: Multiorganversagen Mechanismen, klinische Manifestation und Therapie- strategien, Der Anaesthesist Heft: Band 50, Nummer 11: Seiten: 648 - 655
3. Vincent J-L., et al. (1996): The SOFA (Sepsis-related Organ Failure Assessment) Score to Describe Organ Dysfunction/-Failure, Intensive Care Medicine 22: 707–710, http://www.sfar.org

4. Steltzer, H., Trummer, B., Höltermann, W., Kolousek, G., Friedrich, P., Lewandowski, K., Adlassnig, K.-P., Hammerle, A.F. (1999) Knowledge-based diagnosis and therapeutic recommendations with fuzzy-set theory methods in patients with acute lung failure (ARDS). Anästhesiologie Intensivmedizin, Notfallmedizin Schmerztherapie 34 (4), 218–223.
5. Steimann, F., Adlassnig, K.-P. (1994) Clinical Monitoring with Fuzzy Automata. Fuzzy Sets and Systems 61, 37–42.
6. Steltzer, H., Friedrich, P., Steimann, F., Rossaint, R., Adlassnig, K.-P., Hammerle, A.F. (1996) Definition of Severity and Compiling of Entry Criteria in Patients Undergoing ECLA: The Use of a Fuzzy Logic Expert System. Acta Anaesthesiologica Scandinavica 40 (Supplement 109), 125–126.
7. Trummer, B., Steltzer, H., Kolousek, G., Adlassnig, K.-P., Hammerle, A.F. (1998) Knowledge-Based Definition of ARDS Therapy Entry Criteria Using Fuzzy Set Theory. Acta Anaesthesiologica Scandina-vica 42 (Supplement 112), 238–239.
8. Steimann, F., Adlassnig, K.-P. (1994) Two-Stage Interpretation of ICU Data Based on Fuzzy Sets. In: AI in Medicine: Interpreting Clinical Data, Stanford 1994 Spring Symposium Series, American Association for Artificial Intelligence, Stanford University, 152–156
9. Becker K., Thull B., Käsmacher-Leidinger H., Stemmer J., Rau G., Kalff G., Zimmermann H-J., Design and validation of an intelligent patient monitoring and alarm system based on a fuzzy logic process model. Artificial Intelligence in Medicine 11 (1997) 33-53.
10. Dojat M., Pachet F., Guessoum Z., Touchard D., Harf A., Brochard L., NéoGanesh: a working system for the automated control of assisted ventilation in ICUs. Artificial Intelligence in Medicine 11 (1997) 97-117.
11. Dirankov D, Hellendorn H, Reinfrank M., An Introduction to Fuzzy Control, Springer Verlag, Berlin 1993
12. Hiesmayr M., Enhanced ventilatory efficiency of airway pressure release ventilation (APRV) versus SIMV after cardiac surgery, International Meeting of the American Thoracic Society, Seattle, (1995) 43-47.
13. Kollef M. H., Shapiro S. D., Silver P., St. John R. E., Prentice D., Sauer S., Ahrens T. S., Shannon W., Baker-Clinkscale D., ‘A randomized, controlled trial of protocol-directed versus physician-directed weaning from mechanical ventilation’, Critical Care Medicine 25 (1997) 567-574.
14. Linkens D.A., Abbod M.F., Peacock J.E., Intelligent System for Monitoring and Control in Anaesthesia. Medical & Biological Engineering & Computing, Supplement 2, Part I, (1999) 14 -19.
15. Ho J. L., Park J., Chen G., Robust Fuzzy Control of Nonlinear Systems with Parametric Uncertainties. IEEE Transaction on Fuzzy Systems, Vol. 9, No. 2. April 2001
16. Randolph A., Clemmer T., East T., et al: Evaluation of compliance with a computerized protocol: weaning from mechanical ventilator support using pressure support. Comput Methods Programs Biomed 57: 201-215, 1998.
17. Schuh Ch., Koller W., Zelenka Ch., Kolb M., Hiesmayr M., Adlassnig K.-P.: ‘Knowledge Acquisition for Crisp- and Fuzzy-Controlled Weaning in Intensive Care Units’, Proceedings of the World Automation Congress - WAC’ 2000, TSI Press, Albuquerque, (2000) 583-588.
18. Eddy D. M., Clinical decision making from theory to practice: a collection of essays from the Journal of the American Medical Association. Sudbury, MA: Jones and Bartlett; 1996.
19. Shan S., Lakshminarayan S., Petty T. Weaning from Mechanical Ventilation; JAMA 235 (1976) 2208-2212.

20. Slutsky A. S., Consensus conference on mechanical ventilation, Intensive Care Medicine, 20 (1994) 64-79.
21. Helgason CM, Jobe TH. Causal interactions, fuzzy sets and cerebrovascular ‘accident’: the limits of evidence-based medicine and the advent of complexity-based medicine. *Neuroepidemiology* 1999; 18:64–67.
22. Ohayon MM. Improved decision-making processes with the fuzzy logic approach in the epidemiology of sleep disorders. *J Psychosom Res* 1999; 47:297–311.
23. Morris A: Computerized protocols and bedside decision support. Crit. Care Clin. 15: 523-545, 1999.
24. Putensen C., Wrigge H. New modes of mechanical ventilation. *Intensive Care Med 1998*; 24 (12): 1342.
25. Esteban A, Alia I, Tobin MJ, Gil A, Gordo F, Vallverdu I, et al. Effect of spontaneous breathing trial duration on outcome of attempts to discontinue mechanical ventilation. Spanish Lung Failure Collaborative Group. *Am J. Respir. Crit. Care Med* 1999; 159 (2):512-8.
26. Velasevic DM, Saletic DZ, Saletic SZ. A fuzzy sets theory application in determining the severity of respiratory failure. *Int. J Med Inf.* 2001; 63:101–107.
27. Long W. The Probability of Disease. In: *Symposium on Computer Applications in Medical Care*. IEEE, 1991:619–623.
28. Miller R, Pople HE, Myers J. INTERNIST-1. An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine. *New England Journal of Medicine* 1982; 307:468-476.
29. Spiegelhalter D. J., Bayesian Analysis in Expert Systems. MRC Biostatistics Unit, Institute of Public Health, Cambridge, 1992.
30. Bone R.C., Balk R.A., Cerra F.B., Dellinger R.P., et al. (1992): American College of Chest Physicians / Society of Critical Care Medicine Consensus Conference: Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis, Critical Care Medicine 20: 864–875

Part VII

Joint Model-Based and Data-Based Learning: The Fuzzy Logic Approach

Process Monitoring Using Residuals and Fuzzy Classification with Learning Capabilities

Joseph Aguilar-Martin[1], Claudia Isaza[1,*], Eduard Diez-Lledo[1], Marie Veronique LeLann[1], and Julio Waissman Vilanova[2]

[1] LAAS-CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse, France
[2] Centro de Investigación en Tecnologías de Información y Sistemas ICBI-UAEH, *CU, Carr. Pachuca-Tulancingo Km. 4.5 C.P. 42084, Pachuca, Hidalgo, México*
{aguilar, cisaza, ediezlle, mvlelann}@laas.fr, julio@uaeh.reduaeh.mx

Abstract. This paper presents a monitoring methodology to identify complex systems faults. This methodology combines the production of meaningful error signals (residuals) obtained by comparison between the model outputs and the system outputs, with a posterior fuzzy classification. In a first off-line phase (learning) the classification method characterises each fault. In the recognition phase, the classification method identifies the faults. The chose classification method permits to characterize faults non included in the learning data. This monitoring process avoids the problem of defining thresholds for faults isolation. The residuals analysis and not the system variables themselves, permit us to separate fault recognition from system operation point influence. The paper describes the proposed methodology using a benchmark of a two interconnected tanks system.

Keywords: Faults Identification, Faults Isolation, Residuals, Fuzzy Classification.

1 Introduction

Most industrial plants have a high degree of complexity, because of the great number of elementary systems interconnected. Nevertheless in many cases there exists the possibility to obtain a simulation model for some unitary processes. Therefore the monitoring of the plant can be based on the individual fault detection for each sub-process. The system faults could then be detected by means of residuals generated from the comparison of system and model outputs.

The residuals are "artificial signals" reflecting the potential faults of the system. These signals should be close to zero when no fault occurs but show "significant" values when the underlying systems change [1]. There exist many techniques for residual generation. One of the most common is the theory of observers, which consists on defining a set of observers one of them sensitive to one fault [2]. However, this approach becomes complex when applied to non-linear dynamic systems composed of different subsystems. Another approach is the Kalman filter, which estimates the system state based on the normal model of the system [3]. The system should be

[*] Professor at the Universidad de Antioquia, Medellín-Colombia.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 275–284, 2007.
springerlink.com

modelled on the basis of all the available possible system models under all possible changes. The parameter estimation is as well usual in residual generation. But once more, accurate dynamic models are required to apply those methods, which are computationally very intensive for large processes [4]. Hardware redundancy and voting schemes are based in sensor redundancy and are widely used in industry. The problem of those methods is the ocurrence of false alarms [5].

The Parity Relations approach consists on checking the parity (consistency) of the plant models with sensor outputs (measurements) and process inputs. Under ideal steady operating conditions, the so-called residuals or the value of the parity equations is zero. Many studies have been carried out showing great results in fault isolation. [6],[7]. Despite of that, it is absolutely necessary to define a threshold for each operation point and failure.

In this paper we present a monitoring methodology which compares the output model signals with the actual measures to generate meaningful error signals (residuals). The residuals are classified to identify and isolate the system faults. This method identifies each faulty mode without defining fault isolation thresholds. Moreover, a fuzzy classification of the residuals allows including expert knowledge in an initial learning phase and permits results interpretation. The fuzzy classification method used in this work is LAMDA (Learning Algorithm for Multivariable Data Analysis) [8]. This approach has a similar performance to other recognized classification methods (e.g. GK Means, ANN, etc). Its easy parameters interpretation is interesting to the learning process when knowledge of the data space distribution is minimal. The LAMDA classes definition permits an immediate relation with the process variables, this feature is desirable to the fault cause identification. Moreover, the LAMDA algorithm offers the possibility of including classes not a priori learned. Therefore, patterns of residuals not presented in the learning phase may generate automatically a new class in the recognized phase.

The monitoring methodology is proposed as follows: (i) first of all, an elementary process model is developed, (ii) selection and pre-processing of interpretable signals for the comparison between model and real plant, (iii) description of failures that should detected and functional states, (iv) after that, learning procedure to create the space partition (v) finally, functional state recognition procedure. To illustrate this methodology we shall follow all those steps applied to the hydraulic system of two connected tanks shown in figure 1 [9]. The paper is organised as follows: Section II presents the model and the variables selected. Section III describes the LAMDA fuzzy classification method. Section IV shows the residuals signals and the results for the learning and recognition phases, and finally the conclusions and perspectives are presented.

2 Model Development

The objective of this system is to produce a continuous flow Q_o. The two tanks T_1 and T_2 are linked trough a pipe transporting a flow Q_{12}. The tank T_1 receives water by means of a pump P_1 The level set point is h_{1c}, its real level is controlled by a proportional and integral regulator PI acting on the pump and modifying the flow Q_p.

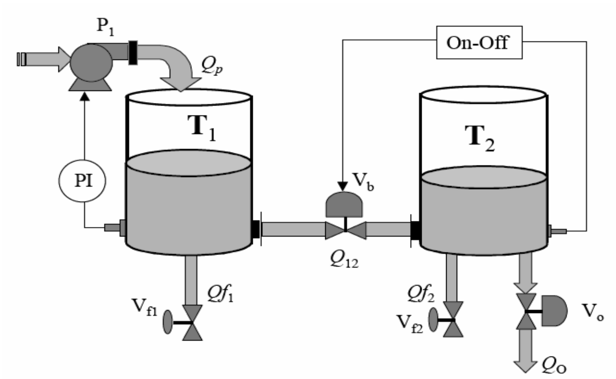


Fig. 1. Scheme of the two-tank System

The flow Q_{12} between both tanks is regulated by means of an ON/OFF controller of valve *Vb* that tries to maintain the level *h2* of tank *T2* at some value between 0.09 m and 0.11 m. The output flow *Q*o depends of valve *Vo* that is here considered always open. Valves *Vf1* and *Vf2* are set to simulate respectively leaks on the tanks, they remain closed in the theoretical faultless model. A table must be built, where the variables and constants, their values and units will be quoted. The building of that table is essential for the analysis of the completeness of the problem.

Table 1. Model Values and Process Parameters

Symbol	Description	Value	Units
C_{vb}	Hydraulic Flow of valve V_b	1.5938 10^{-4}	$m^{3/2}$ / s
U_b	Position of the valve V_b	On-Off	{0,1}
C_{vo}	Coefficient of flow in V_o	1.5964 10^{-4}	$m^{3/2}$ / s
U_o	Position of the valve V_o	On-Off	{0,1}
U_p	Control of the pump flow	variable	m^3 / s
$A_{(1,2)}$	Cross-section of the cylindrical tank $T_{(1,2)}$	1.54 10^{-2}	m^2
$h_{(1,2)}$	Water level for tank $T_{(1,2)}$	variable	m
$h_{(1,2)max}$	Maximal water level for tank $T_{(1,2)}$	0.6	m
Q_p	Pump P_1 flow	variable	m^3 / s
Q_{pmax}	Maximum pump P_1outflow	0.01	m^3/ s
$Q_{f(1,2)}$	Leakage flow in tank $T_{(1,2)}$ (failure)	0.0001	m^3 / s
h_c	PI controller set point	variable	m

2.1 State Equations of the System

Let us consider the classical state space form of a dynamical system given by equation 1

$$\begin{cases} \dot{x} = f(x,u) \\ y = g(x) \end{cases} x(0) = x_0 \tag{1}$$

The state variable is here the vector $x=[h_1\ h_2]^T$. It is constrained by $0 \le h_i \le h_{imax}$. The measurement vector is given by the measures of the water heights h_1 y h_2 on both tanks. Possible measurement errors ε are considered so that the state observation equation is (2)

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} . \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{2}$$

The control variables are gathered in the vector $u=[Q_P\ U_b\ U_o\ U_P]^T$. The dynamic equations of the system express the volume flow balance given in (3).

$$A_1\dot{h}_1 = Q_p - Q_{12} \text{ and } A_2\dot{h}_2 = Q_{12} - Q_0 \tag{3}$$

The flow of the pump is proportional to the control action $Q_p = C_p \cdot U_p$. The characteristic relation between the pump control signal and its real flow is constrained to lie in the interval $\lfloor 0, Q_{p\max} \rfloor$. The leakage flows Q_{f1}, Q_{f2} will only be added as constant values when failures are to be simulated. The PI regulator is represented by $U_p = K_p(h_{1c} - h_1(t)) + K_i \int (h_{1c} - h_1(t))dt$. As the output valve is open at the value 1, we have $Q_0 = C_{vo} \cdot \sqrt{h_2}\ .Uo$. If the output valve is all time opened (Uo=1m^3/s), and the level of the tank T2 is set around the reference, a mean flow of Qo=5.048e-5m^3/s is expected. Level h1c is the set point of T1and the applied PI parameters are Kp = 10^{-3}m^{-1} and Ki = 5 .10^{-6}(m.s)$^{-1}$. A non linear relation connects the flows with the state and the opening valve is assumed to be proportional to the control action. That relation is $Q_{12} = U_b \cdot \sqrt{h_1}$.

For the treatment of complex non-linear systems a representation in linear submodels is suitable. However, when subsystems should be based on complex physical equations, those subsystems could be defined alternatively using fuzzy logic techniques as rule based models (e.g. Mamdani-Sugeno type) or fuzzy models obtained from historical data [10]. The non-linear complete system could be represented by a Takagi-Sugeno (TS) fuzzy model [11]. The linear submodels are identified by means of input-output data clustering. The parameters of the classes are used to tune the Takagi-Sugeno fuzzy model [10]. The global model output is the aggregation of submodels outputs in function of the activation degree of each submodel. Consequently,

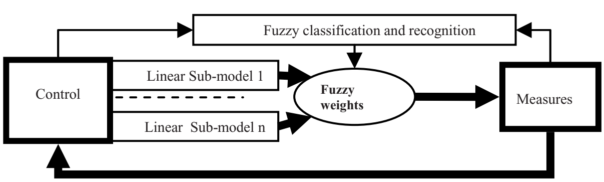


Fig. 2. Scheme of Fuzzy Model for Monitoring Systems

a complex system can be represented with an aggregation of linear submodels when the complex model is not available. Figure 2 shows the general representation of a fuzzy model.

2.2 Interpretable Signals Selection

Thanks to all the information collected up to this point we are enabled to build a simulation model. The benchmark model used in simulations is presented in Figure 3.a. The model presents the connections between the subsystems. That makes unnecessary a mathematical representation of the complete system. Then the residuals in several situations can be visualized in Figure 3.b., where the 7 observed variables Q_p, U_b, U_o, y_1, y_2, h_{1c}, and U_p are shown. It must be remarked that in the situations considered, the variables U_o and h_{1c} remain unchanged hence providing little information. Therefore, it has been decided to exclude them. The ON/OFF regulator exhibits variable U_b excessively sensitive to disturbances: Moreover this control variable has little effect on the global performance since this kind of regulator is extremely robust. Finally only the 4 variables Q_p, y_1, y_2 and U_p will be used for monitoring. In many practical cases a pre-processing of the selected variables might be necessary, such as filtering of spurious values or outliers elimination. For simplicity this action is here avoided.

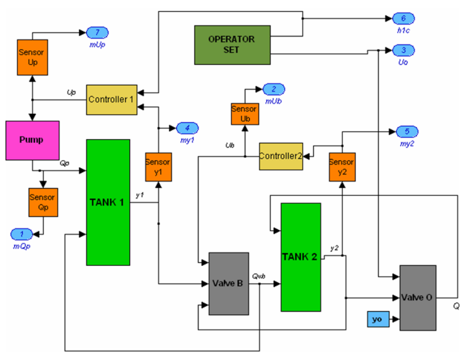


Fig. 3a. Model System

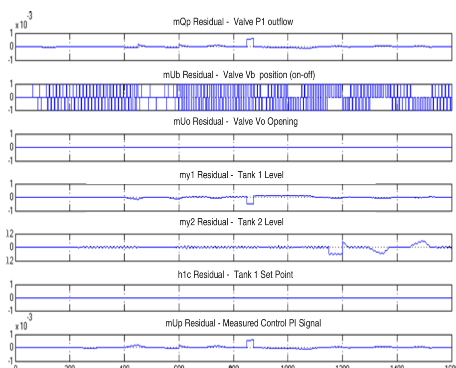


Fig. 3b. Two Tanks System Signals

3 Learning Algorithm

The vector of residuals is classified so that each fault mode could be associated to one class. Consequently the fault identification would be realised unequivocally. The classes corresponding to normal operation modes are as well included.

3.1 Fuzzy Classification Algorithm

The algorithm used is LAMDA [8], it is based on the evaluation of the adequacy of an element to each class. The global adequacy (GAD: *Global Adequacy Degree*) of an element is a function of the marginal adequacies (MAD: *Marginal Adequacy Degree)* representing the adequacy of each data vector component to a given class. The function that aggregates the MADs to form the GAD is called a *connective*; it belongs to Fuzzy Logic operator family. An element is assigned to the class with maximum GAD. A Non Informative Class (NIC) is always provided so that all elements in the space hold the same adequacy to the NIC. The NIC gives a minimum threshold for the class attribution. Using passive recognition an element is associated to the NIC when it is considered non-recognized. If the element is recognized, it is then associated to an existing class. In active learning the parameters of this class can be updated with the element values, in the other hand if the element is not recognized a new class is created.

In self-learning a new class would be created and initialised with that element and the corresponding NIC parameters. Therefore, with the self-learning mode it is possible to start a classification with no previous information, having only the NIC class when the first element is processed, either to introduce the expert knowledge. Figure 4 presents a flow diagram of the algorithm principle.

In a single direction *j* of the data space the adequacy of an element *x* to a class C is a parametric function characterized by a vector of parameters ρ. Then the MAD's are marginal adequate functions ($0 \le MAD(x_j|C) \le 1$). It is assumed that $\bar{x}_j$ is a prototype for class *C* in direction *j* (C_j). such that $MAD(\bar{x}_j|C) \ge MAD(x_j|C)$. Moreover, as all the data given by measurement devices is such that a maximum and a minimum value exist, without loss of generality, it will be assumed that $0 \le x_j \le 1$. Whenever the sensor gives a value y_j it will be transformed into $x_j = (y_j - y_{j\min})/(y_{j\max} - y_{j\min})$.

Fuzzy binomial function: Given a $x = \begin{cases} 0 \\ 1 \end{cases}$ and class *C* is characterized by $\rho = prob[x=1]$ then $MAD\ [x|\rho] = \begin{cases} 1-\rho \\ \rho \end{cases}$. A monotonous function satisfying those boundary conditions can be taken as fuzzy measure distribution that interpolates this binomial probability for $x \in [0,1]$. Following this requisite the function given in equation 4 is proposed:

$$MAD_{Xcj}\left[x_j \middle| \rho_{cj}\right] = \rho_{cj}^{\,xj}\left(1-\rho_{cj}\right)^{(1-xj)} \tag{4}$$

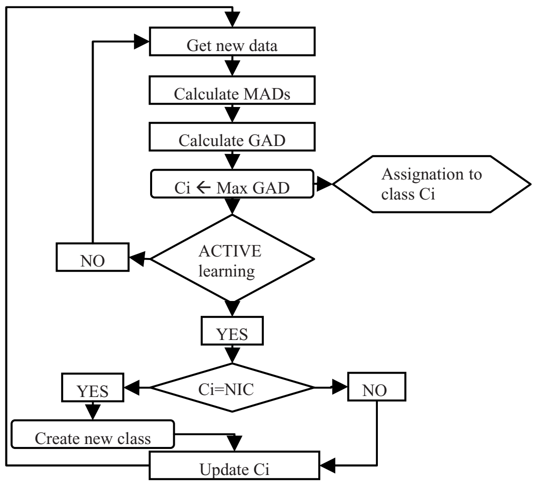


Fig. 4. Algorithm Flow Chart

3.2 Classification Parameters Estimation (Learning)

The estimation of the MAD function parameters is a statistical problem. It is possible to determine the parameters for each class using two methods: direct or iterative. For the direct approach let us consider a sequence of measurements $\{x(1),\ldots x(t),\ \ldots,\ x(T_c)\}$ assigned to a class, that for simplicity will be called C and T_c is the number of elements assigned to it. This sequence of x values are indexed by they order of arrival to class C, by a discrete variable t that will be called time. The parameter ρ can be estimated by the mean value of the sequence:

$$\rho_C(T_c) = \frac{1}{T_c}\sum_{t=1}^{t=T_c} x_j(t) \tag{5}$$

The parameter ρ_{cj} can be obtained iteratively (active learning) by equation 6

$$\rho_{cj}(T_c+1) = \rho_{cj}(T_c) + \frac{1}{T_c+1}\left(x_j(T_c+1) - \rho_{cj}(T_c)\right) \tag{6}$$

3.3 Fuzzy Aggregation Functions

An element x is a vector $x=[x_1 \ \ldots \ x_n]$ so, once the marginal adequacies are obtained there is a MAD vector $MAD(x\,|C) = [MAD(x_1|C) \ \ \ldots \ \ MAD(x_n|C)]$. An aggregation function γ combines the marginal adequacies such that the GAD value represents the membership degree of each element to each class. [12]

$$GAD(x|C) = \gamma(MAD(x_1|C),\ldots,MAD(x_n|C)) \tag{7}$$

Fuzzy Logic connectives are fuzzy versions of the Boolean logic connectives, particularly, intersection (*AND*) γ and union (*OR*) β. In Fuzzy Logic γ is a t-norm and β its dual. Two choices are selected in this work. As t-norms are associative, the extension for *N* arguments is straightforward.

AND	OR
$\gamma(a,b) = a.b$	$\beta(a,b) = a + b - a.b$
$\gamma(a,b) = \boldsymbol{min}[a,b]$	$\beta(a,b) = \boldsymbol{max}[a,b]$

To obtain a *compensated* aggregation function and provide classification algorithms with compensated properties; *mixed connectives* consist in a lineal interpolation between t-norm and t-conorm $\lambda(a,b) = \alpha.\gamma(a,b) + (1-\alpha).\beta(a,b)$, where $0 \le \alpha \le 1$. Parameter α is called *exigency*. When $\alpha = 1$ (the greatest exigency) the AND operator is applied, the intersection being the most restrictive set.

4 Example Results

4.1 Failures Description

By using the simulation model presented in Section 2 it is possible to have a faultless reference of the selected signals. To generate the necessary data for learning, we duplicate the simulation model: a first one faultless called "model", and a second one in which we activate the possibility of simulating failures “System". By comparing the corresponding observable signals, "error" signals are generated. As 4 variables have been selected, a 4 dimensional "error" or "residual" time varying vector is obtained. A sampling time Δ is chosen and the corresponding first order hold is applied. Given a simulation time period of duration *T* a sequence of *N*-dimensional (here *N=4*) values of the residual signal can be stored in a file having the form of a table whose dimensions are $N \times \frac{T}{\Delta}$.

The failures to be considered involve the sensors, the PI regulator, the valve B and leaks on the tanks. Table 2 shows the failures as well as their initial and end instants, during the learning period given in number of samples.

Table 2. Failures Description

Fault	Element	Data Set 1	Data Set 2	Class
0	Normal operation	1-	1-	**1**
1	Sensor Qp	100-150	200-250	**2**
2	Sensor Up	250-300	550-600	**3**
3	Pump	400-450	900-950	**4**
4	Control PI	550-600	1250-1300	**5**
5	TANK1	700-750	1600-1650	**6**
6	Sensor y1	850-875	1950-1975	**7**
7	Sensor Ub	1000-1050	2075-2125	**8**
8	Sensor y2	1150-1200	2625-2675	**9**
9	TANK 2	1300-1350	2975-3025	**10**
10	Valve B	1450-1500	3325-3375	**11**

4.2 Learning and Failures Recognition

A simulation is done while the failures are introduced only to the "system" at the instants shown in Table 2. In Figure 5 the trajectories of the normalized residual vector are shown in the lower part. Looking at Figure 5 one can remark that:

- Some failures are better recognized than others
- When normal operation is recognized it corresponds to the real situation
- Some false alarms occur mostly for failure 6, a signals pre-treatment can be suitable.
- The black line al level zero corresponds to unrecognised situations (NIC).

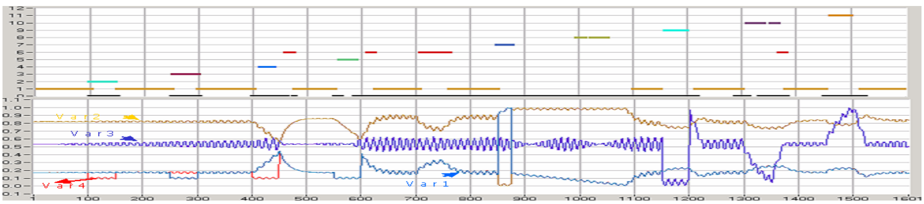


Fig. 5. Variables and Classes. **Variables: Var1. Controller PI signal**, Var.2 Water Level T1(my1),Var 3.Water Level T2 (my2), Var 4. Pump flow P1 (mQp).

The next figure shows graphically, the prototypes of each class presented to the operator. This description allows the operator to identify each failure taking in account just the variables distribution of each class.

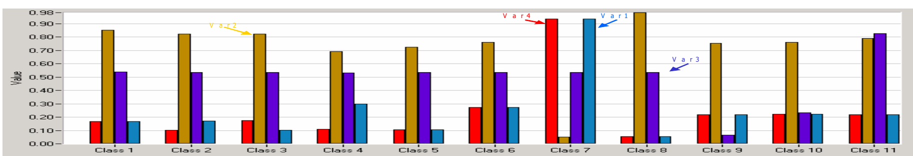


Fig. 6. Classes Prototypes. **Variables: Var1. Controller PI signal**, Var.2 Water Level T1(my1), Var 3.Water Level T2 (my2), Var 4. Pump flow P1 (mQp).

To illustrate the correct behavior of the monitoring system, an experiment is done in which 3 changes of set point are introduced. Figure 7 shows the recognized data results. The faults time are presented in table 2 (Data set 2).

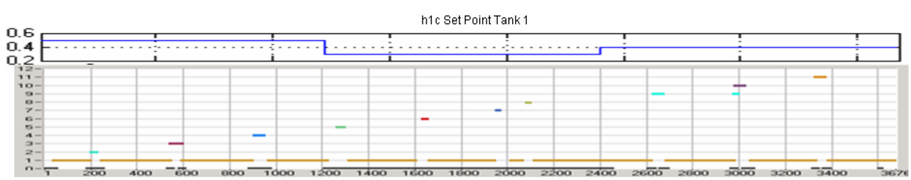


Fig. 7. Faults Recognition –Operation Point Changes

5 Conclusions and Perspectives

The combination of a dynamic model and a data mining procedure based on learning classification algorithms gives a powerful methodology for process monitoring. Moreover the introduction of fuzzy concepts provides the method with versatility and power of interpretation.

The methodology proposed for fault detection is not dependant from the residuals generation technique. Therefore, we found interesting the possibility of using modelling methods based in historical data for complex systems. The use of an adaptive fuzzy model would give us the possibility to include the experience of the process expert and forecast operation modes. The advantage of the presented classification method is the learning possibilities using the non informative class (NIC). Thanks to this, the monitoring system becomes more robust and more adaptable.

The model, often non linear, has the Takagi –Sugeno structure, and the partition of the state space can use the fuzzy partition obtained by a fuzzy learning algorithm as LAMDA. The learning classification and recognition part of the method, by its fuzzy nature, gives, not only information about the present situation, but also the membership for other possible classes, the vector of instantaneous memberships can be used to give a weight to the decision to be taken as reaction to the failure detection. This option is considered an interesting topic for a future study.

References

1. Venkatasubramanian V., Rengaswamy R.. Yin K, Kavuri S.: A Review of Process Fault Detection and Diagnosis Part I. Computers and Chemical Engineering, Vol.27 (2003)
2. Frank P. M.:. Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-Based Redundancy- A Survey and Some New Results, Automatica 26, (1990) 459-474.
3. Basseville M., Benveniste A.: Detection of Abrupt Changes in Signals and Dynamic Systems, Lecture Notes in Control and Information Sciences: 77, Springer-Verlag. (1986).
4. Young, P.: Parameter Estimation for Continuous Time Models a Survey, Automatica 17 (1), (1981) 23-39.
5. Desai M., Ray, A.: A Fault Detection and Isolation Methodology-Theory and Application, Proceedings of American Control Conference, San Diego (1984) 262-270.
6. Yin K.: Minimax Methods for Fault Isolation in the Directional Residual Approach, Chemical Engineering Science 53, (1998) 2921-2931.
7. Gertler J., Singer D.: A New Structural Framework for Parity Equation-Based Failure Detection and Isolation, Automatica 26, (1990) 381-388.
8. Aguilar-Martin J., Lopez de Mantaras R.: The Process of Classification and Learning the Meaning of Linguistic Descriptors of Concepts, Approximate Reasoning in Decision Analysis North Holland (1982) 165-175
9. Ould Boumama B., Mrani Alaoui R., Taillibert P., Staroswiecki M.: Diagnosis of a Two Tank System, CHEM FDI Benchmark, CNRS, France.
10. Babuska R.: Fuzzy Modeling for Control, Control Engineering Technology and Systems, Delft University of Technology, Kluwer Academic Publishers (1998).
11. Takagi T., Sugeno M.: Fuzzy Identification of Systems and its Applications to Modeling and Control, IEEE Transactions on Systems, Man and Cybernetics, 15, (1985) 116-132.
12. Piera N., Aguilar-Martin J.: Controlling Selectivity in Non-standard Pattern Recognition Algorithms, IEEE Transactions on Systems, Man, and Cybernetics, Vol.21 N.1 (1991)

Part VIII

Fuzzy/Possibilistic Optimization

Possibilistic Worst Case Distance and Applications to Circuit Sizing

Eva Sciacca[1], Salvatore Spinella[2], and Angelo Marcello Anile[1]

[1] Dipartimento di Matematica e Informatica, Università degli studi di Catania, Viale A. Doria 6, I 95125 Catania, Italy
sciacca@dmi.unict.it
[2] Consorzio Catania Ricerche, via A. Sangiuliano 262, I 95124 Catania, Italy
spins@unical.it

Abstract. The optimization methodology proposed in this work is inspired to [1] and is named Possibilistic Worst-Case Distance (PWCD). This scheme has been tested on an application related to the MOS device sizing of a two stage Operational Transconductance Amplifier circuit (OTA) [2]. In order to model the uncertainties arising from circuit parameter simulations the fuzzy set theory, introduced by Zadeh [3], has been used. A linearization of the circuit performances as function of circuit parameters has been fitted as suitable approximation in a finite range, this choice was suggested to reduce the computational cost related to simulations of the real design. By means of linearization the circuit performances were fuzzyfied and a possibility measure of performance failure was minimized. The proposed case study will show that the possibilistic approach to the worst case analysis, even though less accurate for indirect yield estimation with respect to the probabilistic one, can identify an optimal design in yield terms. Furthermore the possibilistic methodology allows to develop calculation without any statistical hypothesis or sensitive analysis.

1 Introduction

The new technologies of MOS with shrinking dimensions have accelerated in the last decade and most likely they point at a typical dimension significantly lower than 80-100 nm in the next few years. Under 80-100 nm the process variations start to build-up and have a sizable effect on circuit's performances. Therefore the exploration of advanced circuits design is mandatory in order to anticipate the challenge of future behaviour of circuit based on nanoscale CMOS devices. Examples of these challenges are parasitic, process variation and transistor reliability. It is necessary to have models for MOS devices able to predict accurately new physical effects arising from this miniaturization and easy to integrate into simulation flow.

2 Worst Case Analysis and Possibility

The analysis of the *worst case* scenario is a fundamental step of any design process [4, 5]. The worst case of a manufacturing process is the failure of a performance specification in the inspection test after manufacturing. In order to prevent this event,

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 287–295, 2007.
springerlink.com

many simulations are carried out on the design to increase the manufacturing yield [5]. The computational cost of these simulations are very high because they simulate statistically process variations and increase the yield asymptotically with respect to a random process or deduce an indirect measure for the yield through statistical assumptions [6].

One assumes a simulation process involves:

- **controllable design parameters $\mathbf{X_c}$**, which is possible to define in reliable way with respect to the manufacturing process, for instance the geometrical structure of a device;
- **uncontrollable model parameters $\mathbf{X_u}$**, which are known in terms of confidence interval for the simulated model, their values are fit from experimental data for each technology node and then fixed over a range of process variations;
- **operational parameter $\mathbf{X_o}$**, which are operational range where the performances must be maintained, for instance the temperature or the power supply.

With respect to these classes of parameters the performances can be defined in functional terms as:

$$P = y(X_c, X_u, X_o) \tag{1}$$

The uncertainty concerning the uncontrolled parameters and the necessity to cover a wide range of operational parameters compel to consider these parameters as fuzzy numbers and to interpret the previous formulation as:

$$\tilde{P} = \tilde{y}(X_c, \tilde{X}_u, \tilde{X}_o) \tag{2}$$

where $\tilde{P}$ is the fuzzy representation of the performances and only the controllable parameters are considered crisp.

In order to deal with design specifications it is necessary to compare the fuzzy numbers representing the performances with crisp numbers representing the design constraints and give a measure of satisfaction of these constraints. For this purpose the possibility measure of failure with respect to the specification constraints can give useful information to improve the yield and design [7]. Note that a fuzzy number may also be considered as the trace of a possibility measure Π on the singletons (single elements) x of the universal set X. When a possibility measure is considered, its possibility distribution π is then interpreted as the membership function of a fuzzy number $\tilde{B}$ describing the event that Π focuses on, as follows:

$$\Pi(\{x\}) = \pi(x) = \tilde{B}(x), \qquad \forall x \in X \tag{3}$$

The possibility measure of a crisp number being smaller or equal to a fuzzy number B is then defined as follows [8]:

$$\Pi_{\tilde{B}}([x, +\infty)) = \sup_{y \geq x} \tilde{B}(y), \qquad \forall x \tag{4}$$

Based on 3 and 4, given a representation of the performance P and maximal failure specification performance P_f then the possibility measure of failure of this performance is a measure for the event $P > P_f$, hence

$$\Pi_{\tilde{P}-P_f}([0,+\infty)) = \sup_{y \geq 0}(\tilde{P} - P_f)(y), \qquad \forall x \tag{5}$$

From possibility measures of all performance failures it is possible to deduce a vector of measures $(p_1, \ldots, p_n)$ for a given design. A metric L_n can summarize all of them and it is used as target of the optimization process, for instance $n = 2$ represents the Euclidian norm.

3 Fuzzyfication Through Linearization and Sampling

In order to model with fuzzy numbers the uncertainty arising from simulation design a linearization of the performance can be used as suitable approximation. The linearization is fitted with respect to the uncertain parameters (uncontrollable and operational) to give an estimation of the behaviour of the performance as function of them. The scheme is the following

$$\overline{P} = \overline{Y}_{X_c}(X_u, X_o) = q + X_u \cdot A_u + X_o \cdot A_o \tag{6}$$

where the linearization $\overline{Y}$ of the performance P is carried out as linear function of $\boldsymbol{X_u}$ and $\boldsymbol{X_o}$ in a finite parameter range. The linearization coefficients represented in the vectors $\boldsymbol{A_u}$, $\boldsymbol{A_o}$ and the constant q are computed through a latin hyper-cube sampling in order to reduce the number of samples in a N-dimensional sparse space. This sampling scheme generates a multidimensional distribution of parameters where there is only one sample in each line parallel to the axes. Notice that these kinds of rough analysis are often used in circuit design.

From this linearization, the fuzzyfication is carried out using a uniform random sampling on the parameter ranges and giving a fuzzy representation of the linearized performance enveloping its sampling by interval. The fuzzy map is constructed by α-level considering the minimum median interval which envelops a fraction (1- α) of the linearized performance [9].

The possibilistic worst-case parameter sets for all specifications determine the worst-case behaviour with an accuracy according to the underlying method for performance evaluation, i.e., exactly when using numerical simulation or approximately when using performance macromodeling techniques.

4 The OTA

The BSIM model card involved in circuit simulation has hundred of parameters to characterize the I-V curve of a MOSFET device, but only a tens of them are critical and determine the technology scaling design. In particular the channel length (L_{eff}), the oxide capacity (T_{ox}), the threshold tension (V_{th0}), and the drain source resistance (R_{dsw}) are essential to characterize the technology process [10]. Hence the uncertainty of these parameters must be taken into account in order to determine the optimal device sizing to reach good performances in a technology scaling.

The computational cost related to the statistical representation of the technology parameters requires a methodology to reduce the number of simulations. Furthermore a design methodology should avoid to point deterministically towards unfeasible over-designs because this could have the opposite effect blocking the optimization process at initial stages. The previous linearization and fuzzyfication was found useful for that purpose.

Table 1. OTA parameters: type, ranges, unit

Parameter	Type	Ranges	Unit
temp	Operational	0 - 50	*Degree*
V_{V1}	Operational	3.5 – 4.2	*V*
L_{eff}	Technological	0.9 ± *perc*	*μm*
NMOS T_{ox}	Technological	9. ± *perc*	*nm*
NMOS V_{th0}	Technological	0.6322 ± *perc*	*V*
NMOS R_{dsw}	Technological	650 ± *perc*	*ohm · μm*
PMOS T_{ox}	Technological	9. ± *perc*	*nm*
PMOS V_{th0}	Technological	-0.6733 ± *perc*	*V*
PMOS R_{dsw}	Technological	460 ± *perc*	*ohm · μm*
$W_{1b}= W_{1a}$	Geometrical	0.6 – 20	*μm*
W_3	Geometrical	0.6 – 20	*μm*
W_5	Geometrical	0.6 – 20	*μm*
W_4	Geometrical	0.6 – 20	*μm*
$W_{2b}= W_{2a}$	Geometrical	0.6 – 20	*μm*
C	Geometrical	1 – 15	*pF*
R	Geometrical	2- 40	*KΩ*

Formally, let **W**, *C* and *R* be the design parameters respectively the MOSFET widths, the capacity and the resistance of the compensation net. Let **o** be the operational parameters (temperature *temp* and supply voltage V_{V1}). And let **t** be the process parameters (L_{eff}, T_{ox},V_{th0}, R_{dsw}). The uncertainty of process parameters is expressed as percentage of a reference value [11] and it defines a range of uncertainty. Tested values are ± 3%, ± 5%, and ± 7%. Table 1 shows the ranges which envelope the uncertainty of process and operational parameters and ranges for geometrical design parameters.

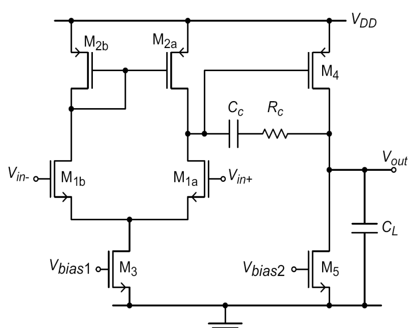


Fig. 1. Two stage OTA net topology

The figure 1 shows the circuit net of the two stage OTA [2]. Some of the main performances of this very useful electronic component are:

- **Gain@100Hz**,which is the base gain of the amplification;
- **Phase margin**, which is related to the circuit stability and to the parasitic effects like cross coupling (the main reason of circuit failure);
- **Unity gain frequency**, defined as the frequency value where the gain is equal to one;
- **Power dissipation,** which is a very important performance for all devices supplied by batteries.

The circuit performances represented by a vector P are computed by simulation Y

$$P = Y(w,t,o) \tag{7}$$

Where:

- w represent the geometrical parameters of the MOSFET devices and RC components of the net (*controllable*),
- t represent the uncertain technological parameters related to the manufacturing (*uncontrollable*),
- o represent the operational parameters whose range must be covered with acceptable performances (*operational*).

Therefore it is possible to proceed with linearization and fuzzyfication introduced in the previous section in order to obtain a fuzzy vector representing the performances of a given OTA circuit based on specific sizing defined by controllable parameters. It is possible to define a functional operator *Linearize* as:

$$a = Linearize_w(Y, I_o, I_t) \tag{8}$$

where I_o (operational) and I_t (technological) are the intervals defined in table 1 by ranges. The fuzzyfication is then carried out by sampling and building the median interval as described in the previous section.

5 PWCD Pseudo-code

The following pseudo-code explains the main functionality of the Possibilistic Worst-Case Distance methodology, that is the procedure which computes the optimization function to be minimized by the Simplex method with Simulated Annealing [12]:

FuzzyCircuitFun(***w***, $\boldsymbol{inf_{t,o}}$, $\boldsymbol{sup_{t,o}}$)

1. coefficients:= *LinearizeFun*(***w***, $\boldsymbol{inf_{t,o}}$, $\boldsymbol{sup_{t,o}}$, *CircuitFun*)
2. r:= random variables between $\boldsymbol{inf_{t,o}}$ and $\boldsymbol{sup_{t,o}}$ uniformly generated
3. $\tilde{P}$:=*FuzzyfyFun* (coefficients, r)
4. Calculate the measures of possibility of failure given $\tilde{P}$ and P_{fi}
5. Get the sum of the failure possibility measures.

- *LinearizeFun* is the procedure which implements the linearization operator of the circuit performance responses (see equation 8).
- *LatinizeFun* computes latin hyper-cube sampling with a number of points twice as parameter dimensions.
- *FuzzyfyFun* computes the fuzzyfied circuit performance $\hat{P}$ by equation 2.
- *CircuitFun* is the function to interface the circuit simulator (see equation 7).

6 Results

The test carried out is aimed to guarantee the specification performances with an optimal choice of the geometrical parameters taking into account the uncertainty of the uncontrollable and operational parameters. An optimization performed with the Possibilistic Worst-Case Distance (PWCD) (see equation 5) on the specification given in table 2.

Table 2. Performance failures for the tested circuit

Performance P	***Failure P_f***	***Unit***
Gain @ 100 Hz	< 60	dB
Phase margin	< 60	Degree
Unity gain frequency	< 20	MHz
Power dissipation	> 0.67	mW

The function to optimize was

$$\sum_{i\in I} \Pi_{(P_i - P_{fi})}([0,+\infty)) \tag{9}$$

where I is the set of performance P_i to guarantee and P_{fi} is the specification failure. The sum of possibilities is the L_1 norm inside the space of the problem objectives possibilities. This choice allows the characterization of convex regions of the multi-objective problem with a suitable merit function. The circuit simulator used is Spice in the implementation named *ngspice* [13] with BiSim3 MOSFET model card.

The graphic in figure 2 shows the comparison between Possibilistic Worst-Case Distance methodology and the most common design methodology used in microelectronic industry named “Nominal Over-Design”. It is shown the comparison of the two methodologies in terms of the resultant yield. The Nominal Over-Design methodology fixes every objective to a secure value with reference to the nominal specifications. In this test case the objectives were increased of a 10% with regard to the minimum thresholds and decreased of a 10% with regard to the maximum thresholds. Both methodologies used the Nelder-Mead simplex optimization algorithm with Simulated Annealing [14]. The scheduling of annealing temperature allows to coordinate the convergence of optimization variables making global the searching procedure.

A circuit is classified as “acceptable” if every performance specification is satisfied. In the microelectronic industry context, the “yield” is the ratio between acceptable

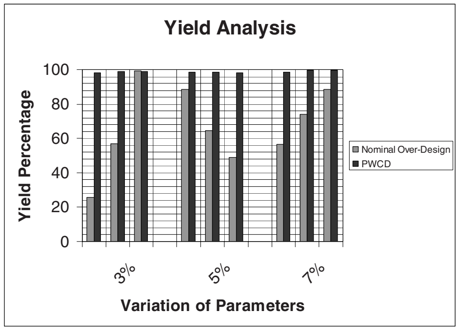


Fig. 2. Yield results of Nominal Over-Design and Possibilistic Worst-Case Distance methodologies carried out by means of Montecarlo simulations

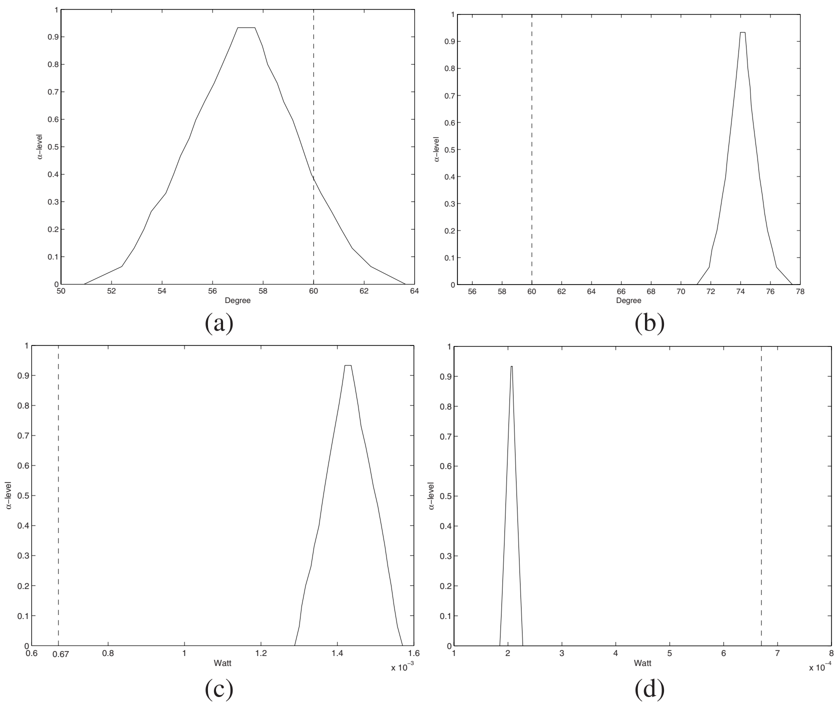


Fig. 3. (a), (b) represent the fuzzy number Phase margin (solid line) with respect to the failure value of 60 *Degree* (dashed line) before (a) and after (b) optimization methodology. (c), (d) represent the fuzzy number Power dissipation (solid line) with respect to the failure value of 0.67 *mW* (dashed line) before (c) and after (d) optimization methodology.

circuits over the whole production of circuits [5]. The yield value was computed by means of Montecarlo simulations. Three independent tests were carried out considering a statistical distribution of the technological parameters of 3%, 5% and 7% with respect to the nominal default value of the model card [11]. Every test executed 2000 simulations.

The graphic in figure 2 shows that Nominal Over-Design methodology (light grey) involves a large variability of yield values. Therefore, this methodology, widely utilized, gives no guarantee of robustness. On the contrary, PWCD methodology (dark grey) shows high values of yield and therefore robust yield results in relation to the considered uncertainty.

The graphics in figure 3 show the fuzzy numbers of the circuit performances *Phase margin* and *Power dissipation* as against the failure specification shown in table 2. Graphics on the left side show these fuzzy numbers before optimization on initialization step of the simplex optimization algorithm, while graphics on the right side show fuzzy numbers of the circuit performances after the optimization process.

This typology of electronic design takes into account the trade-off between different objectives during searching of the optimal design. In general, given a starting configuration, only few constraints are satisfied. The optimization must find a new configuration to satisfy all constraints simultaneously with a safety margins. The effectiveness of the methodology is shown by the graphics on the right side which point out the constraint satisfaction. Graphics in figure 4 show the fuzzy numbers representing the performances *Unity Gain Frequency* and *Gain @ 100 Hz* at different technological parameter variations of 5%, 7% and 9%. Notice that the increasing of parameter uncertainties changes the value of the performances and their relative uncertainties.

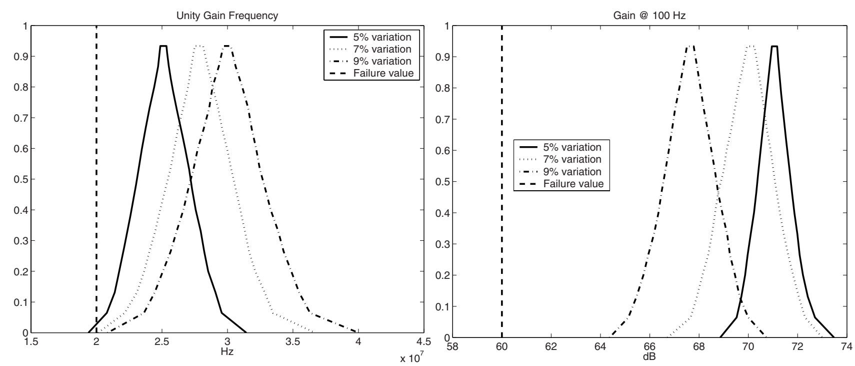


Fig. 4. Fuzzy numbers representing the performances Unity Gain Frequency and Gain @ 100 Hz at different technological parameters variations (5%, 7%, 9%)

7 Conclusions

The Possibilistic Worst-Case Distance optimization methodology made use of concepts from fuzzy set and possibilistic theory to model uncertainty of circuit parameters in order to evaluate and to minimize the Worst-Case Distance. Briefly, this methodology showed the following advantages:

- the uncertainty arising from circuit design has been modelled by means of a methodology that avoid statistical hypothesis or sensitive analysis;
- the new optimization methodology has a good behaviour over a wide range of process and design conditions;

- PWCD features are suitable to be integrated into optimization flows. The problem specification of worst-case distance is in accordance with common circuit design problem specifications;
- the methodology made use only of the circuit simulation scheme without any other analysis tools of analog circuits;
- the methodology provides a real approach to reduce the computational effort of a design based on fuzzy set to model uncertainty.

References

1. Venter, G., Haftka, R.(1999): Using response surface approximations in fuzzy set based design optimization. Structural and Multidisciplinary Optimization 18(4) (December) 218 - 227
2. Hjamlmarson, E., (2003), Studies on design automation of analog circuits - the design flow. PhD thesis, Department of Electrical Engineering. Linlköpings universitet, Linlköpings, Sweden
3. Zadeh, L.A. (1968).: Fuzzy algorithms. Information and Control 12(2) 94 - 102
4. . Antreich, K., Graeb, H., Wieser, C. (1993): Practical methods for worst-case and yield analysis of analog integrated circuits. International Journal of High Speed Electronics and Systems 4(3) 261 - 282
5. Graeb, H., Wieser, C., Antreich, K. (1993): Improved methods for worst-case analysis and optimization incorporating operating tolerances. 142 - 147
6. Abdel-Malek, H., Bandler, J. (1980): Yield optimization for arbitrary statistical distributions: Part i - theory, part ii - implementation. IEEE Transactions on Circuit and System CAS - 27(4)
7. Zadeh, L. (1978): Fuzzy sets as the basis for a theory of possibility. Fuzzy Sets and Systems 1 3 - 28
8. Dubois, D., Prade, H.: Possibility theory: An approach to computerized processing of uncertainty. New York: Plenum Press (1988)
9. Spinella, S., Anile, A.: Modeling uncertain sparse data with fuzzy b-splines. Reliable Computing 10(5) (October 2004) 335 - 355
10. Zhao, W., Cao, Y. (2006): New generation of predictive technology model for sub-45nm design exploration. isqed 0 585 - 590
11. Cao, Y., Sato, T., Sylvester, D., Orshansky, M., Hu, C. (2000): New paradigm of predictive mosfet and interconnect modeling for early circuit design. Proc. of IEEE CICC
12. Medeiro-Hidalgo, F., Dominguez-Castro, R., Rodríguez-Vázquez, A., Huertas, J.L. (1992): A prototype tool for optimum analog sizing using simulated annealing. Proc. IEEE Int. Symp. Circuits Syst. (May) 1933 - 1936
13. NGSPICE. Web site of ngspice at http://ngspice.sourceforge.net/
14. Nelder, J., Mead, R.: A simplex method for function minimization. Computer Journal 7 (1965) 308 – 313

An Algorithm to Solve Two-Person Non-zero Sum Fuzzy Games

Wanessa Amaral and Fernando Gomide

Department of Computer Engineering and Automation, Faculty of Electrical and Computer Engineering, State University of Campinas, 13083-970 Campinas, SP, Brazil
wmamaral@gmail.com, gomide@dca.fee.unicamp.br

Abstract. The paper presents an algorithm to solve two-person nonzero-sum fuzzy games using decomposition of a bilinear programming model into a series of linear programming models. Despite considerable research and advances in the area, most computational algorithms developed require solution of nonlinear optimization models. The approach discussed here provides a mechanism to translate the original fuzzy model into a family of its α-cut equivalents and a decomposition scheme to split them in linear models. A simple example is given to illustrate the algorithm.

Keywords: Fuzzy games, nonzero sum games, game theory, bilinear algorithm.

1 Introduction

This paper concerns non-cooperative two player nonzero-sum games with fuzzy payoffs. Non-cooperative games assume that there is no communication between players or, if there is, players do not agree on bidding strategies and act rationally. Game theory research plays an important role in decision making theory and in many practical situations especially in economics, mechanism design and market analysis, multi-agent systems, deregulated energy markets, biology and evolutionary modeling and computation, to mention a few.

In real situations, however, it is difficult to know the values of payoffs exactly. Recently, much attention has been given on game problems with fuzzy payoff matrices to approach both, zero and non-zero sum games.

In parallel to theoretical developments, methods for solving fuzzy game problems have also been proposed. For instance, Campos (1989) introduces five different ranking functions to handle fuzzy payoff matrices and develops linear programming models to solve zero sum games. Nishizaki and Sakawa (1997) suggest an approach to solve bi-matrix games with fuzzy payoffs in the form of L-R fuzzy numbers. The α-cut approach is used together with possibility and necessity measures to define a solution concept, the α-possible equilibrium. Lee-Kwang and Lee (1999) suggest a method to rank fuzzy numbers using satisfaction functions and illustrate the use of the method to solve fuzzy decision-making problems and games with profit and loss. Maeda (2003) shows that equilibrium strategies of two-person zero-sum fuzzy games can be characterized through (α,β)-Nash equilibrium strategies of a family of parametric bi-matrix

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 296–302, 2007.
springerlink.com

games with crisp payoffs. Chen and Larbani (2006) formulate multiple attribute decision-making problems as two-person zero-sum games and develop a method to solve fuzzy two-person zero-sum games using α-cuts instead of ranking.

In this paper we suggest a α-cut based algorithm to solve non-cooperative, two-person nonzero sum fuzzy game problems. The algorithm explores the bilinear nature of an associated optimization model and a decomposition scheme to translate the original bilinear model into a set of α-parameterized linear optimization models.

The paper proceeds as follows. After this introduction, next section briefly reviews basic notions of fuzzy games and section 3 details the algorithm and its characteristics. Section 4 gives a simple illustrative example and Section 5 concludes the paper suggesting issues for further consideration.

2 Fuzzy Sets and Games

In this section basic notions on fuzzy games are briefly reviewed. A comprehensive treatment of single and multiobjective fuzzy games is found in (Nishizaki and Sakawa, 2001).

A fuzzy set $\tilde{a}$ is defined by a membership function mapping the elements of a universe U in the unit interval [0, 1]:

$$\tilde{a} : U \rightarrow [0,1] \tag{1}$$

A fuzzy number is a fuzzy set $\tilde{a}$ on the set of real numbers such that (Nguyen & Walker, 1999),

1. There exists a unique real number x such that $\tilde{a}(x) = 1$;
2. $\tilde{a}_\alpha$ must be a closed interval for every $\alpha \in (0,1]$;
3. The support of $\tilde{a}$ must be bounded,

where $\tilde{a}_\alpha = \{x \mid A(x) \geq \alpha\}$ is the α-cut of $\tilde{a}$.

This paper consider games with fuzzy payoff matrices $\tilde{A}$ of the form

$$\tilde{A} = \begin{bmatrix} \tilde{a}_{11} & \cdots & \tilde{a}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{a}_{m1} & \cdots & \tilde{a}_{mn} \end{bmatrix} \tag{2}$$

where $\tilde{a}_{ij}$, i=1,...,m, j = 1,...,n are triangular fuzzy numbers of the form $\tilde{a}_{ij} = (l, m, u)$ as elements. In what follows we assume the reader to be familiar with the notion of payoff matrices for two person zero sum games.

Nonzero-Sum Games

Two payoff matrices $\tilde{A}$ and $\tilde{B}$, one for each player, characterize two-person non zero-sum games.

$$\tilde{A} = \begin{bmatrix} \tilde{a}_{11} & \cdots & \tilde{a}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{a}_{m1} & \cdots & \tilde{a}_{mn} \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} \tilde{b}_{11} & \cdots & \tilde{b}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{b}_{m1} & \cdots & \tilde{b}_{mn} \end{bmatrix} \tag{3}$$

where $\tilde{a}_{ij}$, $\tilde{b}_{ij}$ are payoffs that players I and II receive when player I chooses the pure strategy i and player II chooses pure strategy j.

Nash equilibrium solutions for non zero sum game are pairs of m-dimensional vectors x^* and n-dimensional vectors y^* such that (Maeda, 2003):

$$\begin{aligned} x^*\tilde{A}y^* &\geq x\tilde{A}y^* \quad \forall x \in X \\ x^*\tilde{B}y^* &\geq x^*\tilde{B}y \quad \forall y \in Y \end{aligned} \tag{4}$$

where x^* is an element of X the set of all mixed strategies available for player I

$$X = \left\{ x = (x_1, \ldots, x_m) \in R^m \mid \sum_{i=1}^{m} x_i = 1, x_i \geq 0, i = 1, \ldots, m \right\} \tag{5}$$

and y* is an element of Y, the set of all mixed strategies available for player II

$$Y = \left\{ y = (y_1, \ldots, y_n) \in R^n \mid \sum_{i=1}^{n} y_i = 1, y_i \geq 0, i = 1, \ldots, n \right\} \tag{6}$$

A point $(x^*\tilde{A}y^*, x^*\tilde{B}y^*)$ is the value of the game.

A point $(x^*, y^*) \in X \times Y$ is a fuzzy Nash equilibrium strategy if and only if x^* is an optimal solution to the following fuzzy linear programming problem with parameter y^*

$$\begin{aligned} \max_{x} \quad & x\tilde{A}y^* \\ st: \quad & x \in X \end{aligned} \tag{7}$$

and y^* is an optimal solution for the following fuzzy linear programming problem with parameter x^*.

$$\begin{aligned} \max_{y} \quad & x^*\tilde{B}y \\ st: \quad & y \in y \end{aligned} \tag{8}$$

Therefore, we note from (7) and (8) that to compute fuzzy Nash equilibrium strategies we must solve the following fuzzy bilinear programming problem

$$\begin{aligned} \max_{x,y} \quad & x\tilde{A}y^* + x^*\tilde{B}y \\ st: \quad & x \in X \\ & y \in Y \end{aligned} \tag{9}$$

Clearly, zero-sum fuzzy games are particular instances of (9). Therefore, algorithms to solve (9) also serve to solve zero-sum games as well.

3 Bilinear Algorithm

The fuzzy bilinear optimization problem (9) cannot be solved using the conventional optimization methods. From a result of we notice that problem (9) is equivalent to the following bilinear fuzzy optimization problem:

$$\begin{aligned}
\max_{x,y,p,q} \quad & x\tilde{A}y + x\tilde{B}y - p - q \\
st: & \tilde{A}y \le pe^m \\
& \tilde{B}'x \le qe^n \\
& \textstyle\sum_{i=1}^m x_i = 1 \\
& \textstyle\sum_{j=1}^n y_j = 1 \\
& x_i \ge 0, i = 1,\dots,m. \\
& y_j \ge 0, \ j = 1,\dots,n.
\end{aligned} \tag{10}$$

where e^m and e^n are m and n-dimensional vectors whose elements are all ones. Therefore, and idea is to handle fuzziness using α-cuts first, and next solve a family of α-parameterized bilinear optimization problems to find equilibrium solutions. This approach has been discussed in the literature in the realm of zero-sum games (Chen & Larbani, 2006). In this paper a similar α-cut approach is used to translate the fuzzy bilinear problem (10) in a set of α-parameterized conventional bilinear problems.

The α-cut of a fuzzy number $\tilde{a}$ can be represented by the interval $[\tilde{a}^L_\alpha, \tilde{a}^U_\alpha]$, where $\tilde{a}^L_\alpha = \inf(\tilde{a})_\alpha$ and $\tilde{a}^U_\alpha = \sup(\tilde{a})_\alpha$, respectively, the lower and the upper bounds of $\tilde{a}_\alpha$. Similarly, we can represent the α-cut of a fuzzy matrix $\tilde{A}_{m\times n}$ by $\tilde{A}_\alpha = (A^L_\alpha, A^U_\alpha)$, where $A^L_\alpha = [a_{ij}{}^L{}_\alpha]_{mxn}$, and $A^R_\alpha = [a_{ij}{}^U{}_\alpha]_{mxn}$ are real valued matrices composed by the lower and upper bounds of the respective elements of matrix $\tilde{A}_\alpha$. Therefore, the bilinear model (10) becomes

$$\begin{aligned}
\max \quad & x\,[A^L_\alpha, A^U_\alpha]\,y + x\,[B^L_\alpha, B^U_\alpha]\,y - p - q \\
st: & [A^L_\alpha, A^U_\alpha]\,y \le pe^m \\
& [B^L_\alpha, B^U_\alpha]'\,x \le qe^n \\
& \textstyle\sum_{i=1}^m x_i = 1 \\
& \textstyle\sum_{j=1}^n y_j = 1 \\
& x_i \ge 0, i = 1,\dots,m. \\
& y_j \ge 0, \ j = 1,\dots,n.
\end{aligned} \tag{11}$$

Given values for $\alpha \in [0,1]$, equilibrium solutions (x^*, y^*) can be found solving (11) for each α. Because (10) is a fuzzy two-person non-zero-sum game in which both players want maximize their profit, the task of player I is to solve the bilinear optimization problem (12) and, similarly, the task of player II is to solve the bilinear problem (13):

$$\begin{aligned} \max \quad & x[A_\alpha^L + B_\alpha^L]\,y - p \\ st: & A_\alpha^L\; y \le pe^m \\ & \textstyle\sum_{j=1}^n y_j = 1 \\ & y_j \ge 0,\; j = 1,\dots,n. \end{aligned} \quad \text{and} \quad \begin{aligned} \max \quad & x[A_\alpha^U + B_\alpha^U]\,y - p \\ st: & A_\alpha^U\; y \le pe^m \\ & \textstyle\sum_{j=1}^n y_j = 1 \\ & y_j \ge 0,\; j = 1,\dots,n. \end{aligned} \tag{12}$$

$$\begin{aligned} \max \quad & x[A_\alpha^L + B_\alpha^L]\,y - q \\ st: & B_\alpha^{L\,\prime}\, x \le qe^n \\ & \textstyle\sum_{i=1}^m x_i = 1 \\ & x_i \ge 0,\; i = 1,\dots,m. \end{aligned} \quad \text{and} \quad \begin{aligned} \max \quad & x[A_\alpha^U + B_\alpha^U]\,y - q \\ st: & B_\alpha^{U\,\prime}\, x \le qe^n \\ & \textstyle\sum_{i=1}^m x_i = 1 \\ & x_i \ge 0,\; i = 1,\dots,m. \end{aligned} \tag{13}$$

To solve the bilinear models (12) and (13) we use a decomposition algorithm suggested in (Bazaraa, 1979).

Consider the bilinear problem $\min \phi(x, y) = c^t x + d^t y + x^t Hy$, subject to $x \in X$ and $y \in Y$, where X and Y are bounded polyhedral sets in R^m and R^n, respectively. The algorithm is as follows:

Initialization Step: Select an $x_1 \in R^m$ and $y_1 \in R^n$. Let k=1 and go to the main step:

Main Step:

1 - Solve the linear program $\min\; d^t y + x_k^t Hy$ subject to $y \in Y$. Let $\hat{y}$ be an optimal solution. Let y_{k+1} be as specified bellow and go to step 2:

$$y_{k+1} = \begin{cases} y_k \text{ if } \phi(x_k, \hat{y}) = \phi(x_k, y_k) \\ \hat{y} \text{ if } \phi(x_k, \hat{y}) < \phi(x_k, y_k) \end{cases} \tag{14}$$

2 - Solve the linear program $\min\; c^t x + x^t Hy_{k+1}$ subject to $x \in X$. Let $\hat{x}$ be an optimal solution. Let x_{k+1} be as specified bellow and go to step 3:

$$x_{k+1} = \begin{cases} x_k \text{ if } \phi(\hat{x}, y_{k+1}) = \phi(x_k, y_{k+1}) \\ \hat{x} \text{ if } \phi(\hat{x}, y_{k+1}) < \phi(x_k, y_{k+1}) \end{cases} \tag{15}$$

3 - If $x_{k+1} = x_k$ and $y_{k+1} = y_k$: stop, with (x_k, y_k) as a Kuhn-Tucker point. Otherwise, replace k by k + 1, and go to step 1.

The algorithm starts with x and y chosen randomly and decompose the bilinear program problem into two coupled linear optimization problems. The basic idea is to solve iteratively the two linear programming problems until the solution does not improve further. Amaral (2006) shows that this algorithm converges to a Karush-Kuhn-Tucker point.

4 Numerical Example

Consider the following fuzzy payoff matrices

$$\tilde{A} = \begin{bmatrix} (0,1,2) & (-1,0,1) \\ (0.5,2,3.5) & (-2,-1,0) \end{bmatrix} \quad \text{and} \quad \tilde{B} = \begin{bmatrix} (2,3,4) & (1,2,3) \\ (-1,0,1) & (-0.5,1,2.5) \end{bmatrix} \tag{16}$$

The equilibrium solution when we solve the problems (12) and (13) using the bilinear algorithm for selected values of $\alpha \in [0,1]$ is shown in Figure 1. The value of the game can be interpreted as around 0.5 for player I, and around 1.5 for player II.

It is interesting to notice that payoff matrices (16) are fuzzified instances of a conventional nonzero sum game example addressed in (Basar, 1982). The value of the game for $\alpha = 1$ is the same as the one obtained by the conventional game.

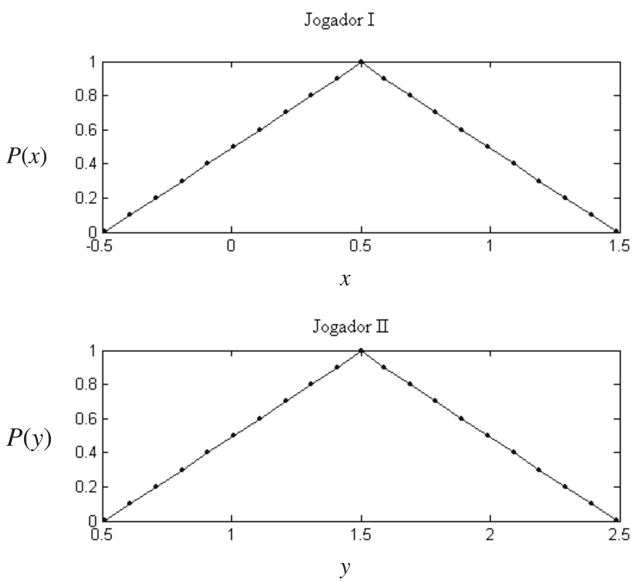


Fig. 1. Fuzzy value of the game

4 Conclusion

In this paper, we suggested a bilinear algorithm to find equilibrium solutions of non-cooperative two-person non-zero-sum games with fuzzy payoff matrices and mixed strategies.

The approach combines the α-cut idea and a decomposition algorithm to solve a bilinear programming problem using a set of linear programming sub-problems. Further experiments have shown that the algorithm performs well and often converges to an equilibrium solution.

Further work shall address characterization and refinements of the solutions produce by the algorithm once we can only show that it converges to Karush-Kuhn-Tucker points and thus, in principle, to local solutions. There are, however, hope that global solutions can be found adding a step to search the extreme points of the Cartesian product of X and Y to find global solutions.

References

1. Basar, T. e Olsder, G. J. (1982). *Dynamic Noncooperative Game Theory.*,Academic Press, New York, USA.
2. Bazaraa, M. S. and Shetty, C. (1979). *Nonlinear Programming Theory and Algorithms*, John Wiley and Sons, New York, USA.
3. Campos, L. (1989). Fuzzy linear programming models to solve fuzzy matrix games, *Fuzzy Sets and Systems*, 32:275-289.
4. Chen, Yuh-Wen e Larbani, Moussa (2006). Two-person zero-sum game approach for fuzzy multiple attribute decision making problems, *Fuzzy Sets and Systems*, 157:34–51.
5. Lee-Kwang, H. and Lee, J. (1999). A Method for Ranking Fuzzy Numbers and Its Application to Decision-Making, *IEEE Trans. on Fuzzy Systems*, 7:677-685.
6. Luenberger, David G. (1984). *Linear and Nonlinear Programming*, 2nd Ed., Addison Wesley, MA, USA.
7. Maeda, T. (2003). On characterization of equilibrium strategy of two-person zero-sum games with fuzzy payoffs, *Fuzzy Sets and Systems*, 139:283–296.
8. Nishizaki, Ichiro and Sakawa, Masatoshi. (2001). Fuzzy and Multiobjective Games for Conflict Resolution, Physica-Verlag, Heidelberg, Germany.
9. Nguyen, H.T. and Walker, E.A. (1999), *A First Course in Fuzzy Logic*, CRC Press, USA.
10. Pedrycz, W. and Gomide, F. (1998). *An introduction to fuzzy sets: Analysis and Design*, MIT Press, Cambridge, USA.
11. Amaral, W. (2006). Fuzzy Games and Solution of Decision-Making Problems with Conflict of Interest, Internal Report, Unicamp-FEEC-DCA.

One-Shot Decision with Possibilistic Information

Peijun Guo

Faculty of Economics, Kagawa University, Takamatsu, Kagawa 760-8523, Japan
guo@ec.kagawa-u.ac.jp

Abstract. In this paper, a new possibilistic decision approach for one-shot decision problem is proposed where two focus points, called active focus point and passive focus point are introduced for balancing satisfaction and plausibility to show which state of nature should be considered for making decision with possibilistic information. Based on the proposed one-shot decision approach, real estate investment problem is analyzed, that is, whether the landowner should construct a house at the present time for sale in the future considering the uncertainty of house price. Uncertainty of house price is characterized by the possibility distribution to reflect the potential of how much the house price being in the future, which is a kind of likelihood to show the similarity between the situation of housing market in the future and in the past time. The proposed model provides insights into individual investment behavior of urban land development in the real world and shows that possibilistic decision analysis based on active and passive focus points is reasonable for such one-shot decision problems, which extensively exist in business and economic society.

Keywords: possibility theory, active and passive focus points, real estate investment.

1 Introduction

Decision analysis can be used to determine an optimal strategy when a decision maker is faced with two or more decision alternatives. In fact, for a selected decision alternative, there are many possible outcomes happening. Such situations can be described by two distinct words, risk and uncertainty [10]. The situations that only involve risk are that the probability of all possible events occurring in the future could be calculated because the events occur with sufficient frequency. On the contrary, uncertainty is associated with partially known information so that an exact probability distribution cannot be obtained. A simple example has been given by Yager [18]. Possibility theory initiated by Zadeh [19] and advanced by Dubois, Prade and Klir [1, 9] offers another formal mathematical tool to characterize and analyze the uncertainty in the real world based on two basic non-additive measures, possibility measure and necessity measure. In the same way as probabilities can be interpreted in different ways (e.g., frequentist view and subjective view), possibility can be explained from several semantic aspects. A first one is the idea of feasibility, such as ease of achievement[19]. A second one is plausibility, referring to the propensity of events to occur, which relates to the concept "potential surprise"[12]. A third one is logical consistency with available information [11, 16]. The last one is the preference, referring to the willingness of agent to make a decision [2]. Necessity is dually related to possibility in the sense that "not A" being not possible means that A is necessary. A semantic analysis

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 303–313, 2007.
springerlink.com

can be done in parallel with possibility, referring to belief, acceptance and priority. Decision analysis based on possibility theory has been researched in the papers [4, 5, 6, 7, 8, 13, 14]. In this paper, an approach for one-shot decision problems that is for situations that can be experienced only once, with partially known information characterized by possibility distributions, is proposed. Two focus points, called active focus point and passive focus point are introduced for balancing satisfaction and plausibility to show which state of nature should be considered for making decision with possibilistic information. Based on the proposed approach, a typical one-shot decision problem, real estate investment problem, is considered. The problem is that the landowner should make the decision whether he should begin to construct a house at the present time for selling at the end of construction. Any delay of sale is not permitted because of huge interest expenses. The uncertainty of price of house at the end of construction is characterized by the possibility distribution to reflect the potential of how much the building price being judged by experts based on the past samples. Two focus points, called active focus point and passive focus point are introduced for balancing satisfaction and plausibility to show which price should be considered for making decision with possibilistic information. A general decision model for real estate investment is proposed where the possibility distributions and satisfaction functions reflecting the satisfaction level of decision makers are not set as specified functions. The proposed model provides novel insights into individual investment behavior of urban land development in the real world and shows that possibilistic decision analysis based on active and passive focus points is much reasonable for such one-shot decision problems, which extensively exist in business and economic society.

2 Possibilistic One-Shot Decision Approach with Active and Passive Focus Points

The first step in the decision analysis for one-shot decision problem is problem formulation. The set of a decision alternative a is denoted as $A=\{a\}$. The uncertain future event is referred to as chance event and its outcomes are referred to as the states of nature. Denoting a state of nature as x, the set of the states of nature is $S=\{x\}$. The consequence resulting from a specific combination of a decision alternative a and a state of nature x is refereed to as a payoff, denoted as $p(x,a)$. The degree of which state of nature being more possible to happen in the future can be characterized by the possibility distribution, defined as follows.

Definition 1. Given a continuous function

$$\pi : S \rightarrow [0,1] \tag{1}$$

$$\text{if } \max_{x\in S} \pi(x) = 1, \tag{2}$$

then the function $\pi(x)$ is called a possibility distribution and the value of $\pi(x)$ is called the possibility degree of x. $\pi(x)=1$ means that it is normal that the state of nature x occurs and $\pi(x)=0$ means that it is abnormal that x occurs. The smaller the possibility degree of x, the more surprised the happening of x. The possibility distribution function can be used to represent the knowledge or judgment of experts. Several kinds of methods for identifying the possibility distribution have been proposed [5, 7, 8, 15].The possibility distribution determines the unique possibility and necessity measures of an event E, denoted as $Pos(E)$ and $Nec(E)$ via the following formulas

$$Pos(E) = \max_{x \in E \subseteq S} \pi(x), \quad (3)$$

$$Nec(E) = 1 - \max_{x \in E^c \subseteq S} \pi(x), \quad (4)$$

$$Pos(\{x\}) = \pi(x) \quad x \in S. \quad (5)$$

The normalized continuous function $u(x,a)$, called the satisfaction function, is used to express the satisfaction level of decision maker for the payoff $p(x,a)$.

In decision analysis, a decision maker must first select a decision alternative, then a state of nature follows, and finally a consequence will occur. In some cases, the selected decision alternative may provide a good or excellent result because of a favorable state of nature. In other cases, a relatively unlikely future event may occur causing the selected decision alternative to provide only fair or even poor result. It is true that there is one and only one of possible states of nature will occur. So that for one-shot decision problem, a decision maker has to decide which state of nature should be considered to balance the satisfaction and plausibility. For example, it is simple that the state of nature with possibility grade 1 is considered for each decision alternative in the sense that the decision maker only focuses the most normal state of nature. In summary, the procedure of one-shot decision analysis is as follows.

Step 1. Decide which state of nature should be focused for each decision alternative. For balancing the satisfaction and plausibility, the following two states of nature of decision alternative a, denoted as $x^*(a)$ and $x_*(a)$, should be considered:

$$x^*(a) = \arg\max_{x \in S} \min(\pi(x), u(x,a)), \quad (6)$$

$$x_*(a) = \arg\min_{x \in S} \max(1 - \pi(x), u(x,a)). \quad (7)$$

Step 2. Optimal decision is chosen based on the selected states of nature in Step 1. Based on $x^*(a)$ and $x_*(a)$, optimal decision is determined to make $u(x^*(a),a)$ or $u(x_*(a),a)$ maximize, that is,

$$a^* = \arg\max_{a \in A} u(x^*(a),a), \quad (8)$$

$$a_* = \arg\max_{a \in A} u(x_*(a),a). \quad (9)$$

Definition 2. Two states $x^*(a)$ and $x_*(a)$, which satisfy (6) and (7) are called active and passive focus points of decision a respectively, and $u(x^*(a),a)$ and $u(x_*(a),a)$ are called active value of decision a and passive value of decision a, respectively. $x^*(a^*)$ satisfying (8) and $x_*(a_*)$ satisfying (9) are called active and passive focus points of states of nature, respectively. a^* and a_* are called active and passive optimal decisions, respectively.

Proposition 1. No other $(\pi(x), u(x,a))$ can strongly dominate $(\pi(x^*(a)), u(x^*(a),a))$, that is, $\pi(x)$ is larger than $\pi(x^*(a))$ and $u(x,a)$ is larger than $u(x^*(a),a)$.

Proposition 2. For the set of states of nature $\{x \mid \pi(x) > \pi(x_*(a))\}$, the relation $u(x,a) \geq u(x_*(a),a)$ must hold.

Proposition 1 shows that it is impossible to increase possibility degree and satisfaction level from $(\pi(x^*(a)), u(x^*(a),a))$ simultaneously. Proposition 2 shows the evaluation of decision a can be improved within $\{x \mid \pi(x) > \pi(x_*(a))\}$. It can be seen that $x^*(a)$ gives more active evaluation, and $x_*(a)$ gives more passive evaluation of decision a in the sense of the existence of improvement margin for evaluation. In the following, a simple example is given to show one-shot decision analysis based on active and passive focus points.

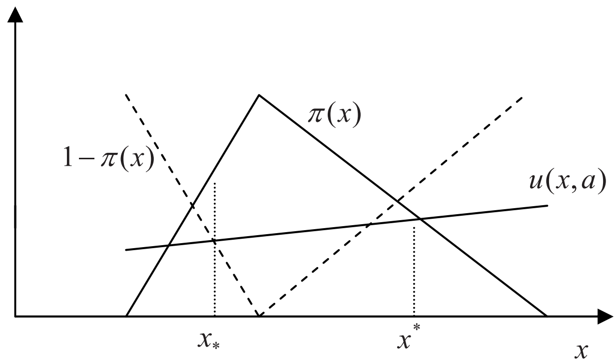
Example. Assume that you have 2,000$ now. There is a game as follows. Throwing a coin, if the outcome is tail then you win 1,000$ otherwise you lose 1,000$ for the result of head. And there is only one chance to play this game. Because the coin is fair, it is obvious that the possibility of tail or head is 1. There is one and only one result you can obtain after the play, that is 3,000$ for result tail or 1,000$ for result head. The satisfaction levels of them are 1, and 0, respectively. From (6) and (7), it is known the active focus point of play is tail, that is,

$$(\pi(tail) \wedge u(tail)) = (\pi(tail) \wedge u(tail)) \vee (\pi(head) \wedge u(head), \tag{10}$$

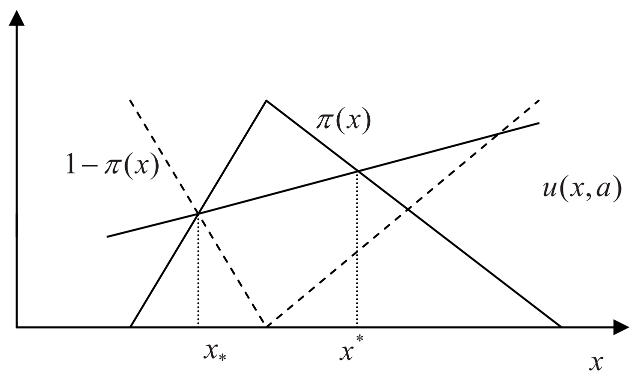
and the passive focus point of play is head, that is,

$$(1-\pi(tail)) \vee u(tail) = ((1-\pi(tail)) \vee u(tail)) \wedge ((1-\pi(head)) \vee u(head)) \tag{11}$$

The results corresponding to x^* and x_* are 3,000$ and 1,000$, respectively. It is straightforward that if the decision is made based on active focus point, this game will be play, but based on passive focus point the game will be given up. It can be easily understood from the above simple example that the possibilistic decision method based on two focus points is very suitable to deal with one-shot decision problem. The formula $\sup_{x \in S} \min(\pi(x), u(x,a))$ and $\inf_{x \in S} \max(1-\pi(x), u(x,a))$ have been initially proposed by Yager [17] and Whalen [15], respectively, axiomatized in the style of Savage by Dubois, Prade and Sabbadin [3]. These two formulas are called optimistic criterion and pessimistic criterion, respectively, because the optimistic criterion will give a higher evaluation of some decision if this decision can lead to a higher utility with a higher possibility; the pessimistic criterion will give a lower evaluation of some decision if this decision can lead to a lower utility with a higher possibility. And they can also be explained by possibility and necessity measures where the utility function is regarded as fuzzy event. However, all explanations are based on commensurability assumption between plausibility and preference, which, in fact, is not easily accepted. In the proposed possibilistic one-shot decision analysis, the focus points of decision alternative firstly are chosen where the formulas (6) and (7) are used without commensurability assumption between the possibility degree and the satisfaction level. Possibility degree and satisfaction level running on their own right are regarded as a pair of the characteristics of the state of nature for choosing the focus points of decision alternative shown by the proposition 1 and 2.



Case 1. $\pi(x^*) \le \pi(x_*)$



Case 2. $\pi(x^*) \ge \pi(x_*)$

Fig. 1. The explanation of relation between $\pi(x^*)$ and $\pi(x_*)$

Theorem 3. For an action a, it is possible that one of the following relations:

$$\pi(x^*(a)) \le \pi(x_*(a)) \tag{12}$$

$$\pi(x^*(a)) \ge \pi(x_*(a)) \tag{13}$$

holds, but the following relation does not hold.

$$u(x^*(a), a) < u(x_*(a), a) \tag{14}$$

Proof. An example has been given in Fig.1 to show that both (12) and (13) are possible to hold. Suppose that (14) holds, there are three cases, that is, Case 1: $u(x^*(a),a) < u(x_*(a),a)$ for $\pi(x^*(a)) < \pi(x_*(a))$;

Case 2: $u(x^*(a),a) < u(x_*(a),a)$ for $\pi(x^*(a)) > \pi(x_*(a))$;

Case 3: $u(x^*(a),a) < u(x_*(a),a)$ for $\pi(x^*(a)) = \pi(x_*(a))$.

Proposition 1 shows that Case 1 is impossible to exist. Proposition 2 shows that Case 2 is impossible to exist. For the case 3, because $\pi(x)$ and $u(x,a)$ are both continuous, $\pi(x^*(a)) = u(x^*(a),a)$ and $\pi(x_*(a)) = u(x_*(a),a)$ hold. Thus it can not be that $\pi(x^*(a)) = \pi(x_*(a))$ whereas $u(x^*(a),a) < u(x_*(a),a)$. □

Theorem 3 clearly shows that the active value of decision is not lower than the passive value of decision.

Theorem 4. The following equation holds:

$$\max_{a\in A} \max_{x\in S} \min(\pi(x), u(x,a)) = \max_{x\in S} \min(\pi(x), \max_{a\in A} u(x,a)), \quad (15)$$

Proof. Set $h(x,a) = \min(\pi(x), u(x,a))$. Then (15) can be rewritten to be $\max_{a\in A} \max_{x\in S} h(x,a) = \max_{x\in S} \max_{a\in A} h(x,a)$, which holds obviously. □

Theorem 5. The following inequality holds:

$$\min_{x\in S} \max(1-\pi(x), u(x,a)) \le \min_{x\in S} \max(1-\pi(x), \max_{a\in A} u(x,a)). \quad (16)$$

Proof. The relation $u(x,a) \le \max_{a\in A} u(x,a)$ makes the following inequality hold:

$$\max(1-\pi(x), u(x,a)) \le \max(1-\pi(x), \max_{a\in A} u(x,a)), (\forall a \in A \text{ and } \forall x \in S). \quad (17)$$

(17) makes the following inequality hold:

$$\min_{x\in S} \max(1-\pi(x), u(x,a)) \le \min_{x\in S} \max(1-\pi(x), \max_{a\in A} u(x,a)), (\forall a \in A). \quad (18)$$

It proves (16). □

Assume that the possibility distribution $\pi(x)$ satisfies the conditions (1) $\pi(x_l) = 0$ (2) $\pi(x_u) = 0$ (3) $\pi(x_c) = 1$ and $\pi(x)$ increases within $[x_l, x_c]$ and decreases within $[x_c, x_u]$, where x_c, x_l and x_u are the center, lower and upper bounds of x, $u(x,a)$ and $\max_{a\in A} u(x,a))$ are continuous, increasing, function of x, respectively. Then the following theorem can be obtained.

Theorem 6. $$\max_{a\in A} \min_{x\in S} \max(1-\pi(x), u(x,a)) = \min_{x\in S} \max(1-\pi(x), \max_{a\in A} u(x,a)). \quad (19)$$

Proof. $1-\pi(x)$ is a continuous, decreasing, function and $\max_{a\in A} u(x,a))$ is a continuous, strictly increasing, function for an arbitrary a within $[x_l, x_c]$. It is obvious that the following relations hold:

$$\max_{a\in A} u(x_l, a) < 1-\pi(x_l) = 1\,. \tag{20}$$

$$0 = 1-\pi(x_c) < \max_{a\in A} u(x_c, a)\,, \tag{21}$$

(20) and (21) show that there is one and only one intersection of $1-\pi(x)$ and $\max_{a\in A} u(x,a)$ within $[x_l, x_c]$. Denote the horizontal coordinate of this intersection as

$$x' \text{ then} \qquad 1-\pi(x') = \max_{a\in A} u(x', a) \tag{22}$$

$1-\pi(x)$ is a decreasing function within $[x_l, x']$, which means for any $x \in [x_l, x']$

$$\max(1-\pi(x), \max_{a\in A} u(x,a)) \geq 1-\pi(x')\,. \tag{23}$$

$\max_{a\in A} u(x,a)$ is an increasing function within $x \in [x', x_u]$, which means for any

$$x \in [x', x_u],\quad \max(1-\pi(x), \max_{a\in A} u(x,a)) \geq \max_{a\in A} u(x', a) \tag{24}$$

Considering (22) the following hold:

$$\min_{x\in S} \max(1-\pi(x), \max_{a\in A} u(x,a)) = 1-\pi(x') = \max_{a\in A} u(x', a)\,, \tag{25}$$

which means that (x', a^*) satisfies the right side of the inequality (19) where $a^* = \arg\max_{a\in A} u(x', a)$ and x' is also the horizontal coordinate of the intersection of $1-\pi(x)$ and $u(x, a^*)$. Similarly, considering that $u(x, a^*)$ is a continuous, strictly increasing, function of x, it is easy to understand that x' satisfies $\min_{x\in S} \max(1-\pi(x), u(x, a^*))$. Recalling (16), (19) can be obtained. □

Corollary 7. The following equations hold.

$$\max_{a\in A} u(x^*(a), a) = \max_{x\in S} \min(\pi(x), \max_{a\in A} u(x,a)) \tag{26}$$

$$\max_{a\in A} u(x_*(a), a) = \min_{x\in S} \max(1-\pi(x), \max_{a\in A} u(x,a))\,. \tag{27}$$

Proof. Considering (6), the following equality holds:

$$\max_{a\in A} \max_{x\in S} \min(\pi(x), u(x,a)) = \max_{a\in A} \min(\pi(x^*(a)), u(x^*(a), a))\,. \tag{28}$$

Because $\pi(x^*(a)) = u(x^*(a), a)$, we have

$$\max_{a\in A} \max_{x\in S} \min(\pi(x), u(x,a)) = \max_{a\in A} u(x^*(a), a)\,. \tag{29}$$

Considering Theorem 4, we have (26). Similarly, it is easy to prove (27). □

3 Possibilistic Decision Analysis for Real Estate Investment

House, in this paper, is characterized by its size, or the number of units, q for simplification. The cost of constructing a house on a given piece of land, C , is a strictly increasing and convex differentiable function of the number of units q ,

that is, $$dC/dq > 0 , \tag{30}$$

$$d^2C/d^2q > 0 . \tag{31}$$

The rationale for the second assumption is that as the number of floors in a building increases and the foundation of the building must be stronger. The profit r that a landowner can obtain by constructing a q -size building is as follows,

$$r = pq - C(q) , \tag{32}$$

where p is the market price per building size at the end of construction. The building size that maximizes the profit of a landowner will satisfy the following maximization problem: $R(p) = \max_q r(p,q) = \max_q pq - C(q)$. (33)

Differentiating (33) with respect to q , it follows that the solution to this maximization problem is to choose a building size, which satisfies the following condition considering the assumption (31),

$$dC(q)/dq = p . \tag{34}$$

Denote the solution of (33) as $q^{\nabla}(p)$, the maximal profit of landowner is as follows:

$$R(p) = \max_q r(p,q) = pq^{\nabla}(p) - C(q^{\nabla}(p)) \tag{35}$$

Theorem 8. $R(p)$ is a strictly increasing and convex continuous function of p .

The possibility distribution of the price p , denoted as π_P is given by the following continuous function:

$$\pi_P : [p_l, p_u] \to [0,1] , \tag{36}$$

where $\pi_P(p_l) = 0$, $\pi_P(p_u) = 0$ and $\exists p_c \in [p_l, p_u]$, $\pi_P(p_c) = 1$. π_P increases within $[p_l, p_c]$ and decreases within $[p_c, p_u]$. p_l and p_u are the lower and upper bounds of prices, respectively, p_c is the most possible price.

Theorem 9. The active focus point of price, denoted as p^* and the passive focus point of price, denoted as p_* are obtained as follows,

$$p^* = \arg \max_{p \in [p_l, p_u]} \min(\pi_P(p), u(R(q^{\nabla}(p), p))) , \tag{37}$$

$$p_* = \arg \min_{p \in [p_l, p_u]} \max(1 - \pi_P(p), u(R(q^{\nabla}(p), p))) , \tag{38}$$

where $q^{\nabla}(p^*)$ and $q^{\nabla}(p_*)$, denoted as q^* and q_*, are called active optimal building size and passive optimal building size, respectively.

Theorem 10. The active focus point of price p^* is the solution of the following equations

$$u(pq - C(q)) = \pi_P(p), \tag{39}$$

$$dC(q)/dq = p, \tag{40}$$

$$p > p_c. \tag{41}$$

Theorem 11. The passive focus point of price p_* is the solution of the following equations

$$u(pq - C(q)) = 1 - \pi_P(p), \tag{42}$$

$$dC(q)/dq = p, \tag{43}$$

$$p < p_c. \tag{44}$$

Theorem 12. The active optimal building size is larger than the passive optimal building size, that is, $q^* > q_*$.

Definition 3. Suppose that there are two possibility distributions, π_A and π_B. If for any arbitrary x $\pi_A(x) \geq \pi_B(x)$ holds, then π_B is said more informed than π_A, which is denoted as $\pi_B \rhd \pi_A$.

Theorem 13. Denote the active optimal building size based on possibility distributions A and B as q_A^* and q_B^*, respectively, the passive optimal building size based on possibility distributions A and B as q_{*A} and q_{*B}, respectively. If $\pi_B \rhd \pi_A$, then $q_A^* \geq q_B^*$ and $q_{*A} \leq q_{*B}$ hold.

Theorem 13 means that increasing the uncertainty of price can make investor increase investment scale $C(q)$ if he considers the active focus point, called as active investor and decrease investment scale if he considers the passive focus point, called as passive investor. In other words, obscure business prospect can stimulate investment of active investors and detent investment of passive investors. These kinds of results accord with the real investment world very much and are helpful for predicting the variation of constructing scale in a perfectly competitive market if knowing what is majority of investors and can provide some useful regulatory policy for urban land development.

Definition 4. Suppose that there are two landowners 1 and 2, the satisfaction functions of them are u_1 and u_2, respectively. If for any profit r, the relation $u_2(r) \geq u_1(r)$ holds, then landowner 2 is said easily satisfied than landowner 1.

Theorem 14. If landowner 2 is more easily satisfied than landowner 1, that is, $u_2(r) \geq u_1(r)$, then for the same possibility distribution of the building price in the future, the optimal building size decided by landowner 2 is always smaller than the one by landowner 1, that is, $q_2^* \leq q_1^*$ and $q_{*2} \leq q_{*1}$.

Theorem 14 means that whatever the landowner is active or passive, the more easily satisfied the investor, the smaller the investment scale. This kind of result accords with the real investment world very much.

4 Conclusions

Possibility theory based decision model is very effective especially for the cases of incomplete information, ignorance, or high cost of information acquisition because uncertainty can be simply characterized by possibility distributions to reflect the judgment of experts based on the past time samples. Decision analysis based on active and passive focus points offers a very powerful approach for one-shot decision problems, which extensively exist in business and economic society.

References

1. D. Dubois and H. Prade (1988), Possibility Theory, Plenum Press, New York.
2. D. Dubois and H. Prade, (1995) Possibility theory as a basis for qualitative decision theory, In Proceedings of the 14th International Joint Conference on Artificial Intelligence 1924-1930.
3. D. Dubois, H. Prade and R. Sabbadin, (2001) Decision-theoretic foundations of possibility theory, European Journal of Operational Research 128 459-478.
4. P. Guo and H. Tanaka, M. Inuiguchi, (2000) Self-organizing fuzzy aggregation models to rank the objects with multiple attributes, IEEE Tran. on SMC, Part A: Systems and Humans 30 573-580.
5. P. Guo, H. Tanaka and H. -J. Zimmermann, (2000) Upper and lower possibility distributions of fuzzy decision variables in upper level decision problems, Fuzzy Sets and Systems 11171-79.
6. P. Guo and H. Tanaka, (2001) Fuzzy DEA: A perceptual evaluation method, Fuzzy Sets and Systems 119149-160.
7. P.Guo, D.Zeng and H.Shishido, (2002) Group decision with inconsistent knowledge, IEEE Transactions on SMC, Part A 32 670-679.
8. P. Guo and H. Tanaka, (2003) Decision analysis based on fused double exponential possibility distributions, European Journal of Oper. Res. 148 467-479.
9. G. J. Klir, and T. Folger, (1988) Fuzzy Sets, Uncertainty and Information, Prentice-Hall, New Jersey,.
10. F. H. Knight, Risk, (1965) Uncertainty and Profit, Harper and Row, NY.
11. H. Prade, (1982) Modal semantics and fuzzy set theory. In Recent Developments in Fuzzy Set and Possibility Theory, R. R. Yager, Ed. 232-246. Pergamon Press, New York.
12. G. L. S., Shackle (1949), Expectation in Economics, Cambridge University Press, Cambridge, UK.

13. H. Tanaka and P. Guo, (1999) Possibilistic Data Analysis for Operations Research, Heidelberg; New York; Physica-Verlag,.
14. H. Tanaka and P. Guo, (1999) Portfolio selection based on upper and lower exponential possibility distributions, Euro. Journal of Oper. Res.114 115-126.
15. T. Whalen, (1984) Decision making under uncertainty with various assumptions about available information, IEEE Tran. on SMC, 14 888-900.
16. R. R. Yager, (1979) Possibilistic decision making, IEEE Transaction on Systems, Man and Cybernetics 9 388-392.
17. R. R. Yager, (1980) A foundation for a theory of possibility, J. Cyber., 10 177-204.
18. R. R. Yager, (2002) On the instantiation of possibility distributions, Fuzzy Sets and Systems 128 261-266.
19. L. A. Zadeh, (1978) Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 3-28.

Portfolio Selection Problem Based on Possibility Theory Using the Scenario Model with Ambiguous Future Returns

Takashi Hasuike and Hiroaki Ishii

Osaka University, Graduate School of Information Science and Technology,
2-1 Yamadaoka, Suita, Osaka, 565-0871, Japan
{thasuike, ishii}@ist.osaka-u.ac.jp

Abstract. In this paper, we propose the solution method about the multiobjective portfolio selection problem, particularly the scenario model to include the ambiguous factors and chance constraints. Generally, mathematical programming problems with probabilities and possibilities are called to stochastic programming problem and fuzzy programming problem, and it is difficult to find its global optimal solution. In this paper, we manage to develop the efficient solution method to find its global optimal solution of such a problem introducing the some subproblems.

Keywords: Portfolio selection problem, Chance constraint, Possibility and Necessity measure, Scenario model.

1 Introduction

When we invest possession funds on finance assets such as stocks, we predict how much profit is provided in the future and must decide the most suitable investment. Of course it is easy to decide the most suitable investment if we know future returns apriori, but they always change, and there may exist the case that an uncertainty from social conditions has a great influence on the future returns. In the real market, there may be also probabilitistic and possibilitistic factors at least from the mental point of investors. Under such probabilitistic and possibilitistic situations, we must consider how to reduce a risk, and it becomes important whether an investment makes profit greatest.

We call the issue of such finance assets selection as a portfolio selection problem, and various studies have been done till now. As for traditional mathematical approach, Markowitz [13] has proposed the mean-variance analysis model, and it is often used in the real financial field (Luenberger [12], Campbell.et al. [2], Elton and Gruber [3], Jorion [5], Steinbach [15]).

But the mean-variance model is formulated as a quadratic programming problem to minimize the variances, and the use of large-scale mean-variance models are restricted to the stock portfolio selection in spite of the recent development on computational and modeling technologies in financial engineering. Therefore, Konno [8,9] and his research group have proposed the mean-absolute derivation model which results in linear programming problem and can be solved faster than a corresponding

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 314–323, 2007.
springerlink.com

mean-variance model. They also have proposed the compact factorization of covariance matrices using the historical dates or the scenarios, and we efficiently solve portfolio selection problems by using this method.

In our previous researches, the returns in the scenario model are considered as fixed values. But when we consider the real world, even if we know all the historical dates, it is difficult to predict the future return as the fixed value, so we need to consider that the future return has the ambiguous. Until now, there are some studies of portfolio selection problems considering ambiguous situations (Leon, et al. [10], Tanaka, Guo, et al. [16,17], Watada [19]). However, there are few studies considering scenario model to portfolio selection problems including fuzzy future returns. Therefore we propose scenario models of portfolio selection problems whose future returns are assumed as L-R fuzzy numbers.When we formulate the portfolio selection problem with randomness and fuzziness, there are several mathematical models other than the mean-variance model and mean-absolute derivation model; semi-variance model (Bawa and Lindenberg [1]), safety first model (Rockafellar [14]), etc. In this paper, we consider mathematical models including randomness and fuzziness based on chance-constraints problems, and find its global optimal solution.These mathematical programming problems with probabilities and possibilities are called to stochastic programming problem and fuzzy programming problem, and there are many basic researches considering them (Katagiri, Ishii, Sakawa [6,7], Vajda [18], Inuiguchi and Ramik [4], Liu [11]). They are usually transformed into nonlinear programming problems and it is difficult to find the global optimal solution of them. In this paper, we also manage to develop the efficient solution method to find a global optimal solution of deterministic equivalent problem with more difficulty constraints

In Section 2, we introduce one of the portfolio selection problems using the scenario, Mean-Variance model. In Section 3, we extend this model of portfolio selection problem for introducing the fuzzy numbers to each return and develop the solution method of this multiobjective mean-variance model. Finally in Section 4, we conclude this paper and discuss future researches.

2 Formulation of Scenario Portfolio Selection Problem

First we introduce one of the traditional mathematical approaches to the portfolio selection problem, Markowitz model [13]. He has proposed the following mathematical programming problem as the portfolio selection problem.

$$\begin{aligned} \text{Minimize} \quad & \mathbf{xVx} \\ \text{subject to} \quad & \bar{\mathbf{r}}\mathbf{x} = r \\ & \mathbf{Ax} = b, \quad x \geq 0 \end{aligned} \tag{1}$$

where $\mathbf{x}$ is n-dimensional decision variable column vector for the decision maker, A is $m \times n$ coefficient matrix, $\bar{\mathbf{r}}$ is assumed to be the mean of n-dimensional Gaussian random variable row vectors and covariance matrix $\mathbf{V}$ is also assumed to be Gaussian random variables $N\left(r_i, \sigma_i^2\right)$. This formulation has long served as the basis of financial theory. This problem is the quadratic programming problem, so we find the optimal solution. However, it is not easy to observe variances of each asset in real market accurately and determine them as the fixed values. Furthermore, when we expect the future return of each product, we don't consider the only one scenario of the future return, but often several scenarios. In this regard, the portfolio selection problem using many scenarios of the future returns has been proposed [8,9]. Now, let r_{ij} be the realization of a random variable about the asset $j,(j=1,\cdots,n)$ of scenario $s,(s=1,\cdots,S)$, which we assume to be available from historical data and from predictions of decision makers. Then the return vector of scenario t is as follows,

$$\mathbf{r}_s = (r_{s1}, r_{s2}, \cdots, r_{sn}),\ s=1,\cdots S \tag{2}$$

where n is the total number of assets. Here, we introduce the probabilities p_s, $(s=1,\cdots,T)$ to occur in each scenario. We also assume that the expected value of the random variable can be approximated by the average derived from these data. Particularly, we let

$$r_j \equiv E[R_j] = \sum_{s=1}^{S} p_s r_{sj} \tag{3}$$

Then the mean $\overline{r_p}$ and variance σ_p^2 derived from the data are as follows.

$$\overline{r_p} = E\left[\sum_{j=1}^{n} R_j x_j\right] = \sum_{j=1}^{n} r_j x_j = \sum_{s=1}^{S} p_s \left(\sum_{j=1}^{n} r_{sj} x_j\right) \tag{4}$$

$$\sigma_p^2 = \sum_{s=1}^{S} p_s \left(r_s - \overline{r_p}\right)^2 \left(= \sum_{s=1}^{S} p_s \left(\left(\sum_{j=1}^{n} r_{sj} x_j\right) - \sum_{s=1}^{S} p_s \left(\sum_{j=1}^{n} r_{sj} x_j\right)\right)^2 \right) = \sum_{j=1}^{n} \sigma_{jk} x_j x_k \tag{5}$$

Therefore, we transformed Markovitz model into the following problem.

$$\begin{array}{ll} \text{Minimize} & \sum_{s=1}^{S} p_s \left(\sum_{j=1}^{n} (r_{sj} - r_j) x_j\right)^2 \\ \text{subject to} & \sum_{j=1}^{n} r_j x_j \geq \rho,\ \ r_j = \sum_{s=1}^{T} p_s r_{sj}, \\ & \sum_{j=1}^{n} x_j = 1,\ \ 0 \leq x_j \leq b_j,\ \ j=1,\cdots,n \end{array} \tag{6}$$

And introducing $z_s = \sum_{j=1}^{n} (r_{sj} - r_j) x_j$, we also transformed problem (6) into the following problem.

$$\begin{array}{ll} \text{Minimize} & \sum_{s=1}^{S} p_s z_s^2 \\ \text{subject to} & z_s - \sum_{j=1}^{n} (r_{sj} - r_j) x_j = 0,\ \ r_j = \sum_{s=1}^{S} p_s r_{sj} \\ & \sum_{j=1}^{n} r_j x_j \geq \rho,\ \ \sum_{j=1}^{n} x_j = 1,\ \ 0 \leq x_j \leq b_j,\ \ j=1,\cdots,n,\ \ s=1,\cdots,S \end{array} \tag{7}$$

This problem is a quadratic programming problem not to include the variances, so we solve the problem more easily than Markowitz model.

3 Fuzzy Extension of Scenario Portfolio Selection Problem and Solution Method Using the Possibility Theory

In previous researches using scenario models of future returns, the each return of scenarios is considered as the fixed value derived from a random variable. But considering the psychology of decision makers and the uncertainty of given information, it is difficult to predict the future return as the fixed value, so we need to consider that the future return has the ambiguous. Therefore we propose the portfolio selection problem using the scenarios with the ambiguous to future returns.

3.1 Introduction of Membership Function to Each Fuzzy Variable and Fuzzy Extension of Portfolio Selection Problem

In this paper, the return including the ambiguous is assumed to the following L-R fuzzy number.

$$\mu_{\tilde{r}_{sj}}(\omega)=\left(\bar{r}_{sj},\alpha_j,\beta_j\right)_{LR}=\begin{cases} L\left(\dfrac{\bar{r}_{sj}-\omega}{\alpha_j}\right), & \left(\omega\le\bar{r}_{sj}\right)\\ R\left(\dfrac{\omega-\bar{r}_{sj}}{\beta_j}\right), & \left(\omega\ge\bar{r}_{sj}\right)\end{cases},\quad j=1,\cdots,n,\quad s=1,\cdots,S \tag{8}$$

where $L(x)$ and $R(x)$ are nonincreasing and nonnegative functions on $[0,\infty)$ satisfying $L(0)=R(0)=1$ and $L(1)=R(1)=0$. Here, for the clarification of the following discussion, we introduce the following α-cut of the membership function.

$$\mu_{\tilde{r}_{sj}}(\omega)=\left[\mu^{(1)}_{\tilde{r}_{sj}}(\alpha),\ \mu^{(2)}_{\tilde{r}_{sj}}(\alpha)\right]=\left[\bar{r}_{sj}-L^*(\alpha)\alpha_j,\ \bar{r}_{sj}+R^*(\alpha)\beta_j\right],\quad j=1,\cdots,n,\quad s=1,\cdots,S \tag{9}$$

where $L^*(\alpha)$ and $R^*(\alpha)$ are pseudo inverse functions of $L(x)$ and $R(x)$ respectively. Furthermore, since each $\tilde{r}_{sj},\left(j=1,\cdots n;\ s=1,\cdots,S\right)$ is the expected return and we usually assume it to a positive value, we assumed a target level α satisfying the following inequality.

$$\bar{r}_{sj}-L^*(\alpha)\alpha_j\ge 0,\quad j=1,\cdots,n,\quad s=1,\cdots,S \tag{10}$$

In this paper, we consider the problem to minimize the target revel f to be less than each objective value of portfolio selection problem in scenarios of future returns. The problem (7) is formulated as the following mathematical programming problem.

$$\begin{aligned}&\text{Minimize} \quad f\\&\text{subject to} \quad \Pr\left\{\sum_{s\in A} p_s\right\} \ge \beta,\ A=\left\{k\left|\left(\sum_{j=1}^{n}\left(\tilde{r}_{kj}-r_j\right)x_j\right)^2 \le f\right.\right\},\\&\qquad \tilde{r}_j=\sum_{s\in A} p_s\tilde{r}_{sj},\ \sum_{j=1}^{n}\tilde{r}_j x_j \ge \rho,\quad \sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le b_j,\quad j=1,\cdots n\end{aligned} \tag{11}$$

This problem includes L-R fuzzy variables $\tilde{r}_{sj},\left(j=1,\cdots n;\ s=1,\cdots,S\right)$, therefore, each objective function, some constraints and parameters including the fuzzy variable $\tilde{r}_{tj}$ are also assumed to fuzzy variables. Consequently, we can't optimize problem (11) without transforming its objective function into another form. In this paper, we consider the necessity measure maximum model to variances of scenarios. We assume f to the target level which is sufficiently small number, and convert problem (11) into the following problem including the chance-constraint.

$$\begin{aligned}&\text{Maximize} \quad \alpha\\&\text{subject to} \quad \Pr\left\{\sum_{s\in A} p_s\right\} \ge \beta,\ A=\left\{k\left|\text{Nes}\left\{\left(\sum_{j=1}^{n}\left(r_{kj}-r_j\right)x_j\right)^2 \le f\right\}\ge\alpha\right.\right\}\\&\qquad \text{Pos}\left\{\sum_{j=1}^{n} r_j x_j \ge \rho\right\}\ge h,\ \tilde{r}_j=\sum_{s\in A} p_s\tilde{r}_{sj},\quad \sum_{j=1}^{n}x_j=1,\quad 0\le x_j\le b_j,\quad j=1,\cdots n\end{aligned} \tag{12}$$

Here, we consider each membership function of fuzzy variables. First, when we consider the membership function of $\tilde{r}_j$, it is given as follows.

$$\mu_{\tilde{r}_j}(\omega)=\left(\sum_{s\in A} p_s\overline{r}_{sj}, P_s\alpha_j, P_s\beta_j\right)_{LR},\quad j=1,2,\cdots,n,\ \ P_s=\sum_{s\in A}p_s \tag{13}$$

From this membership function, that of $Y=\sum_{j=1}^{n}\tilde{r}_j x_j$ and its α-cut become following forms.

$$\begin{aligned}\mu_{\tilde{Y}}(\omega)&=\left(\sum_{j=1}^{n}\left(\sum_{s\in A}p_s\overline{r}_{sj}\right)x_j, \sum_{j=1}^{n}P_s\alpha_j x_j, \sum_{j=1}^{n}P_s\beta_j x_j\right)_{LR}\\&=\left[\sum_{j=1}^{n}\left(\sum_{s\in A}p_s\overline{r}_{sj}\right)x_j - L^*(\alpha)\sum_{j=1}^{n}P_s\alpha_j x_j,\ \sum_{j=1}^{n}\left(\sum_{s\in A}p_s\overline{r}_{sj}\right)x_j + R^*(\alpha)\sum_{j=1}^{n}P_s\beta_j x_j\right]\end{aligned} \tag{14}$$

Next, we consider the membership function of

$$Z_s=\left(\sum_{j=1}^{n}\left(r_{sj}-r_j\right)x_j\right)^2=\left(\sum_{j=1}^{n}\left(r_{sj}-\sum_{s\in A}p_s r_{sj}\right)x_j\right)^2=\sum_{i=1}^{n}\sum_{j=1}^{n}\left(r_{si}-\sum_{s\in A}p_s r_{si}\right)\left(r_{sj}-\sum_{s\in A}p_s r_{sj}\right)x_i x_j .$$

However, this membership function includes the difference and square calculation between two L-R fuzzy numbers, so it is very difficult to apply to them. Therefore we consider the calculation using α-cut to each L-R fuzzy number. Here, we introduce

m which is the number of scenarios satisfying $\Pr\left\{\sum_{s\in A} p_s\right\} \geq \beta$. First, considering inequality (10), the α-cut of each membership function to fuzzy product between r_{sj} and r_{ki} is as follows.

$$\begin{aligned}&\mu_{r_{sj}\bullet r_{ki}}(\omega)=\left[\mu_{r_{sj}}^{(1)}(\alpha)\mu_{r_{ki}}^{(1)}(\alpha),\ \mu_{r_{sj}}^{(2)}(\alpha)\mu_{r_{ki}}^{(2)}(\alpha)\right]\\&\mu_{r_{sj}}^{(1)}(\alpha)=\bar{r}_{sj}-L^*(\alpha)\alpha_j,\ \ \mu_{r_{sj}}^{(2)}(\alpha)=\bar{r}_{sj}+R^*(\alpha)\beta_j\\&k,s=1,2,\cdots,m,\ \ i,j=1,2,\cdots,n\end{aligned} \tag{15}$$

From this α-cut, we find the α-cut of each membership function to $Y_{s1}=\sum_{j=1}^{n}(r_{si}x_i)(r_{sj}x_j)$, $Y_{s2}=(r_{si}x_i)\sum_{j=1}^{n}\sum_{s=1}^{m}p_s r_{sj}x_j$, $Y_{s3}=\sum_{j=1}^{n}\left(\sum_{s=1}^{m}p_s r_{si}x_i\right)(r_{sj}x_j)$, $Y_{s4}=\left(\sum_{s=1}^{m}p_s r_{si}x_i\right)\sum_{j=1}^{n}\left(\sum_{s=1}^{m}p_s r_{sj}x_j\right)$ as the following form.

$$\begin{cases}\mu_{Y_{s1}}(\omega)=\left[\mu_{Y_{s1}}^{(1)}(\alpha),\ \mu_{Y_{s1}}^{(2)}(\alpha)\right]=\left[\mu_{r_{si}}^{(1)}(\alpha)x_i\sum_{j=1}^{n}\mu_{r_{sj}}^{(1)}(\alpha)x_j,\ \mu_{r_{si}}^{(2)}(\alpha)x_i\sum_{j=1}^{n}\mu_{r_{sj}}^{(2)}(\alpha)x_j\right]\\ \mu_{Y_{s2}}(\omega)=\left[\mu_{Y_{s2}}^{(1)}(\alpha),\ \mu_{Y_{s2}}^{(2)}(\alpha)\right]=\left[\mu_{r_{si}}^{(1)}(\alpha)x_i\sum_{s=1}^{m}p_s\sum_{j=1}^{n}\mu_{r_{sj}}^{(1)}(\alpha)x_j,\ \mu_{r_{si}}^{(2)}(\alpha)x_i\sum_{s=1}^{m}p_s\sum_{j=1}^{n}\mu_{r_{sj}}^{(2)}(\alpha)x_j\right]\\ \mu_{Y_{s3}}(\omega)=\left[\mu_{Y_{s3}}^{(1)}(\alpha),\ \mu_{Y_{s3}}^{(2)}(\alpha)\right]=\left[\sum_{s=1}^{m}p_s\mu_{r_{si}}^{(1)}(\alpha)x_i\sum_{j=1}^{n}\mu_{r_{sj}}^{(1)}(\alpha)x_j,\ \sum_{s=1}^{m}p_s\mu_{r_{si}}^{(2)}(\alpha)x_i\sum_{j=1}^{n}\mu_{r_{sj}}^{(2)}(\alpha)x_j\right]\\ \mu_{Y_{s4}}(\omega)=\left[\mu_{Y_{s4}}^{(1)}(\alpha),\ \mu_{Y_{s4}}^{(2)}(\alpha)\right]=\left[\sum_{s=1}^{m}\sum_{s'=1}^{m}p_s p_{s'}\mu_{r_{si}}^{(1)}(\alpha)x_i\sum_{j=1}^{n}\mu_{r_{sj}}^{(1)}(\alpha)x_j,\ \sum_{s=1}^{m}\sum_{s'=1}^{m}p_s p_{s'}\mu_{r_{si}}^{(2)}(\alpha)x_i\sum_{j=1}^{n}\mu_{r_{s'j}}^{(2)}(\alpha)x_j\right]\end{cases} \tag{16}$$

Therefore we get the α-cut of membership function to Z_s as following form.

$$\begin{aligned}\mu_{Z_s}(\omega)&=\left[\mu_{Z_s}^{(1)}(\omega),\ \mu_{Z_s}^{(2)}(\omega)\right]\\&=\left[\sum_{i=1}^{n}\left(\mu_{Y_{s1}}^{(1)}(\alpha)-\mu_{Y_{s2}}^{(2)}(\alpha)-\mu_{Y_{s3}}^{(2)}(\alpha)+\mu_{Y_{s4}}^{(1)}(\alpha)\right),\ \sum_{i=1}^{n}\left(\mu_{Y_{s1}}^{(2)}(\alpha)-\mu_{Y_{s2}}^{(1)}(\alpha)-\mu_{Y_{s3}}^{(1)}(\alpha)+\mu_{Y_{s4}}^{(2)}(\alpha)\right)\right]\end{aligned} \tag{17}$$

Furthermore, we introduce a fuzzy goal to f is represented by

$$\mu_G(\omega)=\begin{cases}1, & \omega\leq f_0\\ \dfrac{f_1-\omega}{f_1-f_0}, & f_0\leq\omega\leq f_1\\ 0, & f_1\leq\omega\end{cases} \tag{18}$$

Using a concept of necessity measure, we define a degree of necessity as follows.

$$N_{\tilde{Z}}(G)=\inf_t \max\left\{\left(1-\mu_{\tilde{Z}_1}(t)\right),\cdots,\left(1-\mu_{\tilde{Z}_m}(t)\right),\mu_G(t)\right\} \tag{19}$$

Therefore, we transform the main problem into the next problem.

$$\begin{aligned}&\text{Maximize}\quad \alpha\\&\text{subject to}\quad N_{\tilde{Z}}(G)\geq\alpha,\\&\qquad \text{Pos}\left\{\sum_{j=1}^{n}r_jx_j\geq\rho\right\}\geq h,\ \tilde{r}_j=\sum_{s=1}^{m}p_s\tilde{r}_{sj},\quad \sum_{j=1}^{n}x_j=1,\quad 0\leq x_j\leq b_j,\quad j=1,\cdots,n\\&\qquad \bar{r}_{sj}-L^*(\alpha)\alpha_j\geq 0,\quad N_{\tilde{Z}}(G)=\inf_t \max\left\{\left(1-\mu_{\tilde{Z}_1}(t)\right),\cdots,\left(1-\mu_{\tilde{Z}_m}(t)\right),\mu_G(t)\right\}\end{aligned} \tag{20}$$

The constraint $N_{\tilde{Z}}(G) \geq \alpha$ implies

$$\begin{aligned}
& N_{\tilde{Z}}(G) \geq \alpha \\
& \Leftrightarrow 1-\mu_{\tilde{Z}_s}(f) < \alpha \ (s=1,2,\cdots m) \Rightarrow \mu_G(f) \geq \alpha \\
& \Leftrightarrow \mu_{Z_s}^{(2)}(1-\alpha) \leq f \ (s=1,2,\cdots m) \Rightarrow (1-\alpha) f_1 + \alpha f_0 \geq f \\
& \Leftrightarrow \mu_{Z_s}^{(2)}(1-\alpha) \leq (1-\alpha) f_1 + \alpha f_0 \ (s=1,2,\cdots m)
\end{aligned} \tag{21}$$

From this inequality, the main problem is transformed into the following problem.

$$\begin{aligned}
& \text{Maximize} \quad \alpha \\
& \text{subject to} \quad \mu_{Z_s}^{(2)}(1-\alpha) \leq (1-\alpha) f_1 + \alpha f_0, \ s=1,\cdots,m \\
& \qquad \text{Pos}\left\{\sum_{j=1}^{n} r_j x_j \geq \rho\right\} \geq h, \ \tilde{r}_j = \sum_{s=1}^{m} p_s \tilde{r}_{sj}, \\
& \qquad \overline{r}_{sj} - L^*(\alpha)\alpha_j \geq 0, \quad \sum_{j=1}^{n} x_j = 1, \quad 0 \leq x_j \leq b_j, \quad j=1,\cdots,n
\end{aligned} \tag{22}$$

3.2 Solution Method of Fuzzy Multiobjective Programming Problem

Problem (22) is a nonlinear programming problem, not linear programming and convex programming problem. However, if the parameter α is fixed, problem (22) is equivalent to a convex programming problem, therefore, we construct an efficient solution method to find a global optimal solution. First, we consider the following problem.

$$\begin{aligned}
& \text{Minimize} \quad M(\alpha) = \max_s \left[\mu_{Z_s}^{(2)}(1-\alpha), \ (s=1,\cdots,m)\right] \\
& \text{subject to} \quad \sum_{j=1}^{n}\left(\sum_{s=1}^{m} p_s \overline{r}_{sj}\right) x_j + R^*(h) \sum_{j=1}^{n} P_s \beta_j x_j \geq \rho, \\
& \qquad \overline{r}_{sj} - L^*(\alpha)\alpha_j \geq 0, \quad \sum_{j=1}^{n} x_j = 1, \quad 0 \leq x_j \leq b_j, \quad j=1,2,\cdots,n \\
& \text{where} \quad \mu_{Z_s}^{(2)}(1-\alpha) = \sum_{i=1}^{n}\left(\mu_{Y_{s1}}^{(2)}(1-\alpha) - \mu_{Y_{s2}}^{(1)}(1-\alpha) - \mu_{Y_{s3}}^{(1)}(1-\alpha) + \mu_{Y_{s4}}^{(2)}(1-\alpha)\right) \\
& \qquad = \left(\sum_{j=1}^{n} \mu_{r_{sj}}^{(2)}(1-\alpha) x_j\right)^2 + \left(\sum_{j=1}^{n}\sum_{s=1}^{m} p_s \mu_{r_{sj}}^{(2)}(1-\alpha) x_j\right)^2 - 2\left(\sum_{j=1}^{n} \mu_{r_{sj}}^{(1)}(1-\alpha) x_j\right)\left(\sum_{j=1}^{n}\sum_{s=1}^{m} p_s \mu_{r_{sj}}^{(1)}(1-\alpha) x_j\right)
\end{aligned} \tag{23}$$

In this problem, each objective function $\mu_{Z_s}^{(2)}(1-\alpha)$ is a quadratic convex function and each constraint is a linear constraint. Consequently, it is equivalent to a multiobjective quadratic programming problem. To solve this problem, we introduce the following subproblem.

$$\begin{aligned}
& \text{Minimize} \quad \mu_{Z_s}^{(2)}(1-\alpha) \\
& \text{subject to} \quad \sum_{j=1}^{n}\left(\sum_{s=1}^{m} p_s \overline{r}_{sj}\right) x_j + R^*(h) \sum_{j=1}^{n} P_s \beta_j x_j \geq \rho, \\
& \qquad \overline{r}_{sj} - L^*(\alpha)\alpha_j \geq 0, \quad \sum_{j=1}^{n} x_j = 1, \quad 0 \leq x_j \leq b_j, \quad j=1,\cdots,n
\end{aligned} \tag{24}$$

This problem is also equivalent to a quadratic convex programming problem, therefore, we find its optimal solution. Here, we assume each optimal value of objective functions $g_k^*, \ (k=1,\cdots,m)$, and consider the value $g_K^* = \max_k \left[g_k^*, \ (k=1,\cdots,m)\right]$. Now we assume an optimal solution of g_K^* to x_K^*. If we substitute x_K^* into other objective functions g_k, $\left(k \neq K\right)$ and $\forall k,\ g_k\left(x_K^*\right) \leq g_K^*$, x_K^* is an optimal solution of the problem (23). However, if not, we consider the following problem about the any objective functions satisfying $g_k\left(x_K^*\right) > g_K^*$.

$$\begin{array}{ll} \text{Minimize} & \mu_{Z_s}^{(2)}\left(1-\alpha\right) \\ \text{subject to} & \sum_{j=1}^{n}\left(\sum_{s=1}^{m} p_s \overline{r}_{sj}\right) x_j + R^*\left(h\right)\sum_{j=1}^{n} P_s \beta_j x_j \geq \rho, \\ & \mu_{Z_{s'}}^{(2)}\left(1-\alpha\right) \leq \mu_{Z_s}^{(2)}\left(1-\alpha\right), \\ & \overline{r}_{sj} - L^*\left(\alpha\right)\alpha_j \geq 0,\quad \sum_{j=1}^{n} x_j = 1,\quad 0 \leq x_j \leq b_j,\quad \delta = \begin{cases} -L^*\left(1-\alpha\right), & i=1 \\ R^*\left(1-\alpha\right), & i=2 \end{cases},\quad s, s' = 1, 2, \cdots m \end{array} \tag{25}$$

This problem is also transformed into the following problem including a parameter λ.

$$\begin{array}{ll} \text{Minimize} & \left(1-\lambda\right)\mu_{Z_s}^{(2)}\left(1-\alpha\right) + \lambda \mu_{Z_{s'}}^{(2)}\left(1-\alpha\right) \\ \text{subject to} & \sum_{j=1}^{n}\left(\sum_{t \in A} p_t \overline{r}_{jt}\right) x_j + R^*\left(h\right)\sum_{j=1}^{n} P_t \beta_j x_j \geq \rho, \\ & \overline{r}_{sj} - L^*\left(\alpha\right)\alpha_j \geq 0,\quad \sum_{j=1}^{n} x_j = 1,\quad 0 \leq x_j \leq b_j,\quad \delta = \begin{cases} -L^*\left(1-\alpha\right), & i=1 \\ R^*\left(1-\alpha\right), & i=2 \end{cases},\quad s, s' = 1, 2, \cdots m \end{array} \tag{26}$$

In this problem, if we fix the value of parameter λ, $\left(0 < \lambda < 1\right)$, it is equivalent to a quadratic convex programming problem, so we find an optimal solution. Furthermore, we find the theory that the optimal value to minimize in any objective functions becomes a global optimal solution of main problem (22). Therefore, we get a global optimal solution using the following solution algorithm.

Algorithm

STEP1: Set $\alpha \leftarrow 1$ and solve the problem (24). If $\forall k,\ g_k\left(x_K^*\right) \leq g_K^*$, x_K^* is an global optimal solution of main problem.

STEP2: Set $\alpha \leftarrow 0$ and solve the problem (24). If $\min_k\left[g_k^*\right] > f_1$, there is no feasible solution and it is necessary to reset a fuzzy goal and the target level f

STEP3: Set $U_\alpha \leftarrow 1,\ L_\alpha \leftarrow 0$

STEP4: Set $\alpha \leftarrow \frac{U_\alpha + L_\alpha}{2}$.

STEP5: Using the bisection algorithm about λ, solve the problem (24) and (26) to find an optimal solution of the problem (23). We assume $f = \left(1-\alpha\right) f_1 + \alpha f$. If $M\left(\alpha\right) \leq f$, set $U_\alpha \leftarrow \alpha$ and return to STEP4. If $M\left(\alpha\right) > f$, set $L_\alpha \leftarrow \alpha$ and return to STEP5. If $M\left(\alpha\right) = f$, then terminate the algorithm. Its solution becomes a global optimal solution of main problem (23).

4 Conclusion

In this paper, we have proposed a fuzzy extension model to minimize each variance about portfolio selection problem including the ambiguous future return and develop an efficient solution method using the sub convex programming problem. We may be able to apply this solution method to the case including not only fuzziness but also both randomness and fuzziness, which is called to fuzzy random variable or random fuzzy variable. Furthermore, as future studies, we need to consider not only mean-variance portfolio selection problem but other portfolio selection models; i.e. mean-absolute model, semi-variance model and safety first model. And we are now attacking the case that optimal solution is restricted to integers, and the case that the period when we need to consider the portfolio selection problem is not only single but also multi-period.

References

[1] V.S.Bawa, E.B.Lindenberg (1977),"Capital market equilibrium in a mean-lower partial moment framework",*Journal of Financial Economics*,5 189-200.
[2] J.Y.Campbell, A.W.Lo. A.C.MacKinlay, (1997) "The Econometrics of Finance Markets", Princeton University Press, Princeton, NJ.
[3] E.J.Elton, M.J.Gruber, (1995) *Modern Portfolio Theory and Investment Analysis*, Wiley, New York.
[4] M.Inuiguchi, J.Ramik,(2000)"Possibilistic linear programming: A brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem",*Fuzzy Sets and Systems*, 111, 3-28,.
[5] P.Jorion, (1992) "Portfolio optimization in practice", *Financial Analysis Journal*, Jan.-Feb., 68-74.
[6] H.Katagiri,H.Ishii,M.Sakawa (2004)."On fuzzy random linear knapsack problems", *Central European Journal of Operations Research*,Vol.12, 1, 59-70,
[7] H.Katagiri, M.Sakawa, H.Ishii, (2005) "A study on fuzzy random portfolio selection problems using possibility and necessity measures", *Scientiae Mathematicae Japonocae,* Vol.65 No.2, 361-369.
[8] H.Konno, (1990) "Piecewise linear risk functions and portfolio optimization", *Journal of Operations Research Society of Japan*, 33 139-159.
[9] H.Konno, H.Shirakawa, H.Yamazaki, (1993) "A mean-absolute deviation-skewness portfolio optimization model", *Annals of Oper. Res.*, 45 205-220,.
[10] T.Leon, V.Liern, E.Vercher, (2002) "Validity of infeasible portfolio selection problems: fuzzy approach", *European Journal of Oper. Res.*, 139, 178-189.
[11] B.Liu, *Theory and Practice of Uncertain Programming*, Physica Verlag, (2002).
[12] David G. Luenberger, (1997). *Investment Science*, Oxford Univ. Press.
[13] H.Markowits, (1952). *Portfolio Selection*, Wiley, New York,
[14] R.T.Rockafellar and S.Uryasev, (2000) "Optimization of conditional value-at-risk", *Journal of Risk*, 2(3) 1-21.
[15] Marc C. Steinbach, (2001) "Markowitz revisited: Mean-variance models in financial portfolio analysis", *SIAM Review*, Vol. 43, No. 1, 31-85.
[16] H.Tanaka, P.Guo, (1999). "Portfolio selection based on upper and lower exponential possibility distributions", *Euro. Journal of Oper. Res.* 114, 115-126.

[17] H.Tanaka, P.Guo, I.B.Turksen, (2000) “Portfolio selection based on fuzzy probabilities and possibility distributions”, *Fuzzy Sets and Systems*, 111, 387-397.
[18] S.Vajda, (1972) *Probabilitstic Programming*, Academic Press.
[19] J.Watada, (1997) “Fuzzy portfolio selection and its applications to decision making”, Tatra Mountains Math. Pub. 13, 219-248.

Optimization of Fuzzy Objective Functions in Fuzzy (Multicriteria) Linear Programs - A Critical Survey

Heinrich J. Rommelfanger

J. W. Goethe University Frankfurt am Main, Institute of Statistics and Mathematics, Mertonstr. 17-23, D-60054 Frankfurt am Main, Germany
Rommelfanger@wiwi.uni-frankfurt.de

Abstract. For calculating a solution of a linear program where coefficients of the objective function(s) may be fuzzy, we have to explain how the optimization of a fuzzy objective can be interpreted. In the literature of fuzzy linear programming, a lot of procedures for substituting fuzzy objectives by crisp ones are proposed. In this paper, a critical survey of these different methods is given.

Keywords: Fuzzy linear programming, fuzzy objective functions, fuzzy inequality relations, compromise solution, set of feasible solutions.

1 Introduction

Modeling economic problems by linear programs, we are often faced with the problem that not all coefficients of the objective function(s) or of the constraints can be determined as real numbers. However, since Lotfi A. Zadeh [19] published his fundamental paper "Fuzzy Sets" in 1965, there exists a convenient and powerful way of modeling vague data according to the subjective imaginations of a decision maker (DM). Nevertheless, one must not forget that the result can only be as precise as the DM can express himself.

A general model of a fuzzy linear programming problem (FLP-problem) is the following system:

$$\tilde{Z}(\mathbf{x}) = \tilde{C}_1 x_1 \oplus \cdots \oplus \tilde{C}_n x_n \to \widetilde{\text{Max}} \tag{1a}$$

subject to

$$\tilde{A}_{i1} x_1 \oplus \tilde{A}_{i2} x_2 \oplus \cdots \oplus \tilde{A}_{in} x_n \ \tilde{\leq}\ \tilde{B}_i\ , \quad i = 1, \ldots, m, \tag{1b}$$

$$x_1, x_2, \ldots, x_n \geq 0 \tag{1c}$$

where $\tilde{A}_{ij}, \tilde{B}_i, \tilde{C}_j,\ i = 1,\ldots,m,\ j = 1,\ldots,n$ are fuzzy sets in **R**. The symbol ⊕ represents the extended addition.

As each real number c can be modeled as a fuzzy number, the general system (1) includes the special cases, that the objective function or all constraints are crisp.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 324–333, 2007.
springerlink.com

The application of FLP-systems offers the advantage that the decision maker can model his problem in accordance to his current state of information. At the same time he is no longer able to use the well known simplex algorithms for computing a solution of his problem. Therefore various procedures for calculating a compromise solution of an FLP-system (1) have been developed, see [13].

In this paper we will concentrate on the handling of fuzzy objectives in literature. The set of feasible solutions X may be described by a crisp or a fuzzy set. Moreover we will assume that all fuzzy coefficients $\tilde{C}_j$ can be described by fuzzy intervals of the L-R-type with the same reference functions, i.e. $\tilde{C}_j = \tilde{C}_j = (\underline{c}_j;\overline{c}_j;\underline{\gamma}_j;\overline{\gamma}_j)_{LR}$. In literature the linear reference function L(u) = R(u) = Max (0, 1 - u) is mostly used.

Then, using Zadeh's extension principle the fuzzy objective function

$$\tilde{Z}(\mathbf{x})=\tilde{C}_1 x_1 \oplus \cdots \oplus \tilde{C}_n x_n$$

can be aggregated to the fuzzy interval of the L-R-type

$$\tilde{Z}(x) = (\underline{c}(x);\overline{c}(\mathbf{x});\underline{\gamma}(x);\overline{\gamma}(x))_{LR},$$

where $\underline{c}(x) = \sum_{j=1}^{n} \underline{c}_j x_j$, $\overline{c}(x) = \sum_{j=1}^{n} \overline{c}_j x_j$, $\underline{\gamma}(x) = \sum_{j=1}^{n} \underline{\gamma}_j x_j$, $\overline{\gamma}(x) = \sum_{j=1}^{n} \overline{\gamma}_j x_j$.

2 The Multiobjective Approach

It seems clear that a fuzzy objective

$$\tilde{Z}(\mathbf{x})=\tilde{C}_1 x_1 \oplus \cdots \oplus \tilde{C}_n x_n \rightarrow \widetilde{\text{Max}}\, x \tag{2}$$

should be interpreted as an multi-objective problem.

Even in the assumed simple case in which the coefficients $\tilde{C}_j$ have the form $\tilde{C}_j = (\underline{c}_j;\overline{c}_j;\underline{\gamma}_j;\overline{\gamma}_j)_{LR}$ and $\tilde{Z}(\mathbf{x})$ can be written as $\tilde{Z}(x) = (\underline{c}(x);\overline{c}(\mathbf{x});\underline{\gamma}(x);\overline{\gamma}(x))_{LR}$, the fuzzy objective function (3) implies that the four goals

$$\begin{array}{ll} \underline{c}(x) \rightarrow \text{Max} & \underline{c}(x) - \underline{\gamma}(x) \rightarrow \text{Max} \\ \overline{c}(x) \rightarrow \text{Max} & \overline{c}(x) + \overline{\gamma}(x) \rightarrow \text{Max} \end{array} \tag{3}$$

should be satisfied simultaneously on the set of feasible solutions X_U. An ideal solution of this problem, i.e. a solution that maximizes all objectives at the same time, does not exist in general.

In the special case of triangular coefficients $\tilde{C}_j = (c_j;\underline{\gamma}_j;\overline{\gamma}_j)$ the set (3) is reduced to the three objectives:

$$z_1(\mathbf{x}) = c(\mathbf{x}) \rightarrow \text{Max} \;\;,\;\; z_2(\mathbf{x}) = \underline{\gamma}(\mathbf{x}) \rightarrow \text{Min} \;\;,\;\; z_3(\mathbf{x}) = \overline{\gamma}(\mathbf{x}) \rightarrow \text{Max} .$$

A. Lai; Hwang [6] proposed to substitute these objectives by fuzzy objective functions with linear membership functions, in which the basic values, the positive and the negative ideal solutions should be calculated as:

$$z_1^{PIS} = c(\mathbf{x}_I) = \underset{x \in X_U}{\text{Max}}\ c(\mathbf{x}) \qquad z_2^{PIS} = \underline{\gamma}(\mathbf{x}_{II}) = \underset{x \in X_U}{\text{Min}}\ \underline{\gamma}(\mathbf{x}) \qquad z_3^{PIS} = \overline{\gamma}(\mathbf{x}_{III}) = \underset{x \in X_U}{\text{Max}}\ \overline{\gamma}(\mathbf{x})$$

$$z_1^{NIS} = \underset{x \in X_U}{\text{Min}}\ c(\mathbf{x}) \qquad z_2^{NIS} = \underset{x \in X_U}{\text{Max}}\ \underline{\gamma}(\mathbf{x}) \qquad z_3^{NIS} = \underset{x \in X_U}{\text{Min}}\ \overline{\gamma}(\mathbf{x}).$$

These definitions, especially those of the values z_1^{NIS}, z_2^{NIS} and z_3^{NIS} are not very convincing. In general, the values z_1^{NIS} and z_3^{NIS} are too small and z_2^{NIS} is too large, see the proposal below.

Therefore the linear membership functions do not seem to be adequately modeled. A better proposal would be:

$$z_1^{NIS} = \underset{x \in X_U}{\text{Min}}\ [c(\mathbf{x}_{II}), c(\mathbf{x}_{III})] \qquad z_2^{NIS} = \underset{x \in X_U}{\text{Max}}\ [\underline{\gamma}(\mathbf{x}_I), \underline{\gamma}(\mathbf{x}_{III})]$$

$$z_3^{NIS} = \underset{x \in X_U}{\text{Min}}\ [\overline{\gamma}(\mathbf{x}_I), \overline{\gamma}(\mathbf{x}_{II})]$$

3 Compromise Objective Functions(s) Based on Defuzzification

B. The first method for getting a compromise solution of (1) was proposed by Tanaka; Ichihashi; Asai [18]. They substitute the fuzzy objective by the crisp compromise objective function

$$Z_1(\mathbf{x}) = \frac{1}{6} \cdot \sum_{j=1}^{n} (2\underline{c}_j + 2\overline{c}_j + \underline{\gamma}_j + \overline{\gamma}_j) \cdot x_j \to \text{Max} \tag{4}$$

C. For controlling the robustness of the solution, Stanciulescu [17] supplements the crisp objective function (4) with three additional objective functions which express three types of uncertainties:

"Impression attached to any fuzzy data"

$$Z_2(\mathbf{x}) = \frac{1}{2} \cdot \sum_{j=1}^{n} (2\overline{c}_j + \overline{\gamma}_j - 2\underline{c}_j - \underline{\gamma}_j) \cdot x_j \to \text{Min} \tag{5}$$

"Fuzziness of a fuzzy data"

$$Z_3(\mathbf{x}) = \frac{1}{2} \cdot \sum_{j=1}^{n} (\overline{\gamma}_j + \underline{\gamma}_j) \cdot x_j \to \text{Min} \tag{6}$$

"Variety of a fuzzy data"

$$Z_4(\mathbf{x}) = \frac{1}{2} \cdot \sum_{j=1}^{n} (\overline{c}_j - \underline{c}_j) \cdot x_j \to \text{Min} \tag{7}$$

Then according to the MAUT approach, a utility function $u_k(Z_k(\mathbf{x}))$ is defined for each objective function $Z_k(\mathbf{x})$ and the global utility function is defined by summing up all these utility functions:

$$U(\mathbf{x}) = \sum_{k=1}^{4} u_k(Z_k(\mathbf{x})) \to \text{Max} \tag{8}$$

4 Compromise Objective Functions(s) Based on σ-Level(s)

D. Carlsson; Korhonen [3] proposed to substitute the maximized fuzzy objective function (2) by the crisp linear objective function

$$Z(\mathbf{x}) = (\bar{c}_1 + R_{C_1}^{-1}(\sigma)\bar{\gamma}_1)\cdot x_1 + \cdots + (\bar{c}_n + R_{C_n}^{-1}(\sigma)\bar{\gamma}_n)\cdot x_n \to \text{Max} \tag{9}$$

where $\sigma \in [0,1]$ is a membership level fixed by the DM. R_{C_j} is the right-hand side reference function of the fuzzy interval $\tilde{C}_j$.

According to the approach of Carlsson; Korhonen, the fuzzy constraints of a FLP-system should be substituted by crisp surrogates based on the same membership level σ. Then, the System (1) with fuzzy sets of the LR-type can be substituted by the crisp LP-problem

$$Z(\mathbf{x}) = (\bar{c}_1 + R_{C_1}^{-1}(\sigma)\bar{\gamma}_1)\cdot x_1 + \cdots + (\bar{c}_n + R_{C_n}^{-1}(\sigma)\bar{\gamma}_n)\cdot x_n \to \text{Max} \tag{10a}$$

subject to

$$\sum_{j=1}^{n}\left(\bar{a}_i(\mathbf{x}) + \bar{\alpha}_i(\mathbf{x})\cdot R_{A_i}^{-1}(\sigma)\right) \le b_i + \beta_i \cdot R_{B_i}^{-1}(\sigma)\,,\ i = 1,2,\ldots,m;\ \ \mathbf{x} \in R_0^n \tag{10b}$$

Carlsson; Korhonen [3] proposed to calculate a set of solutions $\{ (z^*(\sigma), x^*(\sigma)) | \sigma \}$ by changing the parameter $\sigma \in [0,1]$. This set should be presented to the DM who can select his preferred solution by taking into account the additional information.

E. The approach of Buckley [1;2] is very similar to the proceeding of Carlsson; Korhonen [3]. The main difference is the more optimistic interpretation of the inequality relation in the fuzzy constraints.

Here, Buckley uses the crisp surrogate

$$\sum_{j=1}^{n}(\underline{a}_i(\mathbf{x}) - \underline{\alpha}_i(\mathbf{x})\cdot L_{A_i}^{-1}(\sigma)) \le b_i + \beta_i \cdot R_{B_i}^{-1}(\sigma)\,,\ i = 1,2,\ldots,m;\ \ \mathbf{x} \in R_0^n \tag{11}$$

where L_{A_i} is the left-hand side reference function of the fuzzy interval $\tilde{A}_j$ and R_{B_i} is the right-hand side reference function of the fuzzy number $\tilde{B}_i$.

F. Ramik [8] presents an optimistic and a pessimistic approach.

In his optimistic approach, he uses an interpretation of the inequality relation in the fuzzy constraints that is more optimistic than (8). It corresponds to optimistic index of Slowinski [16], see term (27a).

$$Z(\mathbf{x})=(\overline{c}_1+R_{C_1}^{-1}(\sigma)\overline{\gamma}_1)\cdot x_1+\cdots+(\overline{c}_n+R_{C_n}^{-1}(\sigma)\overline{\gamma}_n)\cdot x_n\to \text{Max} \tag{12a}$$

subject to

$$\sum_{j=1}^{n}\left(\overline{a}_i(\mathbf{x})-\overline{\alpha}_i(\mathbf{x})\cdot L_{A_i}^{-1}(\sigma)\right)\le b_i+\beta_i\cdot R_{B_i}^{-1}(\sigma)\ ,\ i=1,2,\ldots,m;\quad x\in R_0^n \tag{12b}$$

The pessimistic approach is similar constructed:

$$Z(\mathbf{x})=(\overline{c}_1-L_{C_1}^{-1}(\sigma)\overline{\gamma}_1)\cdot x_1+\cdots+(\overline{c}_n-L_{C_n}^{-1}(\sigma)\overline{\gamma}_n)\cdot x_n\to \text{Max} \tag{13a}$$

subject to

$$\sum_{j=1}^{n}\left(\overline{a}_i(\mathbf{x})+\overline{\alpha}_i(\mathbf{x})\cdot R_{A_i}^{-1}(\sigma)\right)\le b_i-\underline{\beta}_i\cdot R_{B_i}^{-1}(1-\sigma)\ ,i=1,2,\ldots,m;\quad x\in R_0^n \tag{13b}$$

But this inequality relation is not convincing that on the right hand side the parameter 1-σ is used and not σ. Moreover, in a 7-relation the fuzzy right-hand side must have the form $(b_i;0;\overline{\beta})_{LR}$.

G. Delgado; Verdegay; Vila [5] use the border elements of the σ-cuts $C_j^\sigma=[(\underline{c}_j-\underline{\gamma}_j^\sigma),(\overline{c}_j+\overline{\gamma}_j^\sigma)]$ to construct new crisp objective functions. In doing so all combinations are allowed., e. g.

$$Z(\mathbf{x})=(\overline{c}_1+\overline{\gamma}_1^\sigma)\cdot x_1+(\underline{c}_2-\underline{\gamma}_2^\sigma)\cdot x_2+(\underline{c}_3-\underline{\gamma}_3^\sigma)\cdot x_3+\cdots+(\overline{c}_n+\overline{\gamma}_n^\sigma)\cdot x_n\to \text{Max} \tag{14}$$

Then the compromise objective function is the weighted sum of all (up to 2^n) crisp objective functions.

H. Chanas; Kuchta [4] analyse a LP-system with a fuzzy objective function (1a) and a crisp set of feasible solutions X. For calculating a fuzzy solution $\tilde{X}=\{(\mathbf{x},\mu_X(\mathbf{x})\,|\,\mathbf{x}\in\mathbf{R}_{0+}^n\}$ they associate this problem at first with the interval linear programming problems:

$$\underset{x\in X}{\text{Max}}\left(C_1^\sigma x_1+C_2^\sigma x_2+\cdots+C_n^\sigma x_n\right) \tag{15}$$

where $C_j^\sigma=[(\underline{c}_j-\underline{\gamma}_j^\sigma),(\overline{c}_j+\overline{\gamma}_j^\sigma)]$ are the σ-cuts of $\tilde{C}_j$ with a special $\sigma\in[0,1]$.

For getting a solution of the interval linear program (15), Chanas; Kuchta [4] use the concept of an $t_0-t_1-\text{optimalsolution},0\le t_0<t_1\le 1$.

According to this procedure, the problem (15) is transformed to the following parametric bicriterial linear programming problem, where $\theta=L^{-1}(\sigma)$:

$$\operatorname*{Max}_{\mathbf{x}\in X}\left(\begin{array}{l} f_1(\mathbf{x}) = \sum_{j=1}^{n}\left(\underline{c}_j + t_0\cdot(\overline{c}_j - \underline{c}_j) + (t_0(\underline{\gamma}_j^\sigma + \overline{\gamma}_j^\sigma) - \underline{\gamma}_j^\sigma)\cdot\theta\right)\cdot x_j \\ f_2(\mathbf{x}) = \sum_{j=1}^{n}\left(\underline{c}_j + t_1\cdot(\overline{c}_j - \underline{c}_j) + (t_1(\underline{\gamma}_j^\sigma + \overline{\gamma}_j^\sigma) - \underline{\gamma}_j^\sigma)\cdot\theta\right)\cdot x_j \end{array}\right) \tag{16}$$

Now, using a parametric version of the simplex algorithm a basic solution $\mathbf{x}^0$ of (16) is calculated, which is efficient for the set

$$<\theta_0,\theta_1> \cup <\theta_1,\theta_2> \cup\cdots\cup <\theta_{S-1},\theta_S> \tag{17}$$

where $<\theta_{s-1},\theta_s>$ are subintervals of $[L^{-1}(1),L^{-1}(0)]$ which can be closed or open.

Then, the membership degree of the basic solution $\mathbf{x}^0$ is

$$\mu_X(\mathbf{x}^0) = \sum_{s=1}^{S}\left(L(\theta_{s-1}) - L(\theta_s)\right) \tag{18}$$

I. For solving the fuzzy multi-objective linear programming problem

$$\left.\begin{array}{c} \tilde{C}_{11}x_1 \oplus \tilde{C}_{12}x_2 \oplus\cdots\oplus \tilde{C}_{1n}x_n \\ \vdots \\ \tilde{C}_{K1}x_1 \oplus \tilde{C}_{K2}x_2 \oplus\cdots\oplus \tilde{C}_{Kn}x_n \end{array}\right| \to \widetilde{\mathrm{Max}} \tag{19a}$$

subject to

$$\tilde{A}_{i1}x_1 \oplus \tilde{A}_{i2}x_2 \oplus\cdots\oplus \tilde{A}_{in}x_n \;\tilde{\leq}\; \tilde{B}_i\ ,\quad i = 1,\ldots, m,\ x \in R_0^n \tag{19b}$$

Sakawa, Yano [14, 15] propose an interactive solution process, where the stepwise calculation of the "G-σ-pareto-optimal solution" is restricted to σ-level-sets of the coefficients $\tilde{C}_{kj}$ and $\tilde{A}_{ij}$.

Concerning the restrictions, Sakawa and Yano use the formulae (12b) of the optimistic approach of Ramik. The set of feasible solutions is named as X^σ.

Concerning the fuzzy objective functions, the authors substitute each objective function by a monotone increasing membership function $\mu_{Z_k}(\mathbf{c}_k{'}\cdot\mathbf{x})\,\forall k = 1,\ldots,K$.

For determining a "compromise solution" Sakawa and Yano propose that the DM specifies for each objective a "reference membership values" $\overline{\mu}_{Z_k}$.

Then the global solution function

$$Z = \mathrm{Max}\left([\overline{\mu}_{Z_1} - \mu_{Z_1}(\overline{\mathbf{c}}_1^\sigma{'}\cdot\mathbf{x})],\ldots,[\overline{\mu}_{Z_K} - \mu_{Z_K}(\overline{\mathbf{c}}_K^\sigma{'}\cdot\mathbf{x})]\right) \tag{20}$$

where $\overline{\mathbf{c}}_k^\sigma{'} = (\overline{c}_{k1} + \overline{\gamma}_{k1}^\sigma,\ldots,\overline{c}_{kn} + \overline{\gamma}_{kn}^\sigma)$ is to minimized on X^σ.

J. A similar concept is the "β-possibility efficient solution" of Luhandjula [7].

A vector $\mathbf{x}^o \in X$ is called β-possible efficient for the mathematical program

$$\underset{\mathbf{x}\in X}{\mathrm{Max}}\left(\tilde{Z}_1(\mathbf{x}),\cdots,\tilde{Z}_K(\mathbf{x})\right), \quad \text{where} \quad X \subset \mathbf{R}_+^n \tag{21}$$

if there is no other vector $\mathbf{x} \in X$ such that

$$\mathrm{Poss}\begin{pmatrix} \tilde{Z}_1(\mathbf{x}) \geq \tilde{Z}_1(\mathbf{x}^o),\ldots,\tilde{Z}_{k-1}(\mathbf{x}) \geq \tilde{Z}_{k-1}(\mathbf{x}^o), \\ \tilde{Z}_k(\mathbf{x}) > \tilde{Z}_k(\mathbf{x}^o), \tilde{Z}_{k+1}(\mathbf{x}) \geq \tilde{Z}_{k+1}(\mathbf{x}^o),\ldots,\tilde{Z}_K(\mathbf{x}) \geq \tilde{Z}_K(\mathbf{x}^o) \end{pmatrix} \geq \beta, \tag{22}$$

where Poss denotes possibility.

Obviously, $\mathbf{x}^o$ is β-possible efficient for (22) if and only if $\mathbf{x}^o$ is an efficient solution of the program with interval coefficients

$$\underset{\mathbf{x}\in X}{\mathrm{Max}}\left(c_1^{\beta}{}'\cdot\mathbf{x},\cdots,c_K^{\beta}{}'\cdot\mathbf{x}\right) \tag{23}$$

where $c_k^{\beta}{}' = \left([\underline{c}_{k1}^{\beta} - \underline{\gamma}_{k1}^{\beta}, \bar{c}_{k1}^{\beta} + \bar{\gamma}_{k1}^{\beta}],\ldots,[\underline{c}_{kn}^{\beta} - \underline{\gamma}_{kn}^{\beta}, \bar{c}_{kn}^{\beta} + \bar{\gamma}_{kn}^{\beta}]\right)$ are vectors, which are composed of the β-cuts of the coefficients $\tilde{C}_{kj}$, j = 1,.2,..,n; k = 1,2,..,K.

For calculating an efficient solution of (23), Luhandjula proposes a procedure which is based on the lower and upper bounds $\underline{c}_{kj}^{\beta} - \underline{\gamma}_{kj}^{\beta}$ and $\bar{c}_{kj}^{\beta} + \bar{\gamma}_{kj}^{\beta}$ of the β-cuts of the coefficients $\tilde{C}_{kj}$.

K. In contrary to the above approaches the procedure "α-level related pair formation of Rommelfanger, Hanuscheck, Wolf [11] assumes that the convex fuzzy coefficients $\tilde{C}_j$ of the linear fuzzy objective function $\tilde{Z}(\mathbf{x}) = \tilde{C}_1 x_1 + \cdots + \tilde{C}_n x_n$ can be completely described by a few α-level sets

$$C_j^{\sigma} = \left[\underline{c}_j - \underline{\gamma}_j^{\sigma}; \bar{c}_j + \bar{\gamma}_j^{\sigma}\right] \qquad \text{for } \sigma = \sigma_1,\ldots,\sigma_S \in [0,1], S \in \mathbf{N} \tag{24}$$

Then the infinite set of objective functions in (1a) is reduced to the two extreme cases of each α-level. Due to space restrictions, see the further solution procedure in [11].

5 Satisfying Solutions

As an ideal solution of (19) on a set of feasible solutions does not generally exist, Slowinski [16] and Rommelfanger [9, 10, 12] suggest calculating a satisfying solution. This idea which corresponds to the usual way of acting in practice was for the first time used by Zimmermann [20] for calculating a compromise solution of a crisp linear programming problem with several objective functions.

In analogy to modeling a right-hand side $\tilde{B}_i$, a fuzzy aspiration level $\tilde{N}$ can be described as $\tilde{N} = (n;\underline{\nu};0)_{LR}$. Then, the satisfying condition

$$\tilde{N} \tilde{\leq} \tilde{Z}(\mathbf{x}) \tag{26}$$

is treated as an additional fuzzy constraint.

Since it is easy to extend this approach to multicriteria problems, we present the more general case, getting a compromise solution of the multicriteria fuzzy linear programming problem (19), where the fuzzy sets of the constraints are given as $\tilde{A}_{ij} = \left(\underline{a}_{ij}, \bar{a}_{ij}, \underline{\alpha}_{ij}, \bar{\alpha}_{ij}\right)_{LR}$, $\tilde{B}_i = \left(b_i, 0, \beta_i\right)_{RR}$, i=1,...,m.

K. In order to maximize the consistency between

$\tilde{Z}_k(\mathbf{x}) = (\underline{c}_k(\mathbf{x}); \bar{c}(\mathbf{x}); \underline{\gamma}_k(\mathbf{x}); \bar{\gamma}_k(\mathbf{x}))_{LR}$ and $\tilde{N}_k = (n_k; \underline{\nu}_k; 0)_{RR}$ Slowinski [16] proposed to maximize the ordinate $F_k(\mathbf{x})$ of the intersection point of the right slope of $\tilde{Z}(\mathbf{x})$ with the left slope of $\tilde{N}_k$, i.e. $F_k(\mathbf{x}) = R\left(\dfrac{\bar{c}_k(\mathbf{x}) - n_k}{\bar{\gamma}_k(\mathbf{x}) - \underline{\nu}_k}\right)$.

If the reference function R of fuzzy cost coefficients and goals is linear or piece-wise linear, then $F_k(\mathbf{x})$ takes a linear fractional form: $F_k(\mathbf{x}) = 1 - \dfrac{\bar{c}_k(\mathbf{x}) - n_k}{\bar{\gamma}_k(\mathbf{x}) - \underline{\nu}_k}$.

Moreover, Slowinski [16] proposes to substitute the constraints (19b) by

$$\underline{a}_i(\mathbf{x}) - b_i \leq (\underline{\alpha}_i(\mathbf{x}) + \beta_i) \cdot L^{-1}(\tau), \quad \tau \in \left]0,1\right] \quad \text{optimistic index} \tag{27a}$$

$$\bar{a}_i(\mathbf{x}) + \bar{\alpha}_i(\mathbf{x}) \cdot R^{-1}(\eta) \leq b_i + \beta_i \cdot L^{-1}(\eta), \quad \eta \in \left[0,1\right] \quad \text{pessimistic index} \tag{27b}$$

$$\mathbf{x} \geq \mathbf{0}, \qquad i = 1,2,\ldots,m$$

L. Rommelfanger [9, 10, 12] proposed to use the inequality relation "$\tilde{\geq}_R$" for getting crisp surrogates for the fuzzy inequations (26). This interpretation

$$\tilde{Z}_k(\mathbf{x}) \tilde{\geq}_R \tilde{N}_k \iff \begin{cases} \underline{c}_k(\mathbf{x}) + \underline{\gamma}_k(\mathbf{x}) \cdot L^{-1}(\varepsilon) \geq n_k + \underline{\nu}_k \cdot L^{-1}(\varepsilon) & (28a) \\ \mu_{Z_k}(\underline{c}_k(\mathbf{x})) \to \text{Max.} & (28b) \end{cases}$$

is composed of the "pessimistic index" (28a), which is also used by Slowinski [16], and a new objective function (28b). The membership function μ_{Z_k} is defined as

$$\mu_{Z_k}(\underline{c}_k(\mathbf{x})) = \begin{cases} 0 & \text{if } \underline{c}_k(\mathbf{x}) < n_k - \underline{\nu}_k \\ L\left(\dfrac{n_k - \underline{c}_k(\mathbf{x})}{\underline{\nu}_k}\right) & \text{if } n_k - \underline{\nu}_k \leq \underline{c}_k(\mathbf{x}) \leq n_k \\ 1 & \text{if } n_k < \underline{c}_k(\mathbf{x}) \end{cases} \tag{29}$$

and may be interpreted as a subjective evaluation of the needed quantity

$\underline{c}_k(\mathbf{x}) = \sum_{j=1}^{n} \underline{c}_{ij} x_j$ with regard to the fuzzy aspiration level $\tilde{N}_k$.

Concerning the constraints, Rommelfanger proposed a similar procedure:

$$\tilde{A}_i(\mathbf{x}) \tilde{\leq}_R \tilde{B}_i \Leftrightarrow \begin{cases} \bar{a}_i(\mathbf{x}) + \bar{\alpha}_i(\mathbf{x}) \cdot R^{-1}(\varepsilon) \leq b_i + \beta_i \cdot R^{-1}(\varepsilon) & (30a) \\ \mu_i(\bar{a}_i(x)) \to \text{Max.} & (30b) \end{cases}$$

For getting a compromise solution of the crisp mathematical program

$$\underset{x \in R^n}{\text{Max}} \left(\mu_{Z_1}(\underline{c}_1(\mathbf{x})), ..., \mu_{Z_K}(\underline{c}_K(\mathbf{x})), \mu_1(\bar{a}_i(x)), ..., \mu_m(\bar{a}_i(x)) \right) \tag{31a}$$

subject to

$$\underline{c}_k(\mathbf{x}) + \underline{\gamma}_k(\mathbf{x}) L^{-1}(\varepsilon) \geq n_k + \underline{\nu}_k L^{-1}(\varepsilon), \qquad k = 1, ..., K \tag{31b}$$

$$\bar{a}_i(\mathbf{x}) + \bar{\alpha}_i(\mathbf{x}) \cdot R^{-1}(\varepsilon) \leq b_i + \beta_i \cdot R^{-1}(\varepsilon), \qquad i = 1, ..., m \tag{31c}$$

Rommelfanger proposes to substitute the K+m objective functions of the system (31) by the compromise objective function

$$\lambda(\mathbf{x}) = \text{Min}\,(\mu_{Z_1}(\underline{c}_1(\mathbf{x})), ..., \mu_{Z_K}(\underline{c}_K(\mathbf{x})), \mu_1(\bar{a}_1(\mathbf{x})), ..., \mu_m(\bar{a}_m(\mathbf{x}))) \tag{32}$$

which expresses the total satisfaction of the DM with a solution **x**..

If all membership functions μ_{Z_k} and μ_i are piecewise linear and concave, the system (32, 31b, 31c) is a crisp linear program which can be solved by means of the well known simplex algorithms.

6 Conclusions

The discussion of different approaches to transform fuzzy objectives into crisp surrogates reveals the wide range of practicable methods for solving (multi-objective) fuzzy linear programs. For making the best choice, the decision maker has to consider the different assumptions of the suggested procedures and to compare them with the actual decision problem. In any case the solution should be determined step by step in an interactive process, in which additional information out of the decision process itself or from outside should be used. In doing so inadequately modeling of the real problem can be avoided and information costs will, in general, be decreased.

References

1. Buckley, J.J. (1989): Solving possibilistic linear programming problems. Fuzzy Sets and Systems 31 329-341
2. Buckley, J.J. (1995): Joint solution to fuzzy programming problems. Fuzzy Sets and Systems 72 215-220
3. Carlsson, C.;Korhonen, P. (1986): A parametric approach to fuzzy linear programming. Fuzzy Sets and Systems 20, 17-30

4. Chanas, S.; Kuchta, D. (1994): Linear programming problems with fuzzy coefficients in the objective function. In: Delgado M.; Kacprzyk J.; Verdegay J.-L.; Vila M.A., Fuzzy Optimization. Physica-Verlag, Heidelberg 148-157
5. Delgado, M., Verdegay, J.L.; Vila, M.A. (1989): A general model for fuzzy linear programming., Fuzzy Sets and Systems 29 21-30
6. Iniguchi, M,; Ramik, J.; Tanino, T.; Vlach, M. (2003): Satisficing solutions and duality in interval and fuzzy linear programming. Fuzzy Sets and Sys. 135 151-177
7. Lai, Y.-J.; Hwang, C.-L. (1992): Interactive fuzzy linear programming. Fuzzy Sets and Systems 45 169-183
8. Luhandjula, M.K. (1987): Multiple objective programming with possibilistic coefficients. Fuzzy Sets and Systems 21 135-146
9. Ramik, J. (2006): Duality in fuzzy linear programming with possibilistic and necessity relations. Fuzzy Sets and Systems 157 1283-1302
10. Rommelfanger H. (1994): Fuzzy decision support systems - Decisions in fuzzy environment (in German). Springer Verlag, Berlin, Heidelberg, second edition
11. Rommelfanger, H. (1995): FULPAL 2.0 - An interactive algorithm for solving multicriteria fuzzy linear programs controlled by aspiration levels. In: Schweigert D. (Ed.), Methods of multicriteria decision theory. Pfalzakademie, Lamprecht 21-34
12. Rommelfanger, H., Hanuscheck, R. (1995); Wolf, J.: Linear programming with fuzzy objectives, Fuzzy Sets and Systems 29 31-48
13. Rommelfanger, H.; Keresztfalvi, T. (1995): Multicriteria fuzzy optimization based on Yager's parametrized t-norm. Foundations of Computing and Decision Sciences 16 99-110
14. Rommelfanger, H.; Slowinski, R. (1999): Fuzzy Linear Programming with single or multiple Objective Functions. In: Slowinski, R. (Ed.) Fuzzy Sets in Decision Analysis, Operations Research and Statistics. Kluwer Academic Publishers Norwell, Massachusetts, USA 1998 179-213
15. Sakawa, M.; Yano, H. (1989): Interactive fuzzy satisficing method for multiobjective nonlinear programming problems with fuzzy parameters. Fuzzy Sets and Systems 30 221-238.
16. Sakawa, M.; Yano, H. (1990): Interactive decision making for multiobjective programming problems with fuzzy parameters. In: Slowinski R. and Teghem J. (Eds.), Stochastic versus fuzzy approaches to multiobjective mathematical programming under uncertainty. Kluwer Academic Publishers, Dordrecht 191-228
17. Slowinski, R. (1990): 'FLIP': an interactive method for multiobjective linear programming with fuzzy coefficients. In: Slowinski R. and Teghem J. (Eds.), Stochastic versus fuzzy approaches to multi-objective mathematical programming under uncertainty. Kluwer Academic Publishers, Dordrecht 249-262
18. Stanciulescu, Ch. (2001): Multiobjective decision support tools using concepts of fuzzy sets. Université catholique de Louvain, Louvain-La-Neuve, Belgium
19. Tanaka, H.; Ichihashi, H.; Asai, K. (1984): A formulation of linear programming problems based on comparison of fuzzy numbers. Control and Cybernetics 13 185-194.
20. Zadeh, L.A. (1965): Fuzzy sets. Information and Control 8 338-353
21. Zimmermann, H.-J. (1978): Fuzzy programming and linear programming with several objective functions. Fuzzy Sets and Systems 1 45-55

Part IX

Algebraic Foundations of Soft Computing

On Extension of LI-Ideal in Lattice Implication Algebra*

Lai Jiajun[1], Xu Yang[2], and Jun Ma[3]

[1, 2] Intelligent Control Development Center, Southwest Jiaotong University, Chengdu 610031, P.R. China
Tel.: 028-86464926 ; Fax: 028-87600760
eagles_laijiajun7309@126.com, xuyang@swjtu.edu.cn
[3] Department of Mathematics, Southwest Jiaotong University, Chengdu 610031, P.R. China

Abstract. Lattice implication algebra is a logical algebraic system which is constructed by combining lattice with implication algebra. In this paper, we focus on the extension of LI-ideal of lattice implication algebras, i.e., weak LI-ideals (briefly, WLI-ideals) and maximal weak LI-ideals. The properties of weak LI-ideals are studied and several characterizations of weak LI-ideals are given. Finally, we study the relationships among WLI-ideals, LI-ideals and Lattice ideals[1].

Keywords: Lattice implication algebra, *LI*-ideals, *WLI*-ideals, maximal *WLI*-ideals, Lattice ideals.

1 Introduction

Non-classical logic has become a considerable formal tool for artificial intelligence to deal with uncertain information and automated reasoning. Many-valued logic is a great extension of classical logic (see e.g. *[1]*), and provide an interesting alternative to the modeling and reasoning of the classical logic. Hence lattice valued plays an important role (see e.g. *[2, 3]*) in the field of many-valued logic. Goguen *[4]*, Pavelka *[5]*, and Novak *[6]* researched on this lattice-valued logic formal systems. Moreover, in order to research the many-valued logical systems whose propositional value is given in a lattice, in 1990, Xu *[7, 8]* proposed the notion of lattice implication algebras and investigated many useful properties. Since then this logical algebra has been extensively investigated by several researchers (see e.g. *[9, 10, 11, 12]*). In *[13]*, Jun et al. defined the concept of *LI*-ideals in lattice implication algebras and discussed its some properties. As the general development of lattice implication algebras, the ideal theory plays an important role (see e.g. *[11, 15,16, 17]*). In this article, as an extension of above-mention work we propose the concept of *WLI*-ideals in lattice implication algebras, study the properties of *WLI*-ideals, and discuss the relationships among those ideals.

---

* This project was supported by the National Science Foundation of P.R. China under Grant No. 60474022, and the specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20060613007.

[1] Brief introduction of author: Lai Jiajun, male, study the direction: Logic algebra.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 337–348, 2007.
springerlink.com

2 Preliminaries

Definition 2.1 [7]**.** A bounded lattice *(L,* $\vee$, $\wedge$, $'$, *O*, *I)* with ordered-reversing involution $'$ and a binary operation $\rightarrow$ is called a lattice implication algebra if it satisfies the following axioms:

(L_1) $x \rightarrow (y \rightarrow z) = y \rightarrow (x \rightarrow z)$,
(L2) $x \rightarrow x = I$,
(L3) $x \rightarrow y = y' \rightarrow x'$,
(L4) $x \rightarrow y = y \rightarrow x = I$ imply $x = y$,
(L5) $(x \rightarrow y) \rightarrow y = (y \rightarrow x) \rightarrow x$,
(L6) $(x \vee y) \rightarrow z = (x \rightarrow z) \wedge (y \rightarrow z)$,
(L7) $(x \wedge y) \rightarrow z = (x \rightarrow z) \vee (y \rightarrow z)$,
for all $x, y, z \in L$.

A lattice implication algebra *L* is called lattice H implication algebra if it satisfies $x \vee y \vee ((x \wedge y) \rightarrow z) = 1$ for all $x, y, z \in L$.

Theorem 2.1 [8] **.** $(L, \vee, \wedge, ', \rightarrow)$ is a lattice H implication algebra if and only if $(L, \vee, \wedge, ')$ is a Boolean lattice, x' is the complement of x and $x \rightarrow y = x' \vee y$ for any $x, y \in L$.

Definition 2.2 [13]**.** Let *A* be a lattice implication algebra. An *LI*-ideal *A* is non-empty subset of *L* such that for any $x, y \in L$,

(1) $O \in A$;
(2) $(x \rightarrow y)' \in A$ and $y \in A$ imply $x \in A$.

Definition 2.3 [18] **.** A non-empty subset *I* of a lattice implication algebra *L* is said to be an implicative *LI*-ideals (briefly, *ILI*-ideals) of *L* if it satisfies the following conditions:

(1) $o \in I$; *(2)* $(((x \rightarrow y)' \rightarrow y)' \rightarrow z)' \in I$ and $z \in I$ imply $(x \rightarrow y)' \in I$ for any $x, y, z \in L$.

Definition 2.4 [19] **.** Let *L* be a lattice, *A* is non-empty subsets of *L*. *A* is called lattice ideals if it satisfies the follows conditions:

(1) for all $x, y \in A$ imply $x \vee y \in A$;
(2) for any $y \in L, x \in A$, $y \leq x$, then $y \in A$.

Theorem 2.2 [13] **.**Let *A* be an *LI*-ideal of a lattice implication algebra *L*. if $x \leq y$ and $y \in A$ then $x \in A$.

Theorem 2.3. Let *A* be a non-empty subset of a lattice *L*. *A* is a lattice filter of *L* if and only if $A' = \{a' : a \in A\}$ is a lattice ideal of *L*.

Proof. Suppose that *A* is a lattice filter of *L*. Then for any $a\in L, b\in A, b\le a$ imply $a\in A$. Since $a\in L\Rightarrow a^{'}\in L, b\in A\Rightarrow b^{'}\in A$, $b\le a\Leftrightarrow b^{'}\ge a^{'}$, hence $a^{'}\in A$.

Let $a^{'}\in A, b^{'}\in A$, then $a\in A, b\in A$. Since *A* is a lattice filter of *L*, we have $a\wedge b\in A$, i.e., $(a\wedge b)^{'}\in A^{'}$, also $a^{'}\vee b^{'}\in A$. Moreover, $A^{'}$ is a lattice ideal of *L*.

Converse can be proved similarly.

Theorem 2.4. Let *A* be a non-empty subset of a lattice implication algebra *L*. Then A is an *ILI*-ideal of *L* if and only if it satisfies for all $z\in A$ and $y,(x\to y)^{'}\in L$, $((x\to y)^{'}\to y)^{'}\le z$ imply $(x\to y)^{'}\in A$.

Proof. Now suppose first that *A* is an *ILI*-ideal of *L*. Let $z\in A$ and $(x\to y)^{'}\in L$. Then $((x\to y)^{'}\to y)^{'}\le z$ implies $(((x\to y)^{'}\to y)^{'}\to z)^{'}=O\in A$. Using definition we obtain $(x\to y)^{'}\in A$.

Conversely, suppose that $((x\to y)^{'}\to y)^{'}\le z$ implies $(x\to y)^{'}\in A$ for all $z\in A$, $y\in L$, and $(x\to y)^{'}\in L$. Since A is a non-empty subset. Hence $((o\to o)^{'}\to z)^{'}\le z$ implies $o\in A$ holds. On the other hand,

$$((x\to y)^{'}\to y)^{'}\le z\Leftrightarrow((x\to y)^{'}\to y)^{'}\to z=I$$
$$\Leftrightarrow(((x\to y)^{'}\to y)^{'}\to z)^{'}=O\in A.$$

Then we have $(((x\to y)^{'}\to y)^{'}\to z)^{'}\in A$ implies $(x\to y)^{'}\in A$ holds. Moreover *A* is an *ILI*-ideal of *L*. this completes the proof.

3 WLI-Ideals of Lattice Implication Algebra and the New Representations of LI-Ideals and Filters in Lattice Implication Algebra

In a lattice implication algebra *L*, we defined binary operations $\otimes$ and $\oplus$ as follows: for any $x, y\in L$,

$$x\otimes y=(x\to y^{'})^{'}$$
$$x\oplus y=x^{'}\to y.$$

Theorem 3.1 [8] **.** In a lattice implication algebra *L*,

(1) $x\oplus x^{'}=I$, $x\otimes x^{'}=O$, $x\oplus I=I$, $x\otimes I=x$, $x\oplus O=x$, $x\otimes O=O$;

(2) $x\otimes(y\vee z)=(x\otimes y)\vee(x\otimes z)$, $x\otimes(y\wedge z)=(x\otimes y)\wedge(x\otimes z)$;

(3) $x\oplus(y\vee z)=(x\oplus y)\vee(x\oplus z)$, $x\oplus(y\wedge z)=(x\oplus y)\wedge(x\oplus z)$, hold for any $x, y, z\in L$.

Theorem 3.2. In a lattice implication algebra *L*,

(1) $x\otimes y\le x\wedge y$; *(2)* $x\oplus y\ge x\vee y$; *(3)* $x\to(x\oplus y)\le x\to y$;
(4) $(x\oplus y)\to y\le x\to y$

hold for any $x,y\in L$.

Theorem 3.3. In a lattice H implication algebra *L*,

(1) $x\oplus y=x\vee y$, $x\otimes y=x\wedge y$;

(2) $x\oplus(y\otimes z)=(x\oplus y)\otimes(x\oplus z)$, $x\otimes(y\oplus z)=(x\otimes y)\oplus(x\otimes z)$;

(3) $(x\oplus y)\to z=(x\to z)\otimes(y\to z)$,
$(x\otimes y)\to z=(x\to z)\oplus(y\to z)$;

(4) $x\oplus(y\otimes z)=(x\oplus y)\otimes(x\oplus z)$, $x\oplus(x\otimes y)=x\otimes(x\oplus y)$;

(5) $x\otimes(y^{'}\oplus z)\le y^{'}\otimes(x\otimes z)$,

hold for any $x,y,z\in L$.

Theorem 3.4. Let *L* be a lattice implication algebra, and $x,y\in L$. Then $x\oplus y=\vee\{x:t\in L,x\le y\to t\}$.

Proof. By $x\to(y\to(x\oplus y))=x\to(y\to(x^{'}\to y))=x\to(x^{'}\to I)=I$, it follows that $x\le(y\to(x\oplus y))$ and hence $x\oplus y\in\{t:t\in L,x\le y\to t\}$.

On the other hand, if $z\in\{t:t\in L,x\le y\to t\}$ then $x\le y\to z$, and hence, $y\to z\le I=y\to(x\oplus y)$, i.e., $z\le x\oplus y$. It follows that $x\oplus y=\vee\{t:t\in L,x\le y\to t\}$. Ending the proof.

Definition 3.1. Let *L* be a lattice implication algebra. An *LI*-ideal *A* is a non-empty subset of *L* such that for any $x,y\in L$,

(1) $o\in A$; *(2)* $y\in A$ and $x\otimes y^{'}\in A$ imply $x\in A$.

Definition 3.2. A non-empty subset *I* of a lattice implication algebra *L* is said to be an implicative *LI*-ideals (briefly, *ILI*-ideals) of *L* if it satisfies the following conditions:

(1) $O\in I$; *(2)* $(x\to y)^{'}\otimes y^{'}\otimes z^{'}\in I$ and $z\in I$ imply $(x\to y)^{'}\in I$ for any $x,y,z\in L$.

Definition 3.3. Let *L* be a lattice implication algebra. A weak filter *A* is a non-empty subset of *L* such that for any $x,y\in L$,

$x^{'}\oplus y\in A$ implies $(x^{'}\oplus y)\oplus x^{'}\in A$.

Definition 3.4. Let *L* be a lattice implication algebra, a subset *A* of *L* is called a weak *LI*-ideals (briefly, *WLI*-ideal) of *L* if it satisfies the following condition:

$(x\to y)^{'}\in A$ implies $(x\to y)^{'}\otimes y^{'}\in A$ holds for all $x,y\in L$.

The following example shows that there exists the *WLI*-ideal in lattice implication algebra.

Example 3.1. Let *A*= *{I, O}* be a set. Now it takes $x = O, y = I$ then $(O \to I)' = O \in A$ implies $(O \to I)' \otimes I' = (O' \oplus I)' \otimes I' = I \otimes O \otimes O = O \in A$; if $x = I, y = O$ then $(I \to O)' = I \in A$ implies Example 3.1 $(I \to O)' \otimes O' = (I' \oplus O)' \otimes O' = I \otimes I \otimes I = I \in A$. Hence *A* is a *WLI*-ideal.

Example 3.2. Let *B*= *{O}* be a set. *B*= *{O}* can be check similarly.

Theorem 3.5. Let *L* be a lattice implication algebra, $A \subseteq L$ is an *LI*-ideal of *L*. Then *A* is a *WLI*-ideal of *L*.

Proof. Suppose that *A* is an *LI*-ideal of *L* and $(x \to y)' \in A$ for all $x, y \in L$. Then

$$
\begin{aligned}
(((x \to y)' \otimes y') \to (x \to y)')' &= ((x \to y)' \otimes y') \otimes (x \to y) \\
&= ((x \to y)' \otimes (x \to y)) \otimes y') \\
&= O \otimes y' \\
&= O \in A,
\end{aligned}
$$

i.e., $(((x \to y)' \otimes y') \to (x \to y)')' \in A$. Thus $(x \to y)' \otimes y' \in A$ as *A* is an *LI*-ideal and $(x \to y)' \in A$. Therefore, *A* is a *WLI*-ideal of *L*. This completes the proof.

It is easy to obtain the next conclusion.

Theorem 3.6. Let *A* and *B* are two *LI*-ideals of lattice implication algebra *L* $A \subseteq B$. If *B* is a *WLI*-ideal of *L*, then so is *A*.

Theorem 3.7. Let f: $L_1 \to L_2$ be an implication homomorphism of lattice implication algebras. If ker(f) $\neq \varnothing$, then $\ker(f)$ is a *WLI*-ideal of L_1.

Proof. Let $(x \to y)' \in \ker(f)$. Then $f((x \to y)') = o$. Hence

$$
\begin{aligned}
f((x \to y)' \otimes y') &= (f((x \to y)') \otimes f(y))' \\
&= (o \to f(y))' \\
&= I' = o,
\end{aligned}
$$

Which implies $(x \to y)' \otimes y' \in \ker(f)$. Therefore, $\ker(f)$ is a a *WLI*-ideal of L_1.

Theorem 3.8. Let $\mathfrak{A}$ is a non-empty family of *WLI*-ideal of lattice implication algebra *L*. Then $\cap\mathfrak{A}$ is also a *WLI*-ideal.

Theorem 3.9. Let *L* be a lattice implication algebra. Every *ILI*-ideal of L is a *WLI*-ideal.

Proof. Suppose that A is an *ILI*-ideal of lattice implication algebra *L*, and $(x \to y)' \in A$ for all $x, y \in L$. Since

$$(((((x \to y)' \otimes y') \otimes o') \otimes o') \to (x \to y)')'$$
$$= (((x \to y)' \otimes y') \otimes o') \otimes (x \to y)$$
$$= (((x \to y)' \otimes (x \to y)) \otimes y' \otimes o')$$
$$= (O \otimes y' \otimes I)$$
$$= (O \otimes I) \otimes y'$$
$$= O \otimes y' = O \in A.$$

Hence we obtain $(x \to y)' \otimes y' \in A$ by *A* is an *ILI*-ideal of *L*. Therefore A is a *WLI*-ideal of *L*.

Theorem 3.10. *A* is a non-empty subset of lattice implication algebra *L* and $A' = \{x' : x \in A\}$, then A' is a *WLI*-ideal of L if and only if A is a weak filter of *L*.

Proof. Assume that for any $x, y \in L$, A is a weak filter of *L* and $x' \oplus y \in A$ implies $(x' \oplus y) \oplus x' \in A$ holds. Then $(x' \oplus y)' \in A'$ implies $((x' \oplus y) \oplus x')' \in A'$, i.e., if $(y' \otimes x) \in A'$ then $((y' \otimes x) \otimes x) \in A'$. Thus A' is a *WLI*-ideal of *L*.

Conversely, let A' is a *WLI*-ideal of *L* and $(x \to y)' \in A'$ implies $((x \to y)' \otimes y') \in A'$ for all $x, y \in L$.

Since $x \to y = y' \to x' \in A$;

$(((x \to y)' \otimes y')' = (x \to y) \oplus y = (y' \to (y' \to x')) \in A$. Moreover, we get $y' \to x' \in A$ implies $(y' \to (y' \to x')) \in A$ holds. Hence *A* is a weak filter of *L*. Ending the proof.

Definition 3.5 [20] **.** A proper *LI*-ideal *B* of lattice implication algebra *L* is said to be a prime *LI*-ideal of if $x \wedge y \in B$ implies $x \in B$ or $y \in B$ for any $x, y \in L$.

Theorem 3.11. Let *A* be a prime *LI*-ideal of lattice implication algebra *L*, then A is a *WLI*-ideal of *L*.

Proof. Let *A* is a prime *LI*-ideal of lattice implication algebra *L,* for any $s, t, y \in L$. Then $(s \to y)' \wedge (t \to y)' \in A$ implies $(s \to y)' \in A$ or $(t \to y)' \in A$. Since

$$((s \wedge t) \to y)' = ((s \to y) \vee (t \to y))' = (s \to y)' \wedge (t \to y)' \in A,$$
$$(((s \wedge t) \to y)' \to y)' = (((s \to y) \vee (t \to y))' \to y)'$$
$$= (y' \to ((s \to y) \vee (t \to y)))'$$

$$= ((y' \to (s \to y)) \vee (y' \to (t \to y)))'$$
$$= ((s \to y)' \otimes y') \wedge ((t \to y)' \otimes y') \in A .$$

If $(s \to y)' \in A$ implies $((s \to y)' \otimes y') \in A$ or
$(t \to y)' \in A$ implies $((t \to y)' \otimes y') \in A$.

Then $(s \to y)' \wedge (t \to y)' \in A$ implies $((s \to y)' \otimes y') \wedge ((t \to y)' \otimes y') \in A$, i.e., $((s \wedge t) \to y)' \in A$ implies $(((s \wedge t) \to y)' \to y)' \in A$ holds. Moreover A is a *WLI*-ideal.

Theorem 3.12. Every lattice ideal in lattice H implication algebra *L* is a *WLI*-ideal of *L*.

Proof. Let *L* be a lattice H implication algebra, *A* is a lattice ideal and $(x \to y)' \in A, y \in A$ for all $x, y \in L$. For

$$y \vee (x \to y)' = y \vee (x' \vee y)' = x \vee y .$$

Hence $x \vee y \in A$. It follows that

$$y \vee ((x \to y)' \otimes y') = y \vee ((x \to y)' \to y)'$$
$$= y \vee (((x' \vee y)')' \vee y)'$$
$$= y \vee ((x' \vee y) \vee y)'$$
$$= y \vee (x \wedge y') = x \vee y .$$

So that $y \vee ((x \to y)' \otimes y') \in A$. Since

$$((x \to y)' \otimes y') \le y \vee ((x \to y)' \otimes y') .$$

Therefore $((x \to y)' \otimes y') \in A$ by *A* is lattice ideal of *L*. This completes the proof.

Corollary 3.13. Let *L* be a lattice H implication algebra, then *LI*-ideal *{o}* of *L* is *WLI*-ideal.

Theorem 3.14. Let *L* be a lattice H implication algebra, if $A(t) = [o, t]$ for all the element t of *L*. Then $A(t)$ is a *WLI*-ideal of *L*.

Proof. Suppose that $(x \to y)' \in A(t)$ for all $x, y \in L$, then $((x \to y)' \otimes t') = o$

$$((x \to y)' \otimes t') = o \Leftrightarrow ((x \to y) \vee t)' = o ,$$

i.e., $(x \to y)' \wedge t' = o$. Since

$$(((x \to y)' \otimes y') \otimes t') = (((x' \oplus y)' \otimes y') \otimes t')$$
$$= (((x' \oplus y)' \otimes t') \otimes y')$$
$$= (O \otimes y')$$
$$= O ,$$

we have $((x \to y)' \otimes y') \in A(t)$. $A(t)$ is a *WLI*-ideal of *L* by Definition 3.1.

Theorem 3.15. Let *L* be a lattice H implication algebra, if *A* is an *LI*-ideal of *L* then $A_t = \{x \in L : (x \otimes t') \in A\}$ is a *WLI*-ideal for any $t \in L$.

Proof. Assume that $(x \to y)' \in A_t$ for all $x, y \in L$, then $((x \to y)' \otimes t') \in A$.

Since
$$((((x \to y)' \otimes y') \otimes t') \to ((x \to y)' \otimes t'))'$$
$$= ((((x \to y)' \otimes y') \otimes t')' \oplus ((x \to y)' \otimes t'))'$$
$$= (((x \to y) \oplus y \oplus t) \oplus ((x \to y)' \otimes t'))'$$
$$= ((x \to y)' \otimes y' \otimes t') \otimes ((x \to y) \oplus t)$$
$$= ((((x \to y)' \otimes y') \to (x \to y)') \vee (((x \to y)' \otimes y') \to t))'$$
$$= ((((x \to y) \oplus y) \oplus (x \to y)') \vee (((x \to y) \oplus y) \oplus t))'$$
$$= ((((x \to y) \oplus (x \to y)') \oplus y) \vee (((x \to y) \oplus y) \oplus t))'$$
$$= ((I \oplus y) \vee (((x \to y) \oplus y) \oplus t))'$$
$$= (I \vee (((x \to y) \oplus y) \oplus t))'$$
$$= o \in A.$$

Note that if *A* is an *LI*-ideal of *L*, then $(((x \to y)' \otimes y') \otimes t') \in A$. Hence $((x \to y)' \otimes y') \in A_t$. Consequently, the result is valid.

Theorem 3.16. Let *L* be a lattice implication algebra, $\{A_i : i \in I\}$ is the set of *WLI*-ideal of *L* for *I* is a index set, then $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$ are *WLI*-ideals.

Proof. Let $(x \to y)' \in \bigcup_{i \in I} A_i$ for all $x, y \in L$, then there exist $i \in I$ such that $(x \to y)' \in A_i$. Since A_i is WLI-ideal, which imply that $((x \to y)' \otimes y') \in A_i$ for some $i \in I$. Hence we get $((x \to y)' \otimes y') \in \bigcup_{i \in I} A_i$. By Definition 3.1, $\bigcup_{i \in I} A_i$ is a *WLI*-ideal of *L*.

Suppose that $(x \to y)' \in \bigcap_{i \in I} A_i$ for any $x, y \in L$, then $(x \to y)' \in A_i$ for any $i \in I$. Since A_i is a *WLI*-ideal of *L*, we have $((x \to y)' \otimes y') \in A_i$ for any $i \in I$. Thus $((x \to y)' \otimes y') \in \bigcap_{i \in I} A_i$. Therefore $\bigcap_{i \in I} A_i$ is a *WLI*-ideal.

Remark: Let *L* be a lattice implication algebra, the intersection of *WLI*-ideals of *L* is also a *WLI*-ideal by Theorem 3.9 Suppose $A \subseteq L$, the maximal *WLI*-ideal containing *A* is called the *WLI*-ideal generated by *A* and denoted by $L < A >$.

Definition 3.6. Let L be a lattice implication algebra, a *WLI*-ideal is called a maximal *WLI*-ideal if it is not L, and it is a maximal element of the set of all *WLI*-ideals with respect to set inclusion.

In what follows, for any $a \in L$,

$$L_a^1 = \{((x \to y)' \otimes y') : x, y \in L, (x \to y)' = a\};$$
$$L_a^2 = \{((x \to y)' \otimes y') : x, y \in L, (x \to y)' \in L_a^1\};$$
$$L_a^3 = \{((x \to y)' \otimes y') : x, y \in L, (x \to y)' \in L_a^2\};$$
$$L_a^4 = \{((x \to y)' \otimes y') : x, y \in L, (x \to y)' \in L_a^3\};$$
$$\vdots \qquad \vdots$$
$$L_a^n = \{((x \to y)' \otimes y') : x, y \in L, (x \to y)' \in L_a^{n-1}\}.$$

It is easy to check

$$((x \to y)' \otimes y') = (((x \to y)' \otimes y') \to o)';$$
$$(((x \to y)' \otimes y') \otimes y') \le ((x \to y)' \otimes y').$$

Hence $L_a^n \subseteq L_a^{n-1} \cdots \subseteq L_a^4 \subseteq L_a^3 \subseteq L_a^2 \subseteq L_a^1$ and denoted by $T_a = \bigcap_{i=1}^{\infty} L_a^i$.

Theorem 3.17. Let f: $L \to \{O, I\}$ be an implication epimorphism of lattice implication algebras. then ker(f) is a maximal proper *WLI*-ideal of L.

Theorem 3.18. Let L be a lattice implication algebra, then T_a is a *WLI*-ideal for any $a \in L$.

Proof. Suppose that $(x \to y)' \in T_a$ for any $x, y \in L$, then there exists $i \ge 1$ and it is the element of the set of $\{0,1,2,3\ldots,\ldots\}$ such that $(x \to y)' \in L_a^i$. Hence $((x \to y)' \otimes y') \in L_a^{i+1}$, i.e., $((x \to y)' \otimes y') \in T_a$. Therefore, T_a is a *WLI*-ideal of L by Definition 3.1.

Theorem 3.19. Let L be a lattice implication algebra, $x, a \in L$, then $x \in T_a$ if and only if there exist $k \in N^+$, x_k, $x_{k-1}, \ldots, x_2, x_1 \in L$, and y_k, $y_{k-1}, \ldots$, $y_2, y_1 \in L$, if it satisfies the follows conditions:

(1) $(x \to y)' = a$;

(2) $(x_i \to y_i)' \in L_a^{i-1}$ and $(x_i \to y_i)' = ((x_{i-1} \to y_{i-1})' \otimes y_{i-1}')$;

(3) $((x_k \to y_k)' \otimes y_k') = x$.

Proof. Assume that conditions hold. Obviously, $x \in T_a$.

Let $x \in T_a$, then there exist $k \in N^+$ such that $x \in L_a^k$ by $T_a = \bigcap_{i=1}^{\infty} L_a^i$, i.e., $\exists x_k, y_k \in L$ such that $x = ((x_k \to y_k)' \otimes y_k')$. Thus we have $(x \to y)' \in L_a^{k-1}$. Since there exist $x_{k-1}, y_{k-1} \in L$ such that $(x_k \to y_k)' = ((x_{k-1} \to y_{k-1})' \otimes y_{k-1}')$ for $x_k \to y_k \in L_a^{k-1}$, and so we get $x_{k-1} \to y_{k-1} \in L_a^{k-2}$. It follows that we can be obtain sequences x_k, $x_{k-1}, \ldots,$ $x_2, x_1 \in L$, and y_k, $y_{k-1}, \ldots,$ $y_2, y_1 \in L$ such that three conditions hold. Ending the proof.

Theorem 3.20. Let L be a lattice implication algebra, then $T_a = < a >$ for any $a \in L$.

Proof. Suppose that $a \in T_a$ then $< a > \subseteq T_a$ by Theorem 3.9. On the other hand, let $a \in T_a$ then there exist $k \in N^+$ such that x_k, x_{k-1}, $x_2, x_1 \in L$, and y_k, y_{k-1}, $y_2, y_1 \in L$ satisfy the following conditions:

(1) $(x \to y)' = a$;
(2) $(x_i \to y_i)' \in L_a^{i-1}$ and $(x_i \to y_i)' = ((x_{i-1} \to y_{i-1})' \otimes y_{i-1}')$ (i= 2,3…);
(3) $((x_k \to y_k)' \otimes y_k') = x$.

Moreover, we have $(x_i \to y_i)' \in < a >$ (i= 1, 2, 3,..., k), i.e., $T_a \subseteq < a >$. Consequently, the result is valid.

Theorem 3. 21. Let *L* be a lattice implication algebra, $A \subseteq L$. Then $< A > = \bigcap_{a \in A} < a >$.

Proof. Since $a \in < a >$ for all $a \in A$, we have $A \subseteq \bigcap_{a \in A} < a >$. Thus $< A > \subseteq \bigcap_{a \in A} < a >$. On the other hand, if $\forall a \in A$ then $< a > \subseteq < A >$. Hence we obtain $\bigcap_{a \in A} < a > \subseteq < A >$. Thus we have $< A > = \bigcap_{a \in A} < a >$.

4 Conclusion

In order to provide a logical foundation for the fuzziness and the incomparability in uncertain information processing reasoning, Xu initiated the notion of lattice implication algebras. And for development of non-classical logical system, it is needed to make clear the structure of lattice implication algebra. It is well known that the ideals with special properties play an important role in the structure of the logic system. The aim of this article is to introduce the concept of *WLI*-ideal in lattice implication algebra, investigate the related properties. And research the relationships of *WLI*-ideal and *LI*-ideal maximal *WLI*-ideal. We believe the research for on the properties of ideals of lattice implication algebra will help to the research of logical system with propositional value.

Acknowledgments. The authors are highly grateful to the referees for their valuable suggestions and comments.

Remark: This work was supported by the National Science Foundation of P.R. China under Grant No. 60474022, and the specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20060613007.

References

1. D.W. Borns, J.M. Mack: An algebraic Introduction on Mathematical logic, Springer, Berlin(1975).
2. L. Bolc, P. Borowik: Many-valued Logics. Springer-Verlag (1992).
3. R. Ben-Eliyahu, R. Dechter: Default reasoning using classical logic, Artificial Intelligence, 84, 113-150(1996).
4. J. A. Goguen: The logic of inexact concepts. Synthese, 19, 325-373 (1969).
5. J. Pavelka: On fuzzy logic Ⅰ,Ⅱ,Ⅲ, Zeitschr. F. Math. Logic and Grundlagend. Math., 25, 45-52, 119-134, 447-464(1979).
6. V. Novak: First-order fuzzy logic. Studia Logica, 46(1),87-109 (1982).
7. Y. Xu: Lattice implication algebras, J. Southwest Jiaotong Univ. 28(1), 20-27(1993) (in Chinese).
8. Y. Xu, D. Ruan, K. Y. Qin, J. Liu: Lattice-Valued Logic, Springer, Berlin(2003).
9. J. Liu, Y. Xu: On filters and structures of lattice implication algebras, Chinese Science Bulletin, Vol.42, No.10, 1049-1052(May 1997) (in Chinese).
10. Y. Xu, K.Y. Qin, J. Liu, Song Z.M.: L-valued propositional logic L_{vpl}, Information Sci. 114, 205-235(1999).
11. Y. Xu, E.E. Kerre, D. Ruan, Z.M. Song: Fuzzy reasoning based on the extension principle, Int. J. Intelligent System, 16(4) 469-495 (2001).
12. Y. Xu, D. Ruan, E.E. Kerre, J. Liu: α-Resolution principle based on first-order lattice-valued logic LF(X), Information Sci., 132, 221-239(2001).
13. Y.B. Jun , E.H. Roh, Y. Xu: LI-ideals in lattice implication algebras. Bull. Korean Math. Soc., 35, 13-24 (1998).
14. Y.B. Jun: On LI-ideals and prime LI-ideals of lattice implication algebras. Bull. Korean Math. Soc., 36(2) 369-380(1999).
15. Y.B. Jun, Y. Xu, K.Y. Qin, E.H. Roh: Intuitionistic fuzzy LI-ideals in lattice implication algebras. Sci. Math., 3(2), 185-192(2000).
16. Y.B. Jun, Y. Xu: Fuzzy LI-ideals in lattice implication algebras. J. Fuzzy Math., 7(4), 997-1003(1999).
17. Y.B. Jun, Y. Xu , K.Y. Qin : Normal fuzzy LI-ideals in lattice implication algebras. J. Fuzzy Math., 7(4), 1045-1052(1999).

18. Y.L. Liu et al.: ILI-ideals and prime LI-ideals in lattice implication algebras, Int. J. Information Sci., 155, 157-175(2003).
19. T.M. Huang, Y. Xu et al. : Application and Introductions of Lattice and Order, Southwest Jiaotong University Press, Chengdu (1998).
20. Y.B. Jun : On *LI*-ideals and prime *LI*-ideals of lattice implication algebras, Bull. Korean Math. Soc. 36(2), 369-380(1999).

Congruence Relations Induced by Filters and LI-Ideals*

Zhiyan Chang and Yang Xu

Department of Mathematics, Southwest Jiaotong University, Chengdu,
610031, P.R. China
changzhiyanyantai@163.com, xuyang@home.swjtu.edu.cn

Abstract. Two equivalent conditions for LI-ideals are given. The lattice implication quotient algebras induced by obstinate filters and LI-ideals are studied, that is, the lattice implication quotient algebras L/F induced by obstinate filters is $\{[0]_F, [1]_F\}$. It is also concluded that there is a bijection between $L(L,A) = \{J : A \subseteq J \subseteq L,\ J$ is a LI-ideal if $L\}$ and the LI-ideals of L/A (the lattice implication quotient algebras induced by LI-ideal A).

1 Introduction

In order to provide a logical foundation for uncertain information processing theory, especially for the fuzziness, the incomparability in uncertain information in the reasoning, Xu [1] proposed the concept of lattice implication algebra, and discussed some of their properties in [1, 6]. Filters and LI-ideals are important algebraic structures of lattice implication algebra and were studied in [2, 3, 4]. Song and Xu [3, 5] research the congruence relation in lattice implication algebra and the relationship between congruence relation and filters, and congruence relation induced by LI-ideals. This paper is an extension of above mentioned work. In section 3, two equivalent conditions for LI-ideals are given. In section 4, we prove that $L/F = \{[0]_F, [1]_F\}$ when F is an obstinate filter. In section 5, we prove that there is a bijection between $L(L,A)$ and the LI-ideals of L/A .

2 Preliminaries

Definition 2.1[1] Let $(L, \vee, \wedge, \neg)$ be a complemented lattice with the universal bounds 0, I, and

$$\rightarrow: \mathrm{L} \times \mathrm{L} \rightarrow \mathrm{L}$$

be a mapping. $(L, \vee, \wedge, \neg, \rightarrow)$ is called a lattice implication algebra if the following conditions hold for any $x, y, z \in L$:

* This work is supported by National Natural Science Foundation of China (Grant No.60474022), and the Special Research Funding to Doctoral Subject of Higher Education Institutes in P.R. China (Grant No. 20060613007).

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 349–357, 2007.
springerlink.com

(1) $x \to (y \to z) = y \to (x \to z)$;
(2 $x \to x = I$;
(3) $x \to y = y' \to x'$;
(4) $x \to y = y \to x = I \Rightarrow x = y$;
(5) $(x \to y) \to y = (y \to x) \to x$;
(6) $(x \vee y) \to z = (x \to z) \wedge (y \to z)$;
(7) $(x \wedge y) \to z = (x \to z) \vee (y \to z)$.

A lattice implication algebra L is called a lattice H implication algebra if it satisfies

$$x \vee y \vee ((x \wedge y) \to z) = I \text{ for all } x, y, z \in L.$$

Theorem 2.1[3] Let L be a lattice implication algebra, then for any $x, y \in L$,

(1) $x \to 0 = x'$, $I \to x = x$
(2) $x \leq y$ if and only if $x \to y = I$
(3) $(x \to y) \to y = x \vee y$, $x \wedge y = ((x \to y) \to x')'$
(4) $(x \to y) \vee (y \to x) = I$
(5) $x \to y \geq x' \vee y$

In a lattice implication algebra L, [2] define binary operations $\otimes$ and $\oplus$ as follows: for any $x, y \in L$

$$x \otimes y = (x \to y')';$$
$$x \oplus y = x' \to y.$$

For more details of lattice implication algebras we refer the readers to [3].

Definition 2.2[2] Let L be a lattice implication algebra. A non-empty subset J of L is called a filter of L if it satisfies the following conditions:

(1) $I \in J$;
(2) for any $x, y \in L$, if $x \in J$ and $x \to y \in J$ imply $y \in J$.

$J \subseteq L$ is said to be an implicative filter of L, if it satisfies the following conditions:

(1) $I \in J$;
(2) for any $x, y, z \in L$, if $x \to (y \to z) \in J$ and $x \to y \in J$ imply $x \to z \in J$.

Definition 2.3[7] Let L be a lattice implication algebra, J a proper filter of L. J is called an obstinate filter if $x \notin J$, $y \notin J$ imply $x \to y \in J$, $y \to x \in J$ for any $x, y \in L$.

Definition 2.4[3] Let L be a lattice implication algebra, P a proper filter of L. P is called a prime filter if $a \vee b \in P$ implies $a \in P$ or $b \in P$ for any $a, b \in L$.

Denote $F(L)$ as the set of all filters on lattice implication algebra L.

Definition 2.5[4] Let L be a lattice implication algebra. A non-empty subset A of L is called an LI-ideal of L if it satisfies:

(1) $0 \in A$;
(2) $(x \to y)' \in A$ and $y \in A$ imply $x \in A$.

Definition 2.6[5] Let L be a lattice implication algebra. $\equiv_\theta \subseteq L \times L$ is said to be a congruence relation on L, if $\equiv_\theta$ satisfies: for any $x, y, z \in L$,

(1) $x \equiv_\theta x$;
(2) $x \equiv_\theta y$ implies $y \equiv_\theta x$;
(3) $x \equiv_\theta y$, $y \equiv_\theta z$ implies $x \equiv_\theta z$;
(4) $x \equiv_\theta y$ implies $x \to z \equiv_\theta y \to z$.

Definition 2.6 shows that $\equiv_\theta$ is an equivalence relation on L, and is compatible w.r.t. $\to$. Theorem 2.2 shows that $\equiv_\theta$ is compatible w.r.t. $\vee$, $\wedge$, $'$. In fact $\equiv_\theta$ is compatible w.r.t. $\otimes$ and $\oplus$ also.

Theorem 2.2[5] If $\equiv_\theta$ is a congruence relation on a lattice implication algebra L, then for any $x, y \in L$,

(1) $x \equiv_\theta y$ if and only if $x' \equiv_\theta y'$;
(2) $x \equiv_\theta y$ if and only if $x \vee z \equiv_\theta y \vee z$ for any $z \in L$;
(3) $x \equiv_\theta y$ if and only if $x \wedge z \equiv_\theta y \wedge z$ for any
.0 $z \in L$.

Theorem 2.3. If $\equiv_\theta$ is a congruence relation on a lattice implication algebra L, then for any $x, y \in L$,
(1) $x \equiv_\theta y$ if and only if $x \otimes z \equiv_\theta y \otimes z$ for any $z \in L$;
(2) $x \equiv_\theta y$ if and only if $x \oplus z \equiv_\theta y \oplus z$ for any $z \in L$.

The proof is clear.

3 Two Equivalent Definitions of LI-Ideals

Theorem 3.1[3] Let L be a lattice implication algebra. Every LI-ideal A is a lattice ideal.

Theorem 3.2[8] Let L be a lattice implication algebra, A an LI-ideal if L. Then A is closed w.r.t. $\vee, \wedge, \oplus, \otimes$, i.e. for any $x, y \in A$, we have $x \vee y$, $x \wedge y$, $x \otimes y$, $x \oplus y \in A$.

Theorem 3.3. Let L be a lattice implication algebra. A non-empty subset A of L is an LI-ideal of L if and only if it satisfies:

(1) $0 \in A$;
(2) If $x, y \in A$, then $x \oplus y \in A$;
(3) $\forall x, y \in L$, if $x \in A$ and $y \leq x$ then $y \in A$.

Proof. Suppose $\forall x, y \in L$, $(x \to y)' \in A$ and $y \in A$.
$(x \to y)' \oplus y = (x \to y) \to y = x \vee y \in A$, and $x \leq x \vee y$, then $x \in A$.
The converse is clear by theorem3.1, 3.2. □

Theorem 3.4. Let L be a lattice implication algebra. A non-empty subset A of L is an LI-ideal of L if and only if it satisfies:

(1) $0 \in A$;
(2) $\forall x, y, z \in L$, if $(x \to y)' \in A$, $(y \to z)' \in A$, then $(x \to z)' \in A$.

Proof. $\forall x, y, z \in L$, if $(x \to y)' \in A$, $(y \to z)' \in A$,

$$\left\{\left[(x \to z)' \to (x \to y)'\right]' \to (y \to z)'\right\}'$$
$$= \left\{(y \to z) \to \left[(x \to y) \to (x \to z)\right]\right\}'$$
$$= \left\{(x \to y) \to \left[(y \to z) \to (x \to z)\right]\right\}'$$
$$= \left\{(x \to y) \to \left[x \to (y \vee z)\right]\right\}'$$
$$= I' = 0 \in A$$

A is an LI-ideal if L, and $(y \to z)' \in A$, so $\left[(x \to z)' \to (x \to y)'\right]' \in A$, and $(x \to y)' \in A$, then $(x \to z)' \in A$.

Conversely, $\forall x, y \in L$, if $(x \to y)' \in A$ and $y \in A$.

$$(x \to y)' \in A,\ y = (y')' = (y \to 0)' \in A, \text{ so } x \in A.$$ □

4 Congruence Relations Induced by Filters

Given a congruence relation $\equiv_\theta$ on lattice implication algebra L, we can construct another lattice implication algebra by $\equiv_\theta$. For any $x \in L$, denote

$$[x]_\theta = \{a \mid a \in L, x \equiv_\theta a\},$$

which is the congruence class of x w.r.t. the congruence relation $\equiv_\theta$. In the following, we sometimes write $[x]_\theta$ as $[x]$ for short. Denote

$$L/\equiv_\theta = \{[x] \mid x \in L\},$$

which is the set of congruence classes of L w.r.t. $\equiv_\theta$ [3].

Theorem 4.1[5] Let L be a lattice implication algebra, and $\equiv_\theta$ a congruence relation on it. In $L/\equiv_\theta$, for any $x, y \in L$,

$$[x] \to [y] \triangleq [x \to y],$$
$$[x] \cup [y] \triangleq [x \vee y],$$
$$[x] \cap [y] \triangleq [x \wedge y],$$
$$[x]' \triangleq [x'],$$
$$[x] \prec [y] \Leftrightarrow [x] \cup [y] = [y]$$

Then $(L/\equiv_\theta, \cap, \cup, ', \to)$ is a lattice implication algebra.

The lattice implication algebra $L/\equiv_\theta$ obtained from Theorem 4.1 is said to be a lattice implication quotient algebra of L w.r.t. $\equiv_\theta$. Define a mapping

$$\pi : L \to L/\equiv_\theta$$

Satisfying: for any $x \in L$,

$$\pi(x) = [x]_\theta.$$

Then π is a natural lattice implication homomorphism.

Denote $CR(L)$ as the set of all congruence relations on lattice implication algebra L.

Lemma 4.1 (1)[3]. For any $\equiv_\theta \in CR(L)$, $[I]_\theta = \{x \in L \mid x \equiv_\theta I\} \in F(L)$;

(2)[6]. For any $F \in F(L)$, $\equiv_F$ is defined as follows: for any $x, y \in L$,

$$x \equiv_F y \ \text{ iff } \ x \to y,\ y \to x \in F.$$

Then $\equiv_F \in CR(L)$, $[I]_F = F$.

In what follows $(L/\equiv_F, \cap, \cup, ', \to)$ is called the lattice implication algebra induced by filter F and denoted by $(L/F, \cap, \cup, ', \to)$.

Lemma 4.2 (1). For any $\equiv_F \in CR(L)$, $[0]_F = \{x' \mid x \in F\} = F'$;

(2). $[0]_F$ is an LI-ideal and it is the smallest element of L/F.

Proof. (1) $x \equiv_F 0$ iff $x \to 0$, $0 \to x \in F$ iff $x' \in F$.

So $[0]_F = \{x \in L | x' \in F\} = \{x' | x \in F\} = F'$.

The proof of (2) is obvious. □

Theorem 4.2[3] Let L be a lattice implication algebra and F a filter of L. Then F is an implicative filter if and only if the lattice implication quotient algebra L/F is a lattice H implication algebra.

Theorem 4.3[3] Let L be a lattice implication algebra and F a proper filter of L. Then F is a prime filter if and only if the lattice implication quotient algebra L/F is totally ordered.

Theorem 4.4. Let L be a lattice implication algebra and F a proper filter of L. Then F is an obstinate filter if and only if the lattice implication quotient algebra $L/F = \{[0]_F, [1]_F\}$.

Proof. Suppose $L/F = \{[0]_F, [1]_F\}$ and for any $x, y \notin F$, where $x, y \in L$. Since $[I]_F = F$, $[I]_F \cap [0]_F = \varnothing$ then $x, y \in [0]_F$. It follows that $x \equiv_F y$, i.e., for any $x, y \notin F$, implies $x \to y \in F$, $y \to x \in F$.

Conversely, suppose F is an obstinate filter. For any $x, y \notin F$, then $x \to y \in F$, $y \to x \in F$, i.e., $x \equiv_F y$. Since $0 \notin F$, then $x \equiv_F 0$, for any $x \notin F$. $[I]_F = F$, hence $[I]_F \cup [0]_F = L$, and since $[I]_F \cap [0]_F = \varnothing$, we have $L/F = \{[0]_F, [1]_F\}$. □

5 Congruence Relations Induced by LI-Ideals

Let A be an LI-ideal of lattice implication algebra L. Define a binary relation $\equiv_A$ on L as follows: for any $x, y \in L$,

$$x \equiv_A y \text{ iff } (x \to y)' \in A,\ (y \to x)' \in A.$$

Lemma 5.1[3] $\equiv_A \in CR(L)$.

Let

$$[x]_A = \{y \in L | x \equiv_A y\},\ L/A = \{[x]_A | x \in L\}.$$

It is clear that $[0]_A = A$ and $[I]_A = \{y \in L | y' \in A\}$. Define binary operation $\cup, \cap, \to$ and an unary operation $'$ on L/A as follows:

$$[x]_A \to [y]_A \triangleq [x \to y]_A,$$
$$[x]_A \cup [y]_A \triangleq [x \vee y]_A,$$

$$[x]_A \cap [y]_A \triangleq [x \wedge y]_A ,$$
$$[x]_A{}' \triangleq [x']_A ,$$
$$[x]_A \prec [y]_A \triangleq [x]_A \cup [y]_A = [y]_A ,$$

for any $[x]_A$, $[y]_A \in L/A$. It can be easily verified that $(L/A, \cap, \cup, ', \to)$ is a lattice implication algebra, which is called the lattice implication quotient algebra of L induced by the LI-ideal A [3].

Let L be a lattice implication algebra, A An LI-ideal of L and L/A the lattice implication quotient algebra induced by A ,and

$$J/A = \{[x]_A \mid x \in J\} .$$

Lemma 5.2. Let
$L(L, A) = \{ J : A \subseteq J \subseteq L,\ J$ is a LI-ideal if $L\}$.
Then for any $J \in L(L, A)$, J/A is an LI-ideal of L/A.

Proof. $[0]_A$ is an LI-ideal and the smallest element of L/A. By J is an LI-ideal of L, $0 \in J$, and then we have $[0]_A \in J/A$.

If $([x]_A \to [y]_A)' \in J/A$, $[y]_A \in J/A$, i.e., $\left[(x \to y)'\right]_A \in J/A$, $[y]_A \in J/A$, then there exist $x_1, x_2 \in J$ such that $[x_1]_A = \left[(x \to y)'\right]_A$, $[x_2]_A = [y]_A$. So $x_2 \equiv_A y$ and $(y \to x_2)' \in A$. It follows that $(y \to x_2)' \in J$. Since $x_2 \in J$ and J is an LI-ideal of L, We have $y \in J$. Similarly, we can prove $(x \to y)' \in J$. Hence $x \in J$, $[x]_F \in J/A$.

Therefore, J/A is an LI-ideal of L/A. □

Lemma 5.3. If T is an LI-ideal of L/A ,

$$J = \{x \mid x \in L, [x]_A \in T\}$$

Then $A \subseteq J \subseteq L$ and J is an LI-ideal of L and $J/A = T$.

Proof. If T is an LI-ideal of L/A, then $0 \in J$ and $A \subseteq J \subseteq L$, by $[x]_A = [0]_A$ for any $x \in A$.

If $(x \to y)' \in J$, $y \in J$ then $\left[(x \to y)'\right]_A \in T$, $[y]_A \in T$, i.e., $([x]_A \to [y]_A)' \in T$, $[y]_A \in T$. Hence $[x]_A \in T$ and $x \in J$. So J is an LI-ideal of L. $J/A = T$ is clear.

□

Lemma 5.4. $J_1 \neq J_2$ if and only if $J_1/A \neq J_2/A$ for any $J_1, J_2 \in L(L, A)$.

Proof. Suppose $J_1, J_2 \in L(L, A)$. If $J_1/A \neq J_2/A$, then $J_1 \neq J_2$ is clear.

If $J_1/A = J_2/A$, but $J_1 \neq J_2$, let $x_1 \in J_1$, $x_1 \notin J_2$, then, by $[x_1]_A \in J_1/A = J_2/A$, there exist $x_2 \in J_2$ such that $[x_1]_A = [x_2]_A$, hence $(x_1 \to x_2)' \in A$. It follows that $(x_1 \to x_2)' \in J_2$.

Since $x_2 \in J_2$ and J_2 is an LI-ideal of L, we have $x_1 \in J_2$. This contradiction shows that the conclusion holds. □

Form Lemma 5.2-5.4 we can obtain the following corollaries.

Corollary 5.1. Let A be an LI-ideal of L, and $\equiv_A$ be the congruence relation of L defined in Lemma 5.1. Define a mapping

$$\text{G: } L(L, A) \to \text{the LI-ideals of } L/A$$

satisfying

$$G(J) = J/A,$$

then G is a bijection.

Corollary 5.2. Let L be a lattice implication algebra, $A \subseteq L$ and $A \neq \varnothing$.
Then $L(L, A) = \{ J : A \subseteq J \subseteq L,\ J \text{ is a LI-ideal if } L\}$ is the set of all LI-ideals on the lattice implication algebra L iff $A = \{0\}$.

Proof. If $L(L, A) = \{ J : A \subseteq J \subseteq L,\ J \text{ is a LI-ideal if } L\}$ is the set of all LI-ideals on the lattice implication algebra L, then for any LI-ideal B of L, $A \subseteq B$. Specially, for $B = \{0\}$, so $A = \{0\}$.

Conversely, it is trivial. □

6 Conclusions

In this paper, we prove that when the filter is "big" enough (here big means the number of elements of filter), the quotient algebras L/F induced by obstinate filters is $\{[0]_F, [1]_F\}$. It is also concluded that there is a bijection between $L(L, A)$ and the LI-ideals of L/A, this bijection helps us to study the property of another when we know one's.

References

1. Y. Xu: Lattice implication algebra. J. Southwest Jiaotong Univ. (1993), Sum. 28(1), 20-27.
2. Y. Xu, K.Y. Qin: On filters of lattice implication algebras. J. Southwest Jiaotong Univ, (1993), 1(2): 251-260.
3. Y. Xu, D.Ruan, K.Y.Qin, J. Liu: Lattice-Valued Logic-An Alternative Approach to Treat Fuzziness and Incomparability. Springer-Verlag, Berlin. (2003).
4. Y.B. Jun, E.H. Roh, Y. Xu: LI-ideal in lattice implication algebra. J. Bull Korean Math Soc, (1998), 35(1): 13-24.
5. Z.M. Song, Y. Xu: Congruence relations on lattice implication algebras. J. Math Appl. (1997), 10(3): 121-124.
6. Y. Xu: Homomorphisms in lattice implication algebras. In: Proc. Of 5th Many-valued Logical Congress of China, Nanjing, China, (1992).
7. X.F. Wang, Y. Xu, Z.M. Song: Some properties of filters in lattice implication algebras. J. Southwest Jiaotong Univ. (2001), 36(5): 536-539.
8. S.W. Chen: Study on Uncertainty Reasoning in Lattice-Valued First-order Logic L_{vfl} Based on Lattice Implication Algebra. Ph.D. Thesis, Southwest Jiaotong Univ. China (2005).

Weak Completeness of Resolution in a Linguistic Truth-Valued Propositional Logic

Yang Xu[1], Shuwei Chen[2], Jun Liu[3], and Da Ruan[4]

[1] Department of Mathematics, Southwest Jiaotong University,
Chengdu, Sichuan 610031, P.R. China
xuyang@home.swjtu.edu.cn
[2] School of Electrical Engineering, Zhengzhou University,
Zhengzhou, Henan 450001, P.R. China
swchen@zzu.edu.cn
[3] School of Computing and Mathematics, Faculty of Engineering,
University of Ulster at Jordanstown, Newtownabbey BT37 0QB, Northern Ireland, UK
j.liu@ulster.ac.uk
[4] The Belgian Nuclear Research Centre (SCK•CEN), Boeretang 200, B-2400 Mol, Belgium
druan@sckcen.be

Abstract. In the present paper, the weak completeness of α-resolution principle for a lattice-valued logic $(L_n \times L_2)P(X)$ with truth value in a logical algebra − lattice implication algebra $L_n \times L_2$, is established. Accordingly, the weak completeness of (Exactly, True)-resolution principle for a linguistic truth-valued propositional logic ℓ based on the linguistic truth-valued lattice implication algebra L-LIA is derived.

1 Introduction

One of the fundamental goals of artificial intelligence (AI) is to build artificially computer-based systems to make computer simulate human's intelligent activities in which reasoning plays an essential role [10, 14]. On the one hand, due to the fact that there are various kinds of uncertainties in the real world, human reasoning is always the reasoning under uncertainties [8, 14]. On the other hand, many intelligent activities of human being are associated with and achieved using natural language. Therefore, it is necessary to investigate natural language based reasoning under uncertainty within the realm of AI [6, 16]. Especially, during the investigation of AI from the logical point of view, it is necessary to investigate the logic systems with linguistic truth values, and furthermore, to investigate reasoning under uncertainty based on linguistic truth-valued logic systems [16].

Although there have been many investigations on the algebraic structure of linguistic truth values [2, 3, 5, 9, 11], logic systems based on linguistic truth values [4, 15, 16], and uncertain reasoning and automated reasoning based on linguistic truth-valued logic systems as well [1, 9, 12, 16], among others, one of the key, substantial and also essential problems is not solved yet. That is, how to choose a comparatively appropriate linguistic truth-valued algebraic structure, which can provide with a

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 358–366, 2007.
springerlink.com

comparatively appropriate interpretation for the logical formulae in linguistic truth value logic systems, and accordingly provide a strict theoretical foundation, as well as a convenient, practical, and effective underlying semantic structures to automated uncertain reasoning based on linguistic truth-valued logic, and various kinds of corresponding intelligent information processing systems. This kind of comparatively appropriate linguistic truth-valued algebraic structures is not established yet so far. It is worthy to state that, as expected, this kind of algebraic structure should satisfy the following assumptions: (1) endow with some kind of well-defined logical algebraic structure; (2) linguistic truth values adopted should be consistent with the meaning of commonly used natural language; (3) linguistic truth values adopted have apparent distinguishability; and (4) the set of linguistic truth value should be a modestly small set of linguistic truth values which can cover commonly used natural linguistic expressions. We have made some explorations [1, 5, 9, 16] to establish this kind of appropriate linguistic truth-valued algebraic structure on the basis of research work about a logical algebra – lattice implication algebra and the corresponding lattice-valued logic systems and reasoning methods [14], but the further research is still ongoing. The present work is one of further steps within this whole research framework.

In [13, 14], we have provided the weak completeness theorem of α-resolution principle for lattice-valued propositional logic LP(X) based on lattice implication algebra L. That is,

Theorem 1. [13] (Weak completeness theorem of α-resolution principle for LP(X)) Let S be a regular generalized conjunctive normal form in LP(X), $\alpha \in L$ a dual molecule and $\vee_{\alpha \in L}(\alpha \wedge \alpha') \le \alpha < I$. Suppose that there exists $\beta \in L$ such that $\beta \wedge (\beta \to \beta') \not\le \alpha$. If $S \le \alpha$, then there exists an α-resolution deduction from S to α-□.

According to Theorem 1, we can get the weak completeness of an α-resolution deduction if we can select $\alpha \in L$ which satisfies the following conditions:

1. C_1. α is a dual molecule;
2. C_2. $\vee_{\alpha \in L}(\alpha \wedge \alpha') \le \alpha < I$;
3. C_3. there exists $\beta \in L$ such that $\beta \wedge (\beta \to \beta') \not\le \alpha$.

In [16], a set of linguistic modifiers AD={Slightly (Sl for short), Somewhat (So), Rather (Ra), Almost (Al), Exactly (Ex), Quite (Qu), Very (Ve), Highly (Hi), Absolutely (Ab)} was defined, and denoted as L_9. The lattice implication algebra defined on the chain Sl<So<Ra<Al<Ex<Qu<Ve<Hi<Ab is called a *lattice implication algebra with modifiers* if its implication is Lukasiewicz implication. We also defined a set of meta truth values MT={True (Tr), False (Fa)}, and denoted it as L_2. The lattice implication algebra (of course a Boolean algebra) defined on the set of meta truth values is called a *meta linguistic truth-valued lattice implication algebra*, where Fa<Tr. The product lattice implication algebra AD×MT (in short $L_9 \times L_2$) is a linguistic truth-valued lattice implication algebra, denote as L-LIA and the lattice-valued logic system with truth-value in L-LIA is denoted as ℓ.

In fact, how many modifiers should be taken is an important issue, but beyond the scope of this paper, will be further investigated in another work. Let L_n: $a_1 \le a_2 \le \ldots \le a_n$,

L_2: $b_1 \le b_2$ be Lukasiewicz lattice implication algebra chains with n and 2 elements respectively. L= $L_n \times L_2$ is the product lattice implication algebra of L_n and L_2 [16]. Focusing on the lattice implication algebra $L_n \times L_2$, in this paper, we firstly introduce the method for selecting the above kind of α in $L_n \times L_2$, then establish the weak completeness of α-resolution principle for lattice-valued propositional logic $(L_n \times L_2)P(X)$ based on $L_n \times L_2$, based on this result, the weak completeness of α-resolution principle for linguistic truth-valued propositional logic ℓ based on linguistic truth-valued lattice implication algebra L-LIA [16] is derived finally.

The remaining of the paper is organized as follows: in Section 2, the resolution level set in lattice-valued propositional logic $(L_n \times L_2)P(X)$ is introduced and provided. In Section 3, the concept of the mp-parallel points of α in $(L_n \times L_2)$ is introduced and the method to determine the set of them is provided. In Section 4, the concept of an appropriate resolution level set in $(L_n \times L_2)P(X)$ and how to determine it is discussed. The weak completeness of α-resolution for $(L_n \times L_2)P(X)$ and the weak completeness of α-resolution for linguistic truth-valued propositional logic ℓ are presented respectively in Section 5. The paper is concluded in Section 6.

2 Resolution Level Set

We firstly address how to determine $\alpha \in L$ which satisfies conditions C_1 and C_2.

Definition 1. Let L be a lattice implication algebra. The set

$$\begin{aligned}\Re &= \{ \alpha \mid \alpha \in L \text{ satisfies } C_1 \text{ and } C_2 \} \\ &= \{ \alpha \mid \alpha \in L,\ \alpha \text{ is a dual molecule, } \vee_{\alpha \in L}(\alpha \wedge \alpha') \le \alpha < I \}\end{aligned}$$

is called *a resolution level set of* LP(X).

2.1 Determination of Dual Molecule

According to Theorem 5 in [16], the set of dual molecules in $L = L_n \times L_2$ is

$$\begin{aligned}&\{(a_k, b_t) \mid (a_k, b_t) \in L_n \times L_2 \text{ satisfies } C_1\} \\ &= \{(a_k, b_t) \mid (k=1, \ldots, n,\ t=2) \text{ or } (k=n,\ t=1)\}\ .\end{aligned}$$

2.2 Solve $\vee_{\alpha \in L}(\alpha \wedge \alpha')$ in $L = L_n \times L_2$

Note that

$$(a_k, b_t) \wedge (a_k, b_t)' = (a_k, b_t) \wedge (a_{n-k+1}, b_{3-t}) = (a_{k \wedge (n-k+1)}, b_{t \wedge (3-t)}) = (a_{k \wedge (n-k+1)}, b_1).$$

So,

$$(a_k, b_t) \wedge (a_k, b_t)' = \begin{cases} (a_k, b_1), \text{ when } k \le (n+1)/2; \\ (a_{n-k+1}, b_1), \text{ when } k \ge (n+1)/2. \end{cases}$$

Therefore, $\vee_{\alpha \in L}(\alpha \wedge \alpha') = \vee_{k \le (n+1)/2}(a_k, b_1) \vee \vee_{k \ge (n+1)/2}(a_{n-k+1}, b_1)$

$$= (a_{(n+1)/2}, b_1) \vee (a_{(n+1)/2}, b_1) = (a_{(n+1)/2}, b_1)\,.$$

2.3 Resolution Level Set of $(L_n \times L_2)P(X)$

It follows from 2.1 and 2.2 that the resolution level set $\Re$ of $(L_n \times L_2)P(X)$ is

$$\Re = \{(a_k, b_t) \mid (a_k, b_t) \in L_n \times L_2 \text{ satisfies } C_1 \text{ and } C_2\}$$
$$= \{(a_k, b_t) \mid (n \neq k \geq (n+1)/2,\ t=2) \text{ or } (k=n,\ t=1)\}.$$

3 Set of mp-Parallel Points

Definition 2. Let L be a lattice implication algebra, $\alpha \in Q \subseteq L$. The set

$$S(\alpha) = \{\beta \mid \beta \in L, \beta \wedge (\beta \to \beta') \not\leq \alpha\}$$

is called *a set of mp-parallel points of* α in Q.

It is clear that α satisfies C_3 if and only if $S(\alpha) \neq \emptyset$.

In $L = L_n \times L_2$, let $(a_k, b_t) \in \Re$. Then we need to determine the following set:

$$S\left((a_k, b_t)\right) = \{\beta \mid \beta \in L_n \times L_2, \beta \wedge (\beta \to \beta') \not\leq (a_k, b_t)\}.$$

According to a result in [16], i.e., if $y = (a_k, b_t) \in L_n \times L_2$, then

$$S\left((a_k, b_t)\right) = \left\{(a_i, b_j) \mid k < i \leq (2n+1)/3,\ j = 1, 2\right\}$$
$$\cup \left\{(a_i, b_j) \mid (2n+1-k)/2 > i > (2n+1)/3,\ j = 1, 2\right\}.$$

Hence, for $(a_k, b_t) \in \Re$, we have:

When t=1, there must be k=n, and therefore,

$$S\left((a_n, b_1)\right) = \left\{(a_i, b_j) \mid n < i \leq (2n+1)/3,\ j = 1, 2\right\}$$
$$\cup \left\{(a_i, b_j) \mid (2n+1-n)/2 > i > (2n+1)/3,\ j = 1, 2\right\}$$
$$= \left\{(a_i, b_j) \mid (2n+1-n)/2 > i > (2n+1)/3,\ j = 1, 2\right\}$$
$$= \emptyset.$$

When t=2, there must be $n \neq k \geq (n+1)/2$, and therefore,

$$S\left((a_k, b_2)\right) = \left\{(a_i, b_j) \mid k < i \leq (2n+1)/3,\ j = 1, 2\right\}$$
$$\cup \left\{(a_i, b_j) \mid (2n+1-k)/2 > i > (2n+1)/3,\ j = 1, 2\right\}.$$

Corollary 1. In $L = L_n \times L_2$, for any positive integer k which satisfies $n \neq k \geq (n+1)/2$, the following conclusion holds:

$$(a_i, b_j) \in S\left((a_k, b_2)\right) \text{ iff } i \text{ satisfies } k < i < (2n+1-k)/2.$$

Proof. Firstly notice that:

(1). $k<(2n+1)/3$ iff $(2n+1)/3<(2n+1-k)/2$.
(2). $k<(2n+1)/3$ iff $k<(2n+1-k)/2$.

From the above analysis, it follows that

$(a_i,b_j)\in S((a_k,b_2))$ iff i satisfies $k<i\leq(2n+1)/3$ or $(2n+1-k)/2>i>(2n+1)/3$.

If $(a_i,b_j)\in S((a_k,b_2))$, then i satisfies $k<i\leq(2n+1)/3$ or $(2n+1-k)/2>i>(2n+1)/3$. So, it follows from (1) that, i satisfies $k<i<(2n+1-k)/2$.

On the contrary, if i satisfies $k<i<(2n+1-k)/2$, then, it follows from (1) and (2) that, i satisfies $k<i\leq(2n+1)/3$ or $(2n+1-k)/2>i>(2n+1)/3$, therefore $(a_i,b_j)\in S((a_k,b_2))$.

Consequently, $(a_i,b_j)\in S((a_k,b_2))$ if and only if i satisfies $k<i<(2n+1-k)/2$.

Corollary 1 shows that (a_k, b_2) satisfies C_3 if and only if there exists i satisfies $k<i<(2n+1-k)/2$ in $L_n\times L_2$.

Theorem 2. In $L=L_n\times L_2$, $S((a_k, b_2))\neq\emptyset$ iff $k<(2n-2)/3$.

Proof. $S((a_k,b_2))=\{(a_i,b_j)\mid k<i\leq(2n+1)/3, j=1,2\}$
$$\cup\{(a_i,b_j)\mid(2n+1-k)/2>i>(2n+1)/3, j=1,2\}.$$

(1). $\{(a_i,b_j)\mid k<i\leq(2n+1)/3, j=1,2\}\neq\emptyset$ iff there exists a positive integer i which satisfies $k<i\leq(2n+1)/3$ iff $k+1\leq(2n+1)/3$ iff $2\leq 2n-3k$ iff $k<(2n-2)/3$.

(2). $\{(a_i,b_j)\mid(2n+1-k)/2>i>(2n+1)/3, j=1,2\}\neq\emptyset$ iff there exists a positive integer i which satisfies $(2n+1-k)/2>i>(2n+1)/3$ iff $(2n+1-k)/2>[(2n+1)/3]+1$ iff $k<2n-1-2[(2n+1)/3]$, where $[(2n+1)/3]$ is the integer part of $(2n+1)/3$.

Given a summary of (1) and (2), it follows that

$S((a_k, b_2))\neq\emptyset$ iff $k<(2n-2)/3$ or $k<2n-1-2[(2n+1)/3]$.

(3). (i) When $n=2$:

$(2n-2)/3=2/3$, so $\max\{k\mid k<(2n-2)/3\}=0$;
$2n-1-2[(2n+1)/3]=1$, and so, $\max\{k\mid k<2n-1-2[(2n+1)/3]\}=0$.

(ii) When $n=3t$ ($t\geq1$, and is an integer):

$(2n-2)/3=2t-(2/3)=2t-1+(1/3)$, and so, $\max\{k\mid k<(2n-2)/3\}=2t-1$;
$2n-1-2[(2n+1)/3]=2t-1$, so $\max\{k\mid k<2n-1-2[(2n+1)/3]\}=2t-2$.

(iii) When $n=3t+1$ ($t\geq1$, and is an integer):

$(2n-2)/3=2t$, so $\max\{k\mid k<(2n-2)/3\}=2t-1$;
$2n-1-2[(2n+1)/3]=2t-1$, so $\max\{k\mid k<2n-1-2[(2n+1)/3]\}=2t-2$.

(iv) When $n=3t+2$ ($t\geq1$, and is an integer):

$(2n-2)/3=2t+(2/3)$, so $\max\{k\mid k<(2n-2)/3\}=2t$;
$2n-1-2[(2n+1)/3]=2t+1$, so $\max\{k\mid k<2n-1-2[(2n+1)/3]\}=2t$.

Summing up (i)-(iv), for any positive integer k, $k<(2n-2)/3$ or $k<2n-1-2[(2n+1)/3]$ iff $k<(2n-2)/3$.

Theorem 2 shows that (a_k, b_2) satisfies C_3 if and only if $k<(2n-2)/3$ in $L_n\times L_2$.

4 Appropriate Resolution Level Set

Definition 3. Let L be a lattice implication algebra. The set

$$\mathcal{P}\mathfrak{R} = \{ \alpha \mid \alpha \in \mathfrak{R}, S(\alpha) \neq \emptyset \}$$

is called an *appropriate resolution level set* of the lattice-valued propositional logic LP(X) based on lattice implication algebra L.

In fact, $\mathcal{P}\mathfrak{R} = \{ \alpha \mid \alpha \in \mathrm{L}, \alpha$ satisfies C_1, C_2, and $\mathrm{C}_3\} = \{ \alpha \mid \alpha \in \mathfrak{R}, \alpha$ satisfies $\mathrm{C}_3\}$.

In $\mathrm{L}=\mathrm{L}_n\times\mathrm{L}_2$, $\mathcal{P}\mathfrak{R} = \{(a_k, b_2) \mid n\neq k\geq(n+1)/2, S((a_k, b_2))\neq\emptyset\}$
$= \{(a_k, b_2) \mid (n+1)/2\leq k<(2n\text{-}2)/3\}$.

In fact, from Theorem 5, $S((a_k, b_2))\neq\emptyset$ iff $k<(2n\text{-}2)/3$, and $(2n\text{-}2)/3<n$. So,

$$\{(a_k, b_2) \mid n\neq k\geq(n+1)/2, S((a_k, b_2))\neq\emptyset\}=\{(a_k, b_2) \mid (n+1)/2\leq k<(2n\text{-}2)/3\} .$$

Therefore, (a_k, b_2) satisfies C_1, C_2, and C_3 if and only if $(n+1)/2\leq k<(2n\text{-}2)/3$ in $\mathrm{L}_n\times\mathrm{L}_2$.

Corollary 2. In $\mathrm{L}=\mathrm{L}_n\times\mathrm{L}_2$, let m be the number of elements of the appropriate resolution level set $\mathcal{P}\mathfrak{R}$ of $(\mathrm{L}_n\times\mathrm{L}_2)\mathrm{P(X)}$.

(1). If n is odd, then $m=[(n\text{-}1)/6]$.
(2). If n is even, then $m=[(n\text{-}4)/6]$.

Proof. (1) If n is odd, notice that for $(a_k, b_2)\in\mathcal{P}\mathfrak{R}$, k satisfies $(n+1)/2\leq k\leq(2n\text{-}1)/3$, then $m=[(2n\text{-}1)/3\text{-}(n+1)/2+1]=[(n\text{-}1)/6]$.

(2) If n is even, notice that for $(a_k, b_2)\in\mathcal{P}\mathfrak{R}$, k satisfies $(n+1)/2+1/2\leq k\leq(2n\text{-}1)/3$, then $m=[(2n\text{-}1)/3\text{-}((n+1)/2+1/2)+1]=[(n\text{-}4)/6]$.

Theorem 3. The appropriate resolution level set $\mathcal{P}\mathfrak{R}\neq\emptyset$ of $(\mathrm{L}_n\times\mathrm{L}_2)\mathrm{P(X)}$ iff $n\geq7$, but $n\neq8$.

Proof. It follows from Theorem 2 that,

$\{(a_k, b_2) \mid n\neq k\geq(n+1)/2, S((a_k, b_2))\neq\emptyset\}\neq\emptyset$ iff there exists a positive integer r such that $r\neq n$, and $(n+1)/2\leq r\leq 2(n\text{-}1)/3$. Hence,

(1) if n is odd, notice that $(n+1)/2$ is a positive integer, then there exists a positive integer r such that $r\neq n$, and $(n+1)/2\leq r\leq 2(n\text{-}1)/3$ iff $(n+1)/2\leq 2(n\text{-}1)/3$ iff $n\geq7$;

(2) if n is even, notice that $((n+1)/2)+(1/2)=(n+2)/2$ is a positive integer, then there exists a positive integer r such that $r\neq n$, and $(n+1)/2\leq r\leq 2(n\text{-}1)/3$ iff $(n+2)/2\leq 2(n\text{-}1)/3$ iff $n\geq10$.

Summing up (1) and (2), the conclusion follows.

Theorem 3 shows that there exists $\alpha\in\mathrm{L}_n\times\mathrm{L}_2$ satisfying C_1, C_2, and C_3 if and only if $n\geq7$, but $n\neq8$.

Lemma 1. In $\mathrm{L}=\mathrm{L}_n\times\mathrm{L}_2$, for any $(a_k, b_2)\in\mathrm{L}_n\times\mathrm{L}_2$ and $(a_r, b_2)\in\mathrm{L}_n\times\mathrm{L}_2$, if $k\leq r$, then $S((a_r, b_2))\subseteq S((a_k, b_2))$.

Theorem 4. In $\mathrm{L}=\mathrm{L}_n\times\mathrm{L}_2$, its appropriate resolution level set $\mathcal{P}\mathfrak{R}\neq\emptyset$ of $(\mathrm{L}_n\times\mathrm{L}_2)\mathrm{P(X)}$ if and only if

(1). $S((a_{(n+1)/2}, b_2))\neq\emptyset$, if n is odd;
(2). $S((a_{(n+2)/2}, b_2))\neq\emptyset$, if n is even.

Proof. If the appropriate resolution level set $\mathcal{P}\mathfrak{R}\neq\emptyset$ of $(L_n\times L_2)P(X)$, then it follows from Theorem 3 that, there exists a positive integer r such that $(n+1)/2\leq r\leq 2(n-1)/3$, $S((a_r, b_2))\neq\emptyset$ iff (1) if n is odd, then $n\geq 7$ iff $S((a_{(n+1)/2}, b_2))\neq\emptyset$ (since $r=(n+1)/2$ satisfies $(n+1)/2\leq r\leq 2(n-1)/3$) ; (2) if n is even, then $n\geq 10$ iff $S((a_{(n+2)/2}, b_2))\neq\emptyset$ (since $r=(n+2)/2$ satisfies $(n+1)/2\leq r\leq 2(n-1)/3$) .

1. Theorem 4 shows that there exists $\alpha\in L_n\times L_2$ satisfying C_1, C_2, and C_3 if and only if $(a_{(n+1)/2}, b_2)$ satisfies C_3 if n is odd; and $(a_{(n+2)/2}, b_2)$ satisfies C_3 if n is even.
From Theorem 3 and Theorem 4, we have

Corollary 3. In $L=L_n\times L_2$, $n\geq 7$, but $n\neq 8$ iff

(1). if n is odd, then $S((a_{(n+1)/2}, b_2))\neq\emptyset$;
(2). if n is even, then $S((a_{(n+2)/2}, b_2))\neq\emptyset$.

1. Corollary 3 shows that in $L_n\times L_2$, $n\geq 7$, but $n\neq 8$ if and only if $(a_{(n+1)/2}, b_2)$ satisfies C_3 if n is odd; and $(a_{(n+2)/2}, b_2)$ satisfies C_3 if n is even.

5 Weak Completeness of α-Resolution

Based on the weak completeness theorem of α–resolution principle in Theorem 1 and notice that $\mathcal{P}\mathfrak{R}=\{(a_k, b_2) \mid (n+1)/2\leq k<(2n-2)/3\}$, the following theorem holds.

Theorem 5. (Weak Completeness I of α-Resolution for $(L_n\times L_2)P(X)$). In $L=L_n\times L_2$, for any (a_k, b_2), if $(n+1)/2\leq k<2(n-1)/3$, then the weak completeness theorem of α-resolution principle holds for $\alpha=(a_k, b_2)$.

From Theorem 2, Theorem 3, $\mathcal{P}\mathfrak{R}=\{(a_k, b_2) \mid (n+1)/2\leq k<(2n-2)/3\}$, and the weak completeness theorem of α-resolution principle in Theorem 1, the following theorem holds.

Theorem 6. (Weak Completeness II of α-Resolution for $(L_n\times L_2)P(X)$). In $L=L_n\times L_2$, if $n\geq 7$, but $n\neq 8$, then there exists k, which satisfies $(n+1)/2\leq k<2(n-2)/3$, such that the weak completeness theorem of α-resolution principle holds for $\alpha=(a_k, b_2)$.

Example. In $L_9\times L_2=L_{18}$, $S((a_5, b_2))=\{(a_6, b_1), (a_6, b_2)\}$, and $S((a_6, b_2))=\emptyset$, so the appropriate resolution level set $\mathcal{P}\mathfrak{R}=\{(a_5, b_2)\}$ of $(L_9\times L_2)P(X)$.

Therefore, based on the above theorems and example, the following theorem holds.

Theorem 7. (Weak Completeness of α-Resolution for ℓ). In ℓ, the weak completeness theorem of α-resolution principle holds for α=(Exactly, True).

Theorem 7 shows it holds for the weak completeness theorem of (Exactly, True)-resolution principle in $(L_9\times L_2)P(X)$.

6 Conclusion

In this paper, the weak completeness of α-resolution principle were presented for lattice-valued propositional logic $(L_n \times L_2)P(X)$ based on a product of two Lukasiewicz lattice implication algebras L_n and L_2. Specially, it then inferred the weak completeness of (Exactly, True)-resolution principle for linguistic truth-valued propositional logic ℓ based on linguistic truth-valued lattice implication algebra L-LIA. All these results will provide a support for further investigating resolution-based automated reasoning method on linguistic truth-valued logic system, which has also been stated in this paper as a necessary and worthy direction in intelligent information process in AI realm.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of P.R. China (Grant No. 60474022), the China-Flanders bilateral scientific cooperation joint project (Grant No. 011S1105), and the Special Research Funding to Doctoral Subject of Higher Education Institutes in P.R. China (Grant No. 20060613007).

References

1. Chen, S.W., Xu, Y., Ma, J.: A Linguistic Truth-Valued Uncertainty Reasoning Model Based on Lattice-Valued Logic. Lecture Notes in Artificial Intelligence 3613 (2005) 276–284
2. Herrera, F., Martínez, L.: A 2-Tuple Fuzzy Linguistic Representation Model for Computing With Words. IEEE Trans. Fuzzy Systems 8 (6) (2000) 746–752
3. Ho, N.C., Wechler, W.: Hedge Algebras: An Algebraic Approach to Structure of Sets of Linguistic Truth Values. Fuzzy Sets and Systems 35 (1990) 281–293
4. Ho, N.C., Wechler, W.: Extended Hedge Algebras and Their Application to Fuzzy Logic. Fuzzy Sets and Systems 52 (1992) 259–281
5. Liu, J., Xu, Y., Ruan, D., Martinez, L.: A Lattice-Valued Linguistic-Based Decision-Making Method. In: Proc. of 2005 IEEE International Conference on Granular Computing (2005) 199–202
6. Ma, J., Chen, S., Xu, Y.: Fuzzy Logic from the Viewpoint of Machine Intelligence. Fuzzy Sets and Systems 157 (2006) 628 – 634
7. Novák, V., Perfilieva, I., Močkoř, J.: Mathematical Principles of Fuzzy Logic. Kluwer, Dordrecht (1999)
8. Novák, V.: Are Fuzzy Sets a Reasonable Tool for Modelling Vague Phenomena. Fuzzy Sets and Systems 156 (3) (2005) 341–348
9. Pei, Z., Xu, Y.: Lattice Implication Algebra Model of a Kind of Linguistic Terms and Its Inference. In: Proc. of the 6th International FLINS Conference (2004) 93–98
10. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice-Hall, Person Education (1995)
11. Truksen, I.B., Kandel, A., Zhang, Y.Q.: Universal Truth Tables and Normal Forms. IEEE Trans. Fuzzy Systems 6 (1998) 295–303

12. Truksen, I.B.: Computing with Descriptive and Verisic Words. In: Proc. of NAFIP'99 (2004) 13–17
13. Xu, Y., Ruan, D., Kerre, E.E., Liu, J.: α -Resolution Principle Based on Lattice-Valued Propositional Logic LP(X). Information Sciences 130 (2000) 195–223
14. Xu, Y., Ruan, D., Qin, K.Y., Liu, J.: Lattice-Valued Logic: An Alternative Approach to Treat Fuzziness and Incomparability. Springer-Verlag, Berlin Heidelberg New York (2003)
15. Xu, Y., Liu, J., Ruan, D., Lee, T.T.: On the Consistency of Rule Bases Based on Lattice-Valued First-Order Logic LF(X). Int. J. Intelligent Systems 21 (2006) 399–424
16. Xu, Y., Chen, S.W., Ma, J.: Linguistic Truth-Valued Lattice Implication Algebra and Its Properties. In: Proc. CESA’06 (2006)

Part X

Fuzzy Trees

Decision-Based Questionnaire Systems

Masoud Nikravesh

EECS Department, University of California, Berkeley
Nikravesh@eecs.berkeley.edu

Abstract. In this paper, we develop a decision-based questionnaire system. Toward this end, we use evolutionary computation techniques. Initially, we work on a first order aggregation model and performed its learning using genetic algorithms, in which these preferences will be represented by a weighting vector associated with the variables involved in the aggregation process. In this model tree nodes represent aggregators, terminals or leaves correspond to variables, and weight values are added to the children branches for each aggregator. The parameters characterizing this multi-aggregation model are aggregators, weights and their combination in form of a tree structure. In this case, the learning process has to find the optimal combination of these parameters based on training data. In this learning process, the evolution principle remains the same as in a conventional GP but the DNA encoding needs to be defined according to the considered problem

1 Introduction

Decision-based questionnaire systems and support systems, searching database records and ranking the results based on multi-criteria queries is central for many database applications used within organizations in finance, business, industrial and other fields. Most of the available systems' 'software' are modeled using crisp logic and queries, which result in rigid systems with imprecise and subjective processes and results. In this chapter we introduce fuzzy querying and ranking as a flexible tool allowing approximation where the selected objects do not need to match exactly the decision criteria resembling natural human behavior. The model consists of five major modules: the Fuzzy Search Engine (FSE), the Application Templates (AT), the User Interface (UI), the Database (DB) and the Evolutionary Computing (EC). We developed the software with many essential features. It is built as a web-based software system that users can access and use over the Internet. The sys-tem is designed to be generic so that it can run different application domains. To this end, the Application Template module provides information of a specific application as attributes and properties, and serves as a guideline structure for building a new application. The Fuzzy Search Engine (FSE) is the core module of the system. It has been developed to be generic so that it would fit any application. The main FSE component is the query structure, which utilizes membership functions, similarity functions and aggregators. Through the user interface a user can enter and save his profile, input criteria for a new query, run different queries and display results. The user can manually eliminate the results he disapproves of or change the ranking according to his preferences. The Evolutionary Computing (EC) module

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 369–379, 2007.
springerlink.com

monitors ranking preferences of the user's queries. It learns to adjust to the intended meaning of the users' preferences.

2 Measure of Association and Fuzzy Similarity

As in crisp query and ranking, an important concept in fuzzy query and ranking applications is the measure of association or similarity between two objects in consideration. For example, in a fuzzy query application, a measure of similarity between two queries and a document, or between two documents, provides a basis for determining the optimal response from the system. In fuzzy ranking applications, a measure of similarity between a new object and a known preferred (or non-preferred) object can be used to define the relative goodness of the new object. Most of the measures of fuzzy association and similarity are simply extensions from their crisp counterparts. However, because of the use of perception based and fuzzy information, the computation in the fuzzy domain can be more powerful and more complex.

Various definitions of similarity exist in the classical, crisp domain, and many of them can be easily extended to the fuzzy domain. However, unlike in the crisp case, in the fuzzy case the similarity is defined on two fuzzy sets. Suppose we have two fuzzy sets *A* and *B* with membership functions $\mu_A(x)$ and $\mu_B(x)$, respectively. The arithmetic operators involved in the fuzzy similarity measures can be treated using their usual definitions while the union and the intersection operators need to be treated specially. It is important for these operator pairs to have the following properties: (1) conservation, (2) monotonicity, (3) commutativity, and (4) associativity. It can be verified that the triangular norm (T-norm) and triangular co-norm (T-conorm) (Nikravesh, 2001a; Bonissone and Decker, 1986; Mizumoto, 1989; Fagin, 1998 and 1999) conform to these properties and can be applied here. A detailed survey of some commonly used T-norm and T-conorm pairs along with other aggregation operators can be find at Nikravesh and Azvine (2002).

In many situations, the controlling parameters, including the similarity metric, the type of T-norm/conorm, the type of aggregation operator and associated weights, can all be specified based on the domain knowledge of a particular application. However, in some other cases, it may be difficult to specify a priori an optimal set of parameters. In those cases, various machine learning methods can be employed to automatically "discover" a suitable set of parameters using a supervised or unsupervised approach. For example, the Genetic Algorithm (GA) and DNA-based computing, as described in later sections, can be quite effective.

3 Evolutionary Computing

In the Evolutionary Computing (EC) module of the BISC Decision Support System, our purpose is to use an evolutionary method to allow automatic adjusting of the user's preferences. These preferences can be seen as parameters of the fuzzy logic model in form of weighting of the used variables. These preferences are then represented by a weight vector and genetic algorithms will be used to fix them.

In the Evolutionary Computation approach, Genetic Programming, which is an extension of Genetic Algorithms, is the closest technique to our purpose. It allows us to learn a tree structure, which represents the combination of aggregators. The selection of these aggregators is included to the learning process using the Genetic Programming.

In this section, we describe the GA (Holland, 1992) and GP (Koza, 1992) application to our problem. Our aim is learning fuzzy-DSS parameters which are the weight vectors representing the user preferences associated to the variables that have to be aggregated on the one hand, and the adequate decision tree representing the combination of the aggregation operators that have to be used on the other hand.

Weight vector being a linear structure, can be represented by a binary string in which weight values are converted to binary numbers. This binary string corresponds to the individual's DNA in the GA learning process. The goal is to find the optimal weighting of the variables. A general GA module can be used by defining a specific fitness function for each application as shown in Figure 1.

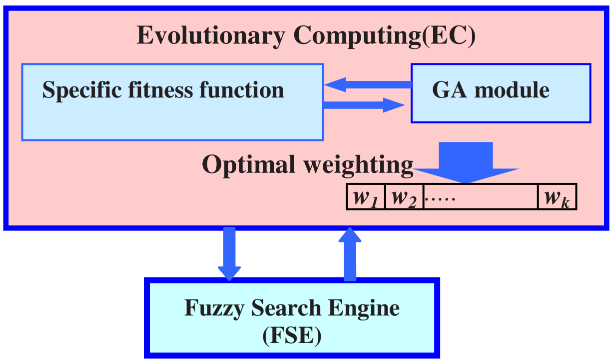


Fig. 1. Evolutionary Computing Module: preferences learning

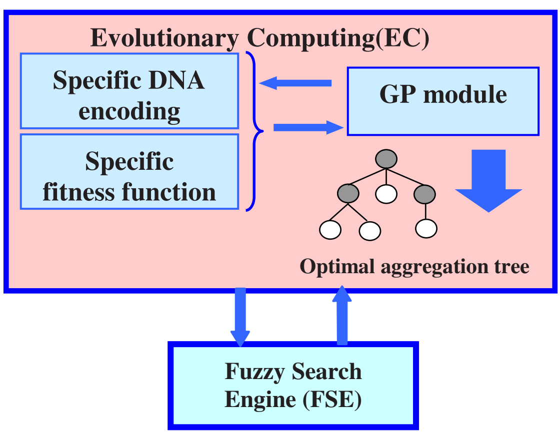


Fig. 2. Evolutionary Computing Module: aggregation tree learning

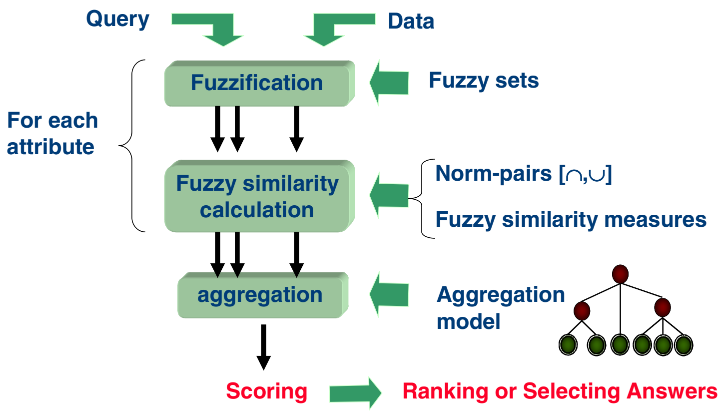


Fig. 3. Multi-Criteria Decision Model -- aggregation tree learning

Aggregators can be combined in the form of a tree structure which can be built using a Genetic Programming learning module (Figures 2 and 3). It consists in evolving a population of individuals represented by tree structures. The evolution principle remains the same as in a conventional GP module but the DNA encoding needs to be defined according to the considered problem. We propose to define an encoding for aggregation trees which is more complex than for classical trees and which is common to all considered applications. As shown in Figure 3, we need to define a specific encoding, in addition to the fitness function specification.

Tree structures are generated randomly as in the conventional GP. But, since these trees are augmented according the properties defined above, the generation process has to be updated. So, we decided to randomly generate the number of arguments when choosing an aggregator as a node in the tree structure. And for the weights, we chose to generate them randomly for each node during its creation.

Concerning the fitness function, it is based on performing the aggregation operation and the root node of the tree that has to be evaluated. For the university admissions application, the result of the root execution corresponds to the score that has to be computed for each value vector in the training data set. The fitness function, as in the GA learning of the user preferences, consists in simple or combined similarity measures. In addition, we can include to the fitness function a complementary measure that represents the individual's size which has to be minimized in order to avoid over-sized trees.

We have described the use of evolutionary computation methods for optimization problems in the BISC decision support system. It is an original idea in combining fuzzy logic, machine learning and evolutionary computation. Another important and unique component of our system is compactification algorithm or Z-Compact [Nikraveh 2005, Zadeh 1976, and Zadeh and Nikravesh 2002]. The fuzzified model of Z-Compact algorithm developed by Nikravesh based on original idea by Lotfi A. Zadeh [Zadeh 1976] and it has been implemented for the first time as part of BISC-DSS for automa-tons multi-agents modeling as part of ONR project [Zadeh and Nikravesh 2002]. The algorithm has been extended to handle fuzzy and linguistic variables and currently is part of the BISC-DSS software and it has been applied in several applications [Nikravesh 2005 and Zadeh and Nikravesh 2002]. Details of step by step of this algorithm can be find [Zadeh and Nikravesh 2002 and Nikravesh 2005].

4 Compactification Algorithm

Let's consider a set of attributes Atrb1, Atrb2,…,AtrbN, each Atrbj defined over a discrete domain . We consider a representation space as the cartesian product of the attribute domains over these attributes, and its mapping into a set of K classes represented by class labels C1, C2, …, CK. The goal is to build a rule-based classifier in which each class is represented by a set of rules, and these rules are the result of the compactification of the original training data for each class. One can use Z(n)-compact algorithm to represent the Data-Attribute-Class (DAC) matrix with rule-base model. Table 2 shows the intermediate results based on Z(n)-compact algorithm for concept 1. Table 3 shows how the original DAC (Table 1) matrix is represented in final pass with a maximally compact representation. Once the DAC matrix represented by a maximally compact representation (Table 3), one can translate this compact representation with rules as presented in Tables 4 and 5. Table 6 shows the Z(n)-compact algorithm.

Table 1. Data-Attribute-Class (DAC) Matrix

Data	$Atrb_1$	$Atrb_2$	$Atrb_3$		C_{class}
D_1	a_1^1	a_1^2	a_1^3		c_1
D_2	a_1^1	a_2^2	a_1^3		c_1
D_3	a_2^1	a_2^2	a_1^3		c_1
D_4	a_3^1	a_2^2	a_1^3		c_1
D_5	a_3^1	a_1^2	a_2^3		c_1
D_6	a_1^1	a_2^2	a_2^3		c_1
D_7	a_2^1	a_2^2	a_2^3		c_1
D_8	a_3^1	a_2^2	a_2^3		c_1
D_9	a_3^1	a_1^2	a_1^3		c_2
D_{10}	a_1^1	a_1^2	a_2^3		c_2
D_{11}	a_2^1	a_1^2	a_1^3		c_2
D_{12}	a_2^1	a_1^2	a_2^3		c_2

As it has been proposed, the DAC entries could not be crisp numbers. The following cases would be possible:

A -- The basis for the kiijjs are [0 and 1]. This is the simplest case and Z(n)-compact will work as presented

B -- The basis for the kiijjs are frequency/probability or any similar statistical based values. In this case, we use fuzzy granulation to granulate into series of granular, two ([0 or 1] or [high and low]), three (i.e. low, me-dium, and high), etc. Then the Z(n)-compact will work as it is presented.

Table 2. Intermediate results for Z(n)-compact algorithm

Data	$Atrb_1$	$Atrb_2$	$Atrb_3$		C_1
D_1	a_1^1	a_1^2	a_1^3		c_1
D_2	a_1^1	a_2^2	a_1^3		c_1
D_3	a_2^1	a_2^2	a_1^3		c_1
D_4	a_3^1	a_2^2	a_1^3		c_1
D_5	a_3^1	a_1^2	a_2^3		c_1
D_6	a_1^1	a_2^2	a_2^3		c_1
D_7	a_2^1	a_2^2	a_2^3		c_1
D_8	a_3^1	a_2^2	a_2^3		c_1
D_2, D_3, D_4	*	a_2^2	a_1^3		c_1
D_6, D_7, D_8	*	a_2^2	a_2^3		c_1
D_5, D_8	a_3^1	*	a_2^3		c_1
D_2, D_6	a_1^1	a_2^2	*		c_1
D_3, D_7	a_2^1	a_2^2	*		c_1
D_4, D_8	a_3^1	a_2^2	*		c_1
D_2, D_6	*	a_2^2	*		c_1
$D_2, D_3, D_4, D_6, D_7, D_8$					

Table 3. Maximally Z(n)-compact representation of DAC matrix

Data	$Atrb_1$	$Atrb_2$		$Atrb_n$	Class
D_1	a_1^1	a_1^2		a_1^3	c_1
D_5, D_8	a_3^1	*		a_2^3	c_1
$D_2, D_3, D_4, D_6, D_7, D_8$	*	a_2^2		*	c_1
D_9	a_2^1	a_1^2		a_1^3	c_2
D_{10}	a_1^1	a_1^2		a_2^3	c_2
D_9, D_{11}	a_2^1	a_1^2		*	c_2

C -- The basis for the kiijjs are set value created based on set theory which can be created based on traditional statistical based methods, human made, or fuzzy-set. In this case, the first step is to find the similarities between set values using statistical or fuzzy similarly measures. BISC-DSS software has a set of similarity measures, T-norm and T-conorm, and aggregator operators for this purpose. The second step is to use fuzzy granulation to granulate the similarities values into series of granular, two ([0 or 1] or [high and low]), three (i.e. low, medium, and high), etc. Then the Z(n)-compact will work as it is presented.

D -- It is also important to note that classes may also not be crisp. There-fore, steps B and C could also be used to granulate concepts as it is used to granulate the key-word entries values (kiijjs).

Table 4. Rule-based representation of Z(n)-compact of DAC matrix

Data	Rules
D_1	If $Atrb_1$ is a_1^1 and $Atrb_2$ is a_1^2 and $Atrb_3$ is a_1^3 THEN Class is c_1
D_5,D_8	If $Atrb_1$ is a_3^1 and $Atrb_3$ is a_2^3 THEN Class is c_1
D_2,D_3, D_4,D_6,D_7, D_8	If $Atrb_2$ is a_2^2 THEN Class is c_1
D_9	If $Atrb_1$ is a_2^1 and $Atrb_2$ is a_1^2 and $Atrb_3$ is a_1^3 THEN Class is c_2
D_{10}	If $Atrb_1$ is a_1^1 and $Atrb_2$ is a_1^2 and $Atrb_3$ is a_2^3 THEN Class is c_2
D_9,D_{11}	If $Atrb_1$ is a_2^1 and $Atrb_2$ is a_1^2 THEN Class is c_2

Table 5. Rule-based representation of Maximally Z(n)-compact of DAC matrix (Alternative representation for Table 4)

Data	Rules
D_1	If $Atrb_1$ is a_1^1 and $Atrb_2$ is a_1^2 and $Atrb_3$ is a_1^3 OR If $Atrb_1$ is a_3^1 and $Atrb_3$ is a_2^3 OR If $Atrb_2$ is a_2^2 THEN Class is c_1
D_5,D_8	
D_2,D_3, D_4,D_6,D_7, D_8	
D_9	If $Atrb_1$ is a_2^1 and $Atrb_2$ is a_1^2 and $Atrb_3$ is a_1^3 OR
D_{10}	If $Atrb_1$ is a_1^1 and $Atrb_2$ is a_1^2 and $Atrb_3$ is a_2^3 OR Concept is c_2
D_9,D_{11}	If $Atrb_1$ is a_2^1 and $Atrb_2$ is a_1^2 THEN Concept is c_2

E -- Another important case is how to select the right attributes in first place. One can use traditional statistical or probabilistic techniques, non traditional techniques such as GA-GP, fuzzy-logic or clustering techniques. These techniques will be used as first pass to select the first set of initial attributes. The second step will be based on feature selection technique based on to maximally separating the classes. This techniques are currently part of the BISC-DSS toolbox, which includes the Z(n)-Compact-Feature-Selection technique (Z(n)-CFS)).

Table 6. Z(n)-Compactification Algorithm

Z(n)-Compact Algorithm:
The following steps are performed successively for each column j; j=1 … n

1. Starting with k_{ii}^{jj} (ii=1, jj=1) check if for any k_{ii}^{1} (ii=1, ..,3 in this case) all the columns are the same, then k_{ii}^{1} can be replaced by *
   - For example, we can replace k_{ii}^{1} (ii=1, …,3) with * in rows 2, 3, and 4 . One can also replace k_{ii}^{1} with * in rows 6, 7, and 8. (**Table 1,** first pass).
2. Starting with k_{ii}^{jj} (ii=1, jj=2) check if for any k_{ii}^{2} (ii=1, ..,3 in this case) all the columns are the same, then k_{ii}^{2} can be replaced by *
   - For example, we can replace k_{ii}^{2} (ii=1, …,3) with * in rows 5 and 8. (**Table 1,** first pass).
3. Repeat steps one and 2 for all jj.
4. Repeat steps 1 through 3 on new rows created Row* (Pass 1 to Pass nn, in this case, Pass 1 to Pass 3).
   - For example, on Rows* 2,3, 4 (Pass 1), check if any of the rows given columns jj can be replaced by *. In this case, k_{ii}^{3} can be replaced by *. This will gives: * k_2^{2} *.
   - For example, on Pass 3, check if any of the rows given columns jj can be replaced by *. In this case, k_{ii}^{1} can be replaced by *. This will gives: * k_2^{2} *
5. Repeat steps 1 through 4 until no compactification would be possible

F -- Another very important case is when the cases are not possibilistic and are in general form " IF Atrbi isri Ai and … Then Class isrc Cj ; where isr can be presented as:

r: = *equality constraint: X=R is abbreviation of X is=R*
r: ≤ *inequality constraint: X ≤ R*
r:⊂ *subsethood constraint: X ⊂ R*
r: blank *possibilistic constraint; X is R; R is the possibility distribution of X*
r: v *veristic constraint; X isv R; R is the verity distribution of X*
r: p *probabilistic constraint; X isp R; R is the probability distribution of X*
r: rs *random set constraint; X isrs R; R is the set-valued probability distribution of X*
r: fg *fuzzy graph constraint; X isfg R; X is a function and R is its fuzzy graph*
r: u *usuality constraint; X isu R means usually (X is R)*
r: g *group constraint; X isg R means that R constrains the attribute-values of the group*

- *Primary constraints: possibilistic, probabilisitic and veristic*
- *Standard constraints: bivalent possibilistic, probabilistic and bivalent veristic*

Once the DAC matrix has been transformed into maximally compact concept based on graph representation and rules based on possibilistic relational universal fuzzy--type I, II, III, and IV (pertaining to modification, composition, quantification, and qualification), one can use Z(n)-compact algorithm and transform the DAC into a decision-tree and hierarchical graph that will represents a Q&A and/or questionnaire

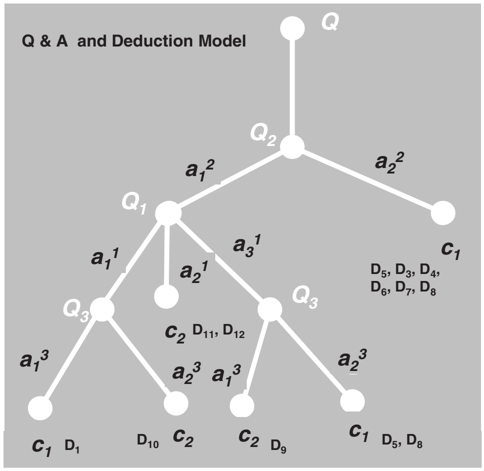


Fig. 4.a. Q & A model of DAC matrix and maximally Z(n)-compact rules and concept-based query-based clustering

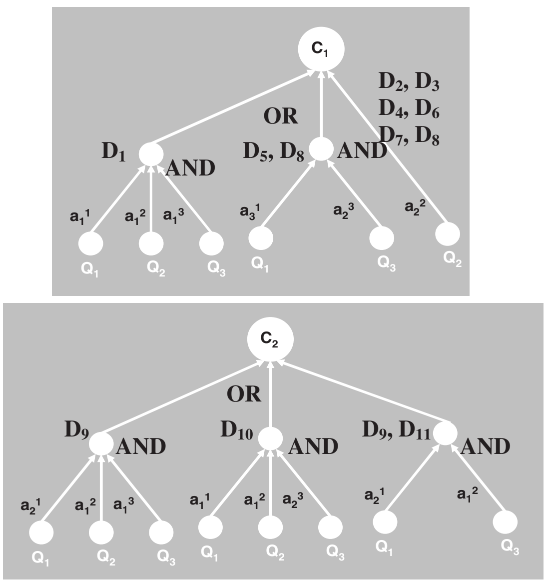


Fig. 4.b. Tree model of DAC matrix for Classes 1 and 2

model (Figure 10). The decision-tree and hierarchical graph will be a more general form of traditional tree since at each node the operation will be more general and tree could be constructed based on "isr" rather than "is" which in this case at each connection we may also have qualifiers such as "most likely", " probably", etc. Finally, the concept of semantic equivalence and semantic entailment based on possibilistic relational universal fuzzy will be used as a basis for question-answering (Q&A), questionnaire and inference from fuzzy premises (Figure 4). This will provide a foundation for approximate reasoning, language for representation of imprecise knowledge, a meaning representation language for natural languages, precisiation of fuzzy propositions expressed in a natural language, and as a tool for Precisiated Natural Language (PNL) and precisation of meaning. The maximally compact dataset based on Z(n)-compact algorithm and possibilistic relational universal fuzzy--type II will be used to cluster the data based on class-based query-based search criteria. For more information regarding technique based on possibilistic relational universal fuzzy--type I, II, III, and IV (pertaining to modification, composition, quantification, and qualification) refer to the following references [10-13].

5 Conclusion

In this paper, we developed a more advanced multi-aggregation model based on a hierarchical decision trees. For the learning process of this model, we developed a technique inspired from genetic programming. In this paper, we develop a decision-based questionnaire system. Another important and unique component of our system is compactification algorithm or Z-Compact. We developed a decision-based questionnaire system based on Z-Compact algorithm. The model has been used for other applications such as the design of a more intelligent search engine which includes the user's preferences and profile (Nikravesh, 2001a and 2001b).

Acknowledgment. Funding for this research was provided by the British Telecommunication (BT), OMRON and the BISC Program of UC Berkeley.

References

1. Fagin R. (1998) Fuzzy Queries in Multimedia Database Systems, Proc. ACM Symposium on Principles of Database Systems, pp. 1-10.
2. Fagin R. (1999) Combining fuzzy information from multiple systems. J. Computer and System Sciences 58, pp 83-99.
3. Koza J. R., Genetic Programming : On the Programming of Computers by Means of Natural Selection, Cambridge, Mass. : MIT Press, USA 1992, 819 pages.
4. Holland J. H.. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelli-gence. MIT Press, 1992. First Published by University of Michigan Press 1975.
5. Mizumoto M. (1989) Pictorial Representations of Fuzzy Connectives, Part I: Cases of T-norms, T-conorms and Averaging Operators, Fuzzy Sets and Systems 31, pp. 217-242.
6. Nikravesh M. (2001a) Perception-based information processing and retrieval: application to user profiling, 2001 research summary, EECS, ERL, University of California, Berkeley, BT-BISC Project. (http://zadeh.cs.berkeley.edu/ & http://www.cs.berkeley.edu/~nikraves/ & http://www-bisc.cs.berkeley.edu/).
7. Nikravesh M. (2001b) Credit Scoring for Billions of Financing Decisions, Joint 9th IFSA World Congress and 20th NAFIPS International Conference. IFSA/NAFIPS 2001 "Fuzziness and Soft Computing in the New Millenium", Vancouver, Canada, July 25-28, 2001.

8. Nikravesh M. and Azvine B. (2002), Fuzzy Queries, Search, and Decision Support System, Journal of Soft Computing, Volum 6, # 5, August 2002.
9. Nikravesh M., Azvine B.,Yagar R., and Zadeh L.A. (2003) "New Directions in Enhancing the power of the Internet", Editors:, to be published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (August 2003)
10. Nikravesh M. (2005), Evolutionary-Based Intelligent Information and Decision Systems, FuzzIEEE, May 22-25, Reno-Nevada, 2005 (Invited Talk)
11. Zadeh L.A. , (1976), A fuzzy-algorithmic approach to the definition of complex or imprecise concepts, Int. Jour. Man-Machine Studies 8, 249-291, 1976.
12. Zadeh L.A. and Nikravesh M., Perception-Based Intelligent Decision Systems; Office of Naval Research, Summer 2002 Program Review, Covel Commons, University of California, Los Angeles, July 30th-August 1st, 2002.

Fuzzy Signature and Cognitive Modelling for Complex Decision Model

Kok Wai Wong[1], Tamás D. Gedeon[2], and László T. Kóczy[3]

[1] School of Information Technology, Murdoch University , South St, Murdoch, Western Australia 6150,
k.wong@murdoch.edu.au
[2] Department of Computer Science, Australian National University, ACT 0200, Australia,
Tom.Gedeon@anu.edu.au,
[3] Department of Telecommunication & Telematics , Budapest University of Technology and Economics , Hungary, and Institute of Information Technology and Electrical Engineering, Széchenyi István University, Győr, Hungary
koczy@ttt.bme.hu,koczy@sze.hu

Abstract. As data is getting more complex and complicated, it is increasingly difficult to construct an effective complex decision model. Two very obvious examples where such a need emerges are in the economic and the medical fields. This paper presents the fuzzy signature and cognitive modeling approach which could improve such decision models. Fuzzy signatures are introduced to handle complex structured data and problems with interdependent features. A fuzzy signature can also be used in cases where data is missing. The proposed fuzzy signature structure will be used in problems that fall into this category. This paper also investigates a novel cognitive model to extend the usage of fuzzy signatures. This Fuzzy Cognitive Signature Modelling will enhance the usability of fuzzy theory in modelling complex systems as well as facilitating complex decision making process based on ill structured information or data.

1 Introduction

Fuzzy control and decision support systems are still the most important applications of fuzzy theory [1, 2]. This is a general form of expert control using fuzzy sets representing vague / linguistic predicates, modelling a system by If ... then rules. In the classical approaches of Zadeh [3] and Mamdani [4], the essential idea is that an observation will match partially with one or several rules in the model, and the conclusion is calculated by evaluation of the degree of these matches and by the use of the matched rules.

Fuzzy modelling has become popular because of its ability to assign meaningful linguistic labels to the fuzzy sets [5] in the rule base [6, 7]. However, a serious problem is caused by the high computational time and space complexity of rule bases describing systems with multiple inputs with proper accuracy. The complexity allows little general systems application (or real time control application) of classical fuzzy algorithms, where the inputs exceed about 6 to 10. These traditional fuzzy systems deal with very simple structured data, where the number of inputs is well defined, and values for each input occur for most or all data items. This further reduces their general applicability.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 380–389, 2007.
springerlink.com

Basically, practical fuzzy rule bases suffer from rule explosion. The number of possible rules necessary is $O(T^k)$ where K is the number of dimensions and T is the number of terms per input. In order to increase the problems solvable by fuzzy rule-base systems, it is essential to reduce T, k, or both. Decreasing T leads to sparse fuzzy systems, i.e. fuzzy rule-bases with "gaps" between the rules [8]. On the other hand, decreasing K leads to hierarchical fuzzy systems [9, 10].

A signature, as an abbreviated but unambiguously characteristic reference to data is widely used in computer based applications for data organization, retrieval, data mining. The abbreviation and conceptual clustering nature also suggests the use of fuzzy signatures. Fuzzy signatures create a natural bridge to verbal classifications, and human estimations. Fuzzy signatures which structure data into vectors of fuzzy values, each of which can be a further vector, are introduced to handle complex structured data [11, 12]. This will widen the application of fuzzy theory to many areas where objects are complex and sometimes interdependent features are to be classified and similarities / dissimilarities evaluated. Often, human experts can and must make decisions based on comparisons of cases with different numbers of data components, with even some components missing. This fuzzy signature tree structure is a generalisation of fuzzy sets and vector valued fuzzy sets in a way modelling the human approach to complex problems.

When dealing with a very large data set, it is possible that there is hidden hierarchical structure that appears in the sub-variable structures. This paper is used to address problems having this characteristic, and the possible use of cognitive modeling in describing the relationships of multiple fuzzy signatures.

2 Fuzzy Signature

The original definition of fuzzy sets was $A : X \to [0,1]$, and was soon extended to *L-fuzzy sets* by Goguen [13],

$$A_S : X \to [a_i]_{i=1}^{k},\ a_i = \begin{cases} [0,1] \\ [a_{ij}]_{j=1}^{k_i} \end{cases},\ a_{ij} = \begin{cases} [0,1] \\ [a_{ijl}]_{l=1}^{k_{ij}} \end{cases},$$

$A_L : X \to L$, L being an arbitrary algebraic lattice. A practical special case, *Vector Valued Fuzzy Sets* was introduced by Kóczy [14], where $A_{V,k} : X \to [0,1]^k$, and the range of membership values was the lattice of k-dimensional vectors with components in the unit interval. A further generalisation of this concept is the introduction of fuzzy signatures and signature sets, where each vector component is possibly another nested vector (right).

Fuzzy signature can be considered as special multi-dimensional fuzzy data. Some of the dimensions are inter-related in the sense that they form sub-group of variables, which jointly determine some feature on a higher level. Let us consider an example. Figure 1 shows a fuzzy signature structure.

The fuzzy signature structure shown in Figure 1 can be represented in vector form as follow:

$$x = \begin{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \\ \begin{bmatrix} x_{21} \\ \begin{bmatrix} x_{221} \\ x_{222} \\ x_{223} \end{bmatrix} \\ x_{23} \end{bmatrix} \\ \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \end{bmatrix}^T$$

Here $[x_{11}\ x_{12}]$ form a sub-group that corresponds to a higher level compound variable of x_1. $[x_{221}\ x_{222}\ x_{223}]$ will then combine together to form x_{22} and $[x_{21}[x_{221}\ x_{222}\ x_{223}]x_{23}]$ is equivalent on higher level with $[x_{21}\ x_{22}\ x_{23}] = x_2$. Finally, the fuzzy signature structure will become $x = [x_1\ x_2\ x_3]$ in the example.

The relationship between higher and lower level is govern by a set of fuzzy aggregations. The results of the parent signature at each level are computed from their branches with appropriate aggregation of their child signature. Let a_1 be the aggregation associating x_{11} and x_{12} used to derive x_1, thus $x_1 = x_{11}\ a_1\ x_{12}$. By referring to Figure 1, the aggregations for the whole signature structure would be a_1, a_2, a_{22}, and a_3. The aggregations a_1, a_2, a_{22}, and a_3 are not necessarily identical or

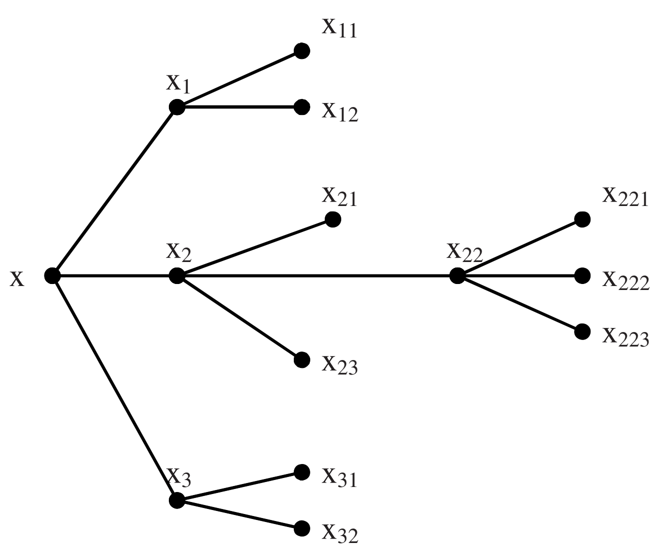


Fig. 1. A Fuzzy Signature Structure

different. The simplest case for a_{22} might be the *min* operation, the most well known t-norm. Let all aggregations be *min* except a_{22} be the averaging aggregation. We will show the operation based on the following fuzzy signature values for the structure in the example.

$$x = \begin{bmatrix} \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} \\ \begin{bmatrix} 0.2 \\ \begin{bmatrix} 0.6 \\ 0.8 \\ 0.1 \end{bmatrix} \\ 0.9 \end{bmatrix} \\ \begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix} \end{bmatrix}^T$$

After the aggregation operation is perform to the lowest branch of the structure, it will be described on higher level as:

$$x = \begin{bmatrix} 0.3 \\ \begin{bmatrix} 0.2 \\ 0.5 \\ 0.9 \end{bmatrix} \\ 0.1 \end{bmatrix}^T$$

Finally, the fuzzy signature structure will be:

$$x = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}^T$$

Each of the signatures mentioned here contains information relevant to the particular data point x_0. By going higher in the signature structure, less information will be kept. In some operations it is necessary to reduce and aggregate information to become compatible with information obtained from another source (some detail variables missing or simply being locally omitted). In such cases interpolation within a fuzzy signature rule base is done, where the fuzzy signatures flanking an observation are not exactly of the same structure. In this case the maximal common sub-tree must be determined and all signatures must be reduced to that level in order to be able to interpolate between the corresponding branches or roots in some cases.

3 Example Application of Fuzzy Signature

Let S_{S_0} denote the set of all fuzzy signatures whose structure graphs are sub-trees of the structural ("stretching") tree of a given signature S_0. Then the signature sets introduced on S_{S_0} are defined by

$$A_{S_0} : X \to S_{S_0} .$$

In this case the prototype structure S_0 describes the "maximal" signature type that can be assumed by any element of X in the sense that any structural graph obtained by a set of repeated omissions of leaves from the original tree of S_0 might be the tree stretching the signature of some A_{S_0} .

$$\begin{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix} \\ a_2 \\ \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} \\ \begin{bmatrix} \begin{bmatrix} a_{411} \\ a_{412} \end{bmatrix} \\ a_{42} \end{bmatrix} \end{bmatrix}$$

An example for the usefulness of this definition is given below. Let us think about some patients, whose daily symptom signatures are based on doctors' assessments according to the following scheme:

$$A_S = \begin{bmatrix} \text{fever} \begin{bmatrix} 8\text{ a.m.} \\ 12\text{ p.m.} \\ 4\text{ p.m.} \\ 8\text{ p.m.} \end{bmatrix} \\ \text{blood pressure} \begin{bmatrix} \text{systolic} \\ \text{diastolic} \end{bmatrix} \\ \text{nausea} \\ \text{abdominal pain} \end{bmatrix} .$$

Let us take a few examples with linguistic values and numerical signatures:

$$A_1 = \begin{bmatrix} \begin{bmatrix} \text{none} \\ \text{none} \\ \text{slight} \\ \text{slight} \end{bmatrix} \\ \begin{bmatrix} \text{normal} \\ \varnothing \end{bmatrix} \\ \text{slight} \\ \text{slight} \end{bmatrix} \mapsto \begin{bmatrix} \begin{bmatrix} 0.0 \\ 0.0 \\ 0.2 \\ 0.2 \end{bmatrix} \\ \begin{bmatrix} 0.5 \\ \varnothing \end{bmatrix} \\ 0.25 \\ 0.25 \end{bmatrix} ,$$

$$A_2 = \begin{bmatrix} \begin{bmatrix} \varnothing \\ \varnothing \\ \text{moderate} \\ \text{moderate} \end{bmatrix} \\ \begin{bmatrix} \text{slightly high} \\ \text{rather high} \end{bmatrix} \\ \text{slight} \\ \text{none} \end{bmatrix} \mapsto \begin{bmatrix} \begin{bmatrix} \varnothing \\ \varnothing \\ 0.4 \\ 0.4 \end{bmatrix} \\ \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} \\ 0.25 \\ 0.0 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} \begin{bmatrix} \text{rather high} \\ \text{high} \\ \text{rather high} \\ \text{rather high} \end{bmatrix} \\ \begin{bmatrix} \text{rather high} \\ \text{very high} \end{bmatrix} \\ \text{none} \\ \varnothing \end{bmatrix} \mapsto \begin{bmatrix} \begin{bmatrix} 0.8 \\ 0.6 \\ 0.8 \\ 0.8 \end{bmatrix} \\ \begin{bmatrix} 0.8 \\ 1.0 \end{bmatrix} \\ 0.0 \\ \varnothing \end{bmatrix}$$

(Ø stands for “not available”). Of course, normally the blood pressure values would initially rather be expressed by the physician as e.g. 120/75, which could then be converted to the linguistic values as appropriate for the patient, taking into account contextual information such as the higher normal blood pressure of infants and children and so on. As for most techniques, there is a significant role for the use of background knowledge of domain experts in data preprocessing.

Note that the structures above are different, which is an important point, namely, real world data is often like this, with missing components, or compressed parts. For patient 2, we have only 2 measurements for fever. The structure of the fuzzy signature contains some information by the association of vector components. The use of aggregation operators allows us to compare components irregardless of the different numbers of sub-components. Such aggregation operators would in general be designed for each vectorial component with the assistance of a domain expert. In this case, let us assume that the time of the day for fever is less significant, while the daily maximum value is most important. (By this assumption, the timing of temperature measurements must therefore be such that it ensure a reasonable coverage of the whole day.) The three signatures will be reduced to the following form, which is their maximal common sub-structure. (Note however, that in the case of the third signature, there is no data available on the presence or absence of abdominal pain, indicated by Ø in the original, nevertheless, this component was not eliminated from the other two, because of the high importance of this information in general.):

$$A_{1f} = \begin{bmatrix} 0.2 \\ \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \\ 0.25 \\ 0.25 \end{bmatrix}, A_{2f} = \begin{bmatrix} 0.4 \\ \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix} \\ 0.25 \\ 0.0 \end{bmatrix}, A_{3f} = \begin{bmatrix} 0.8 \\ \begin{bmatrix} 0.8 \\ 1.0 \end{bmatrix} \\ 0.0 \\ \varnothing \end{bmatrix}$$

The "fever component" can be verbally rewritten as "slight", "moderate" and "rather high", respectively. The signatures above still contain sufficient information about the "worst case fever" of each patient, while the detailed knowledge on the daily tendency of the fever is lost. This hierarchically structured access to the information is a key benefit of the fuzzy signatures.

We could continue this process further completely, and determine an overall "abnormal condition" measure $A_{1o} = [0.25]$, $A_{2o} = [0.4]$, $A_{3o} = [1.0]$. (Here the missing component being completely hidden already.)

4 Cognitive Modelling

Fuzzy signatures, as it has been described in the previous section, can address some issues of granulation and organisation well. In order to better model the human cognitive system, we have divided our cognitive modeling into two main categories. The first category consists of meta-levels of visual representation to model decision and cognitive behavior. For the ease of discussion, we will limit the model to a single meta-level in this paper. In this category the model consists of nodes and pointers to show the concepts and relations. Each node exhibits the behavior of a human cognitive system. Each node consists of three states, the sensory input state IN_i, current state CR_i, and action state AC_i. In the second category, nodes basically consist of the fuzzy signatures as described in the previous section. These signatures contain the knowledge necessary for the node to take any action.

Figure 2 shows a simple Fuzzy Cognitive Modeling. For node i,

$$N_i = (IN_i, CR_i, AC_i)$$

The modeling of the three states can be represented by the original definition of the fuzzy sets which is

$$A : X \rightarrow [0,1]$$

For some current states CR_i, if necessary, they will go down to the fuzzy signature level as

$$CR_i = A_{S_i}$$

where A_{s_i} is the fuzzy signature contributing to the knowledge of the node N_i.

There are basically two modes of operation for each fuzzy cognitive node: static and dynamic mode.

The static mode operation within the nodes is as follows:

- The three states within each node can be linked using fuzzy linguistics rules with antecedents and consequents.
- The antecedent of the fuzzy rules contains either the sensory input state, or the current state, or both the sensory input and the current state, i.e. (IN_i, CR_i).
- The consequent is the action state (AC_i).
- The operations between the antecedent/s and consequent are the ones used with fuzzy rules in general.
- Depending on how the fuzzy rules are constructed, the model may allow missing states within each node.
- When the action state is obtained, it can either propagate to the next node or the node can convert into dynamic mode.

Dynamic mode operations of the nodes are described by the following:

- In this mode, the time factor (*t*) is considered.
- Consequently, the fuzzy signature in the node will be formulated as $A_{s_i}(t)$.
- For (*t+1*), cross check with the fuzzy rules in the Fuzzy Cognitive Meta-level will be performed in order to see if the present node action can propagate to the next node. If not, the node will enter into (*t+2*). This is continued until an action can be propagated to the next connecting node/s, or when there exists a fuzzy rule to resolve the outcome.

For cases where there are more than one input arrows coming into the node, IN_i consists of more than one input component:

$$IN_i = \{IN_{i1}, IN_{i2}, ..., IN_{in}\}.$$

In order to avoid the combinatorial explosion problem within the rules, the relationship among $IN_{i1}, IN_{i2}, ..., IN_{in}$ is governed by a set of fuzzy aggregations, which eventually reduce the input state to a single component. The aggregations among them are not necessarily identical, even in the type. They can be a mixture of t-norms, s-norms, averaging aggregations and so on. Thus,

$$IN_i = IN_{i1} a_{1,2} IN_{i2} a_{2,3} ... a_{n-1,n} IN_{in}$$

Therefore, regardless of how many inputs are fed into the node, they will be resolved into a single fuzzy set before being used by the node. With this flexibility, missing information is allowed when performing modeling. For applications where computing power is crucial, or when the available information or data is massive, the nodes can be arranged in a distributed computing architecture, with each cognitive node being taken care of by separate nodes in the distributed computing cluster.

In those nodes where there is no input state, for example, node N_1 in Figure 2, the input state could be:

$$IN_1 = \varnothing$$

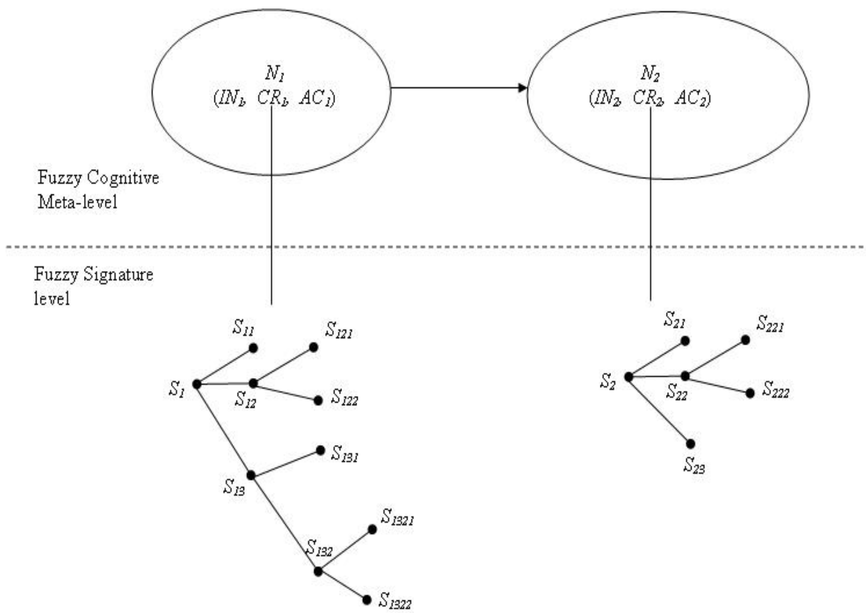


Fig. 2. The basic Efficient Fuzzy Cognitive Model

5 Conclusion

We described a technique for dealing with problems with complex and interdependent features or where data is missing. This was done by using the concept of fuzzy signatures, which extends the idea of vectorial fuzzy sets to components with varying numbers of sub-components. We also introduced a cognitive model to expand the usage of fuzzy signature for cases where complex decision is required. This hierarchical structuring allows the further use of domain experts as the information can be abstracted to higher levels, analogous to patterns of human expert decision making.

In the next phase, investigations will be done for the applicability of this new model to various problems where the complexity of the task, or its indeterministic and/or vague nature traditionally necessitate the involvement of human decision makers, such as in biology and medicine, agriculture, business, management and economics, and many others.

References

1. Chong, A., Gedeon, T.D., Wong, K.W.: Extending the Decision Accuracy of A Bioinformatics System. In Reznik, L., Kreinovich, V. (Eds.) Soft Computing in Measurement and Information Acquisition, Physica-Verlag, Springer-Verlag. (2003) 151-163.
2. Gedeon, T.D.: Clustering Significant Words using their Co-occurrence in Document Sub-Collections. Proceedings 7th European Congress on Intelligent Techniques and Soft Computing (EUFIT'99), Aachen. (1999) 302-306.

3. Zadeh, LA: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics (1973) 28-44.
4. Mamdani, EH and Assilian, S: An experiment in linguistic synthesis with a fuzzy logic controller. Int. Journal of Man-Machine Studies. (1975) 1-13.
5. Zadeh, L.A.: Fuzzy Algorithm. Information and Control. vol. 12. (1968) 94-102.
6. Sugeno, M. and Takagi, T.: A New Approach to Design of Fuzzy Controller. Advances in Fuzzy Sets, Possibility Theory and Applications. (1983) 325-334.
7. Nguyen, H.T., and Sugeno, M., (Eds): Fuzzy Systems: Modeling and Control, The Handbook of Fuzzy Sets Series. (1998) Kluwer Academic Publishers.
8. Kóczy, L.T., and Hirota, K.: Approximate reasoning by linear rule interpolation and general approximation. Int. J. Approx. Reason. Vol. 9. (1993)197-225.
9. Sugeno, M., Murofushi, T., Nishino, J., and Miwa, H.: Helicopter flight control based on fuzzy logic. Proceedings of Fuzzy Engineering toward Human Friendly Systems '91. (1991) 1120-1124.
10. Kóczy, L.T., and Hirota, K.: Approximative inference in hierarchical structured rule bases. Proceedings of Fifth IFSA World Congress. (1993)
11. Kóczy, L.T., Vámos, T., and Biró, Gy.: Fuzzy Signatures. Proceedings of EUROFUSE-SIC '99. (1999) 210-217.
12. Vámos, T., Kóczy, L.T., Biró, Gy.: Fuzzy Signatures in Data Mining. Proceedings of IFSA World Congress and 20th NAFIPS International Conference. (2001) 2842-2846.
13. Goguen, J. A.: "L-fuzzy Sets," J. Math. Anal. Appl. 18, (1967) 145-174.
14. Kóczy, L. T.: "Vectorial I-fuzzy Sets," in: M. M. Gupta & E. Sanchez (eds.): Approximate Reasoning in Decision Analysis, North Holland, Amsterdam, (1982) 151-156.

Part XI

Soft Computing in Petroleum Applications

Estimating Monthly Production of Oil Wells

Rajani Goteti[+], A. Tamilarasan, R.S. Balch[+], S. Mukkamala, and A. H. Sung

Department of Computer Science, New Mexico Tech, Socorro, NM 87801,
Institute for Complex Additive Systems and Analysis, Socorro, NM 87801,
[+] Petroleum Recovery Research Center,
{rajini,ashwin,srinivas,sung}@cs.nmt.edu, balch@nmt.edu

Abstract. This paper describes results concerning the capability of supervised machine learning techniques to predict production potential for a single formation, prior to drilling, over a 16,000 square mile area of SE New Mexico. In this paper a neural network is used to predict production potential for a single formation of SE New Mexico region. The process involved gathering data for use as potential inputs, collecting production data at known wells, selecting optimal inputs, developing and testing various network architectures, making predictions, analyzing and applying the results. This predicted production was further refined by excluding production at locations where the Woodford shale was not present. Results were evaluated by inspecting a map of predicted production and performing statistical testing, including a correlation of predicted and actual production, which produced a correlation coefficient of 0.79. The results were then used by the Devonian FEE Tool, an expert system designed to reduce exploration risk.

1 Introduction

Computational intelligence techniques with many possible applications are develped in the oil and gas industry. In this paper a neural network is used to identify the production at a set of wells with attempted production from the formation. Production was measured in barrels of oil equivalent per month (BOEPM) averaged over the first year 12 months of production.

The data included producing wells and dry holes. Variables considered as potential inputs included the thickness of the primary source rock (Woodford shale), total organic carbon (TOC), production index (PI), paleo thickness, curvature of paleo structure, and permeability. Two variables were derived; the product of Woodford thickness, TOC and PI, termed generative potential, and the log of generative potential. Variables were interpolated using kriging over the study area. Potential input variables were tested against the output to determine the best variables for training. First, a fuzzy ranking algorithm selected Woodford thickness and TOC. A secondary correlation study also recommended Woodford thickness and TOC, and added permeability and generative potential.

Fuzzy Expert Exploration Tool, (FEE Tool) has been developed for the Devonian Carbonates of SE New Mexico by the Reservoir Evaluation and Advanced Computational

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 393–403, 2007.
springerlink.com

Techniques (REACT) group at the Petroleum Recovery Research Center (PRRC) of New Mexico Tech. This tool was modeled on a previous successful expert system developed for the Lower Brushy Canyon formation [1]. The FEE Tool is built around a knowledge base of rules generated through interviews with experts having knowledge of the target formation. Each rule requires an input, either provided by the user or contained in a project database. As with the Lower Brushy Canyon expert system, a desired input value for the Devonian FEE Tool is predicted production. To obtain values of predicted production to use as inputs to the knowledge base rules, neural networks were used. The neural networks provided values of predicted production for each gridpoint of the 16,000 square mile region (one grid square equals 160 acres, with a total of 64,347 points for the region).

To generate relevant values of predicted production for the region, data were collected for use as potential inputs and outputs for training a neural network. Potential inputs included geophysical data, structure data, source rock data and log data. Output data came from production reports for both producing and non-producing wells. For the purposes of this work, production was measured in barrels of oil equivalents or BOE, setting 6 mcf of gas equal to 1 BOE.

Two methods were used to select the best input variables from the available data. These were fuzzy ranking algorithms and linear correlations. Four input variables were selected, which were then used, along with the production data, to develop, train and test the neural network. Once a suitable network was created, it was used to generate predicted production for the entire region. The predicted production was evaluated by viewing maps and histograms of the results, as well as measuring the correlation between the predicted results and the actual production at locations with producing wells. Prior to adding the results to the Devonian FEE Tool database, the data were filtered to remove possibly boundary effects, showing production in regions where the primary source rock and seal (Woodford Shale) was not present [2]. The final Devonian FEE Tool that included these results performed well at reducing drilling risk in the formation [1].

The dataset used for analysis is described in seection 2. A brief description of features used and methods used to select the features is given in section 3. Description of neural network used for prediction is given in section 4. Summary of the results is give in section 5. Conclusions of our work are given in section 6.

2 Data Used for Analysis

Geological Data. The Devonian Carbonates zone in southeast New Mexico is a structural formation dating from the Silurian and Devonian period. Three important characteristics that correlate to production from this formation are the thickness, organic richness and thermal maturity of the Woodford Shale, the principal source rock for the formation.

A database is developed for the region that included the following variables

- Woodford Shale Thickness
- Subsea Elevation (to top of the Woodford Shale)

- Total Organic Carbon (TOC)
- Production Index (PI)
- Generative Potential (GP) (defined for this project as Thickness·TOC·PI)
- Log (GP)
- Permeability (in md)
- Curvature (calculated curvature of subsea elevation)
- Structure (measured as shown in Figure 1)
- Closure (Figure 1)
- Structural Relief (Figure 1)
- Paleostructure (calculated from Woodford and Abo formation tops)

The variables Woodford Shale thickness, subsea elevation, TOC, PI, and permeability were recorded at producing wells throughout the region [3], and then interpolated for the entire region using a kriging technique. The remaining variables were calculated from these variables. The result was a database that included the values of these eleven variables at each of 64,347 grid points (each gridpoint corresponds to a 160 acre square in the region).

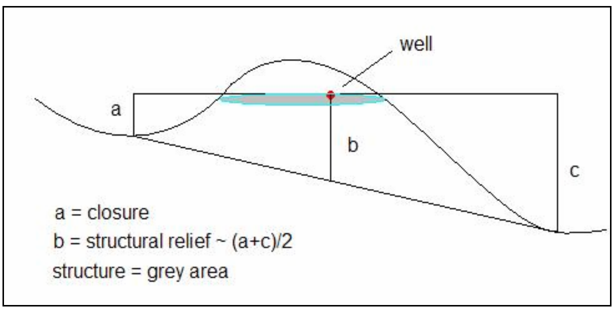


Fig. 1. Graph illustrating how the variables closure, structural relief and structure were estimated based on the formation surface

2.1 Production Data

To train the neural network, a data set containing both the selected input variables and the corresponding production had to be developed. Production data was compiled for 172 wells completed to this formation. This data included gas wells, oil wells, wells with mixed production and unsuccessful wells. For each well, production was measured in barrels of oil equivalent per month, averaged over the first producing year (using 6mcf = 1BOE). Of the wells used, 105 were unsuccessful, 15 had production less than 1000 BOEPM, 31 had production between 1000 and 5000 BOEPM, and 21 had production over 5000 BOEPM with a maximum of 38,970. The wells were distributed throughout the study area as shown in Figure 2.

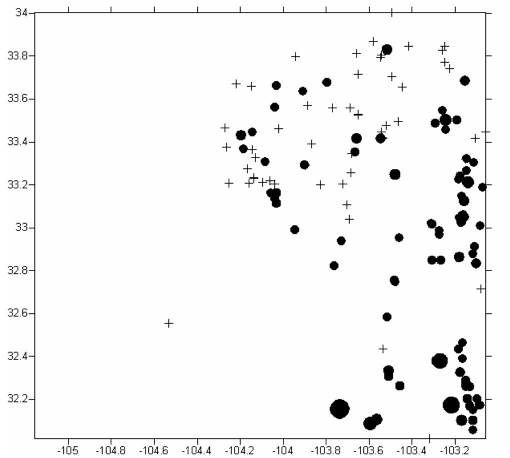


Fig. 2. Location of the producing wells and unsuccessful wells used to train the neural network

3 Selection of Input Variables

The feature selection and ranking of geological data is similar in nature to various engineering problems that are characterized by:

- Having a large number of input variables **x** = (x_1, x_2, …, x_n) of varying degrees of importance to the output **y**; i.e., some elements of **x** are essential, some are less important, some of them may not be mutually independent, and some may be useless or irrelevant (in determining the value of **y**)
- Lacking an analytical model that provides the basis for a mathematical formula that precisely describes the input-output relationship, **y** = ***F*** (**x**)
- Having available a finite set of experimental data, based on which a model (e.g. neural networks) can be built for simulation and prediction purposes
- Excess features can reduce classifier accuracy
- Excess features can be costly to collect
- If real time classification is important, excess features can reduce classifier operating speed independent of data collection
- If storage is important, excess features can be costly to store

The next step in neural network development was to determine which of the eleven variables to use as inputs. Using all of the variables would be cumbersome, and would present problems of overtraining, as the complexity of the network is dependent on the number of records (in this case, 172 records corresponding with the 172 wells). On the other hand, if only one or two variables were sufficient, a neural network would not be the best choice. In order to select the most relevant data to use as input variables, two methods were used; a fuzzy ranking algorithm and a linear correlation.

3.1 Fuzzy Ranking

Fuzzy ranking is applied to select and rank the most important inputs from a list of inputs. This helps in building a robust and reliable neural network for the analysis. Fuzzy ranking is done using fuzzy curves and fuzzy surfaces.

Consider a system where there are n possible input variables, one desired output variable and k data points.

For each of the input variables, a Gaussian fuzzy membership function is generated at each data point as follows:

$$F_{ij}(x_i) = \exp(-(\frac{x_{ij} - x_i}{s})^2) \quad j = 1,2,...k \tag{1}$$

The fuzzy surface, a two dimensional fuzzy curve is defined using two input variables as:

$$s(x_1, x_2) = \frac{\sum_{j=1}^{k} F_{1j}(x_1) \cdot F_{2j}(x_2) \cdot y_j}{\sum_{j=1}^{k} F_{1j}(x_1) \cdot F_{2j}(x_2)} \tag{2}$$

With this fuzzy curves and fuzzy surfaces we can determine which set of inputs predicts the output accurately. The fuzzy ranking algorithm uses the mean square error between the fuzzy curve of the input variable x_i and the output variable y_i. [2].

The software package, Fuzzy Rank, was developed by the REACT group at New Mexico Tech [4]. It is designed to determine which inputs are best at predicting a desired output. The software allows the user to upload input variables along with an output variable, and it ranks the input variables according to the fuzzy ranking algorithm, providing the user with the top ranked outputs. Given an input, output pair of (x_i, y_i), a fuzzy curve is defined by first defining a fuzzy membership function as follows [5]:

$$F_i(x) = e^{-\left(\frac{x_i - x}{b}\right)^2} y_i \tag{3}$$

Here, b is a constant usually taken to be 10% of the range of the input variables. For this reason, the input data is often normalized prior to applying the fuzzy ranking algorithm, thus allowing b to be set to 0.1.

From the fuzzy membership function, the fuzzy curve is defined as:

$$FC(x) = \frac{\sum_{i=1}^{N} F_i(x)}{\sum_{i=1}^{N} F_i(x) / y_i} \tag{4}$$

Two examples of fuzzy curves are given in figure 3. When the input is random, the trend line is close to horizontal, indicating that the input would not be a good predictor of the output. Data that is more likely to be a good predictor of the output has a trend line similar to the one shown in the top example.

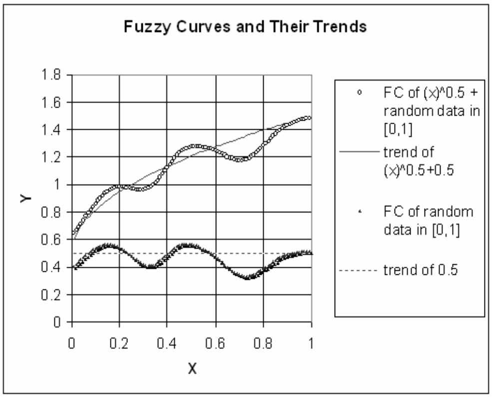


Fig. 3. Two fuzzy curves, one for random data (bottom) and one for data with a trend (top)

One way to evaluate the fuzzy curves is to compute the rank, the distance between the maximum and minimum values on the curve. The higher the rank, the better the data is for modeling.

When the Fuzzy Rank software was applied to the eleven possible input variables, two high ranking variables were selected as having the best combination of uniqueness and utility for arriving at a solution, Woodford thickness and TOC.

3.2 Linear Correlation

In order to look at the data a different way, and select a few more variables to use as inputs, the linear correlation of all eleven variables with the production was computed [7]. The four variables with the largest correlations (or smallest p-values) were selected. As expected, the two variables selected by fuzzy rank were also on this list. The two new variables were permeability and generative potential. It is interesting to note that the permeability correlation is negative. This is due to an excessive amount of water production [3]. The correlations and the p-values are shown in Table 1.

Table 1. Correlation coefficients and p-values for input variables

Variable	Correlation	p-value
Wood. Th.	0.378	0.000
TOC	0.315	0.000
Perm	-0.131	0.086
GP	0.362	0.000

The success of the training of a neural network can be evaluated by looking at the R^2 value and the correlation coefficient. In general, the closer these values are to 1, the better trained the network, however values approaching 1 can also indicate a network has been overtrained. R^2 and the correlation coefficient (r) are defined below [6].

$$R^2 = 1 - \frac{\sum_i (d_i - c_i)^2}{\sum_i (d_i - \bar{c})^2}$$
$$r = \sqrt{R^2} \tag{5}$$
$$\bar{c} = \sum_i c_i / n$$

Where:
d_i is the desired output for record i
c_i is the neural network output for record i
n is the number of records

The results of feature ranking are shown in the table below.

Table 2. Ranking of the features

Rank	Feature
1	TOC
2	paleo1
3	flexure
4	WThick
5	GP
6	structure
7	st_relief
8	perm
9	logGP
10	PI
11	closure

4 Scaled Conjugate Gradient Decent

Moller [9] introduced the scaled conjugate gradient algorithm as a way of avoiding the complicated line search procedure of conventional conjugate gradient algorithm (CGA). According to the SCGA, the Hessian matrix is approximated by [8]

$$E''(w_k)p_k = \frac{E'(w_k + \sigma_k p_k) - E'(w_k)}{\sigma_k} + \lambda_k p_k \tag{6}$$

where E' and E'' are the first and second derivative information of global error function $E\ (w_k)$. The other terms p_k, σ_k and λ_k represent the weights, search direction, parameter controlling the change in weight for second derivative approximation and parameter for regulating the indefiniteness of the Hessian. In order to get a good quadratic approximation of E, a mechanism to raise and lower λ_k is needed when the Hessian is positive definite.

5 Summary Results

This section summarizes results obtained from using neural networks for predicting oil production and subsequent results obtained from fuzzy ranking.

Once the feature has been ranked using the fuzzy ranking we now build the neural network architecture with 3 layers. The first layer is the input layer and the second layer is the hidden layer and the third layer is the output layer. The first 4 most significant features were considered for the experiment and the graph between the actual output and the predicted output is plotted.

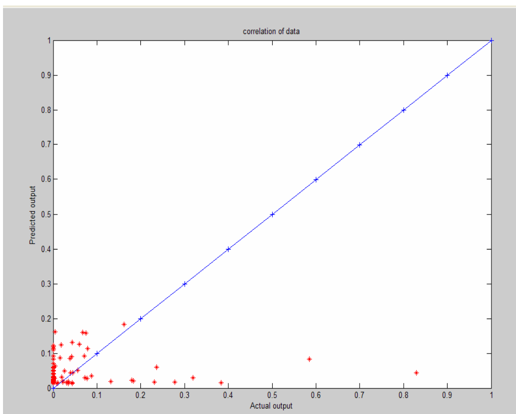


Fig. 4. Actual Output Vs Predicted Output

The graph shows that the most of the predicted output is scattered around the line showing that it is a good prediction. Accuracy is not a good way of measure since it doesn't show local minima and local maxima. Even if there are local minima the accuracy can be as high as 99% so we do not consider the accuracy measure in this case. The local minima occur due to neuron saturation and a local maximum occurs due to neuron overloading.

Neural network predicted production results are evaluated first by mapping them over the region. The producing wells and dry holes were then overlain over the contour map of predicted production. The data was then filtered by zeroing out

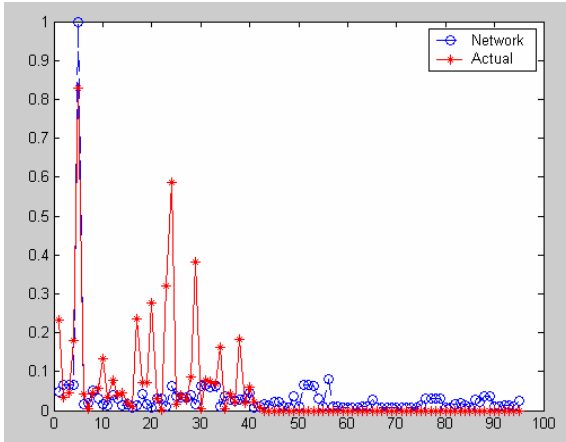


Fig. 5. Oil Prediction Using Scaled Conjugate Gradient Decent(SCG)

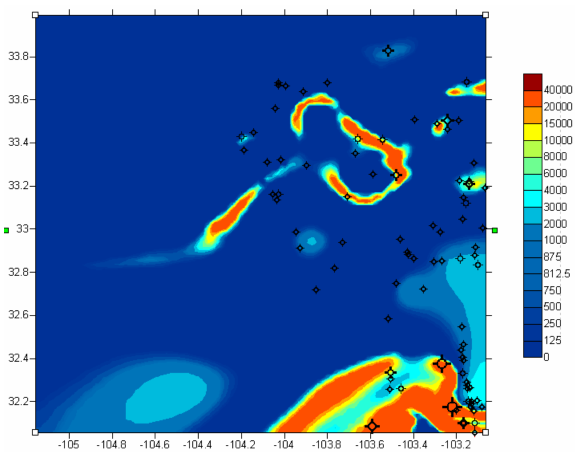


Fig. 6. Oil Prediction Using 611 Data Points

production in a small section of the region where the Woodford shale was not present. As another means of testing the data, the predicted values at the 172 well locations were correlated with the actual values at these locations, back propagation neural network resulted in a correlation of 0.79. The correlation was greatly improved (0.907) by using SCG neural network. Contour map are produced, to show the predicted values, of the producing wells and the dry holes. Contour maps using 611 and 64348 are given in figures 6 and 7.

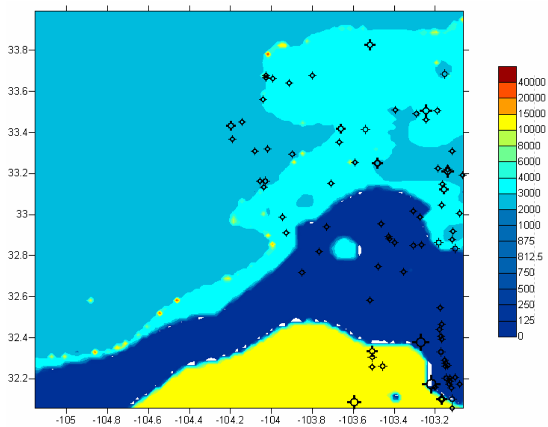


Fig. 7. Oil Prediction Using 64348

5 Conclusions

A few observations are drawn from the results:

- Scaled conjugate decent neural network achieved the best performance among the neural networks in terms of prediction with a very high correlation of (.9075).

Regarding feature selection and ranking, we observe that

Different feature selection techniques (statistical methods) along with a comparative study of feature ranking (fuzzy ranking) for oil prediction are proposed.

- Fuzzy ranking is done using fuzzy curves and fuzzy surfaces. Fuzzy ranking produces largely consistent results. Features ranked as important by fuzzy ranking are used as inputs to the neural networks. This helps in building a robust and reliable neural network for the analysis.
- Using the important features gives the most remarkable performance

We demonstrate that using these fast execution machine learning methods we can achieve high classification accuracies in a fraction of the time required by the well know traditional prdiction methods.

Acknowledgements

Support for this research received from PRRC (Petroleum Recovery Research Center, a division of New Mexico Tech) and ICASA (Institute for Complex Additive Systems Analysis, a division of New Mexico Tech) are gratefully acknowledged.

References

1. Balch, R.S., Broadhead, R.F., Ruan, T., and Schrader, S.M.: "Risk Reduction with a Fuzzy Expert Exploration Tool," Final Report to the U.S. Department of Energy, December 2004.
2. Broadhead, R.F.: "Risk Reduction with a Fuzzy Expert Exploration Tool, Geologic Data Acquisition and Analysis," Report of Project Activities, September 2004.
3. Balch, R.S., Broadhead, R.F., Ruan, T.: "Fourth Annual Progress Report for Risk Reduction with a Fuzzy Expert Exploration Tool," Report to the U.S. Department of Energy, March 2003.
4. Schrader, S.M.: *REACT Software User's Guide*, PRRC, New Mexico Tech, December 2004.
5. Balch, R.S., Stubbs, B.S., Weiss, W.W., and Wo, S.: "Using Artificial Intelligence to Correlate Multiple Seismic Attributes to Reservoir Properties," paper SPE 56733 presented at the 1999 SPE Annual Technical Conference and Exhibition, Houston, October 5-6.
6. Du, Y.: Optimization of Artificial Neural Network Design through Synthetic Datasets Analysis, M.S. Thesis, New Mexico Institute of Mining and Technology, May 2002.
7. Statsoft: *Electronic Textbook, Neural Networks*, http://www.statsoft.com/textbook/stneunet.html, 2003.
8. Moller A F, A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Volume (6), pp. 525-533, 1993.

IRESC: Reservoir Characterization

Masoud Nikravesh

EECS Department
University of California, Berkeley
Nikravesh@eecs.berkeley.edu

Abstract. Reservoir characterization plays a crucial role in modern reservoir management. It helps to make sound reservoir decisions and improves the asset value of the oil and gas companies. It maximizes integration of multi-disciplinary data and knowledge and improves the reliability of the reservoir predictions. The ultimate product is a reservoir model with realistic tolerance for imprecision and uncertainty. Soft computing aims to exploit such a tolerance for solving practical problems. In reservoir characterization, these intelligent techniques can be used for uncertainty analysis, risk assessment, data fusion and data mining which are applicable to feature extraction from seismic attributes, well logging, reservoir mapping and engineering. The main goal is to integrate soft data such as geological data with hard data such as 3D seismic and production data to build a reservoir and stratigraphic model. While some individual methodologies (esp. neurocomputing) have gained much popularity during the past few years, the true benefit of soft computing lies on the integration of its constituent methodologies rather than use in isolation.

1 Introduction

With oil and gas companies presently recovering, on the average, less than a third of the oil in proven reservoirs, any means of improving yield effectively increases the world's energy reserves. Accurate reservoir characterization through data integration (such as seismic and well logs) is a key step in reservoir modeling & management and production optimization.

Soft computing is bound to play a key role in the earth sciences. This is in part due to subject nature of the rules governing many physical phenomena in the earth sciences. The uncertainty associated with the data, the immense size of the data to deal with and the diversity of the data type and the associated scales are important factors to rely on unconventional mathematical tools such as soft computing. Many of these issues are addressed in a recent books, Nikravesh et al. (2003a, 2003b), Wong et al (2001), recent special issues, Nikravesh et al. (2001a and 2001b) and Wong and Nikravesh (2001) and recent papers by Nikravesh et al. (2001c) and Nikravesh and Aminzadeh (2001).

This paper address the key challenges associated with development of oil and gas reservoirs. Given the large amount of by-passed oil and gas and the low recovery factor in many reservoirs, it is clear that current techniques based on conventional methodologies are not adequate and/or efficient. We are proposing to develop the next generation of Intelligent Reservoir Characterization (IRESC) tool, based on Soft

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 404–414, 2007.
springerlink.com

computing (as a foundation for computation with perception) which is an ensemble of intelligent computing methodologies using neuro computing, fuzzy reasoning, and evolutionary computing. We will also provide two real world examples.

2 Intelligent Reservoir Characterization

Figure 1 shows techniques to be used for intelligent reservoir characterization (IRESC). The main goal is to integrate soft data such as geological data with hard data such as 3-D seismic, production data, etc. to build reservoir and stratigraphic models. In this case study, we analyzed 3-D seismic attributes to find similarity cubes and clusters using three different techniques: 1. k-means, 2. neural network (self-organizing map), and 3. fuzzy c-means. The clusters can be interpreted as lithofacies, homogeneous classes, or similar patterns that exist in the data. The relationship between each cluster and production-log data was recognized around the well bore and the results were used to reconstruct and extrapolate production-log data away from the well bore. The results from clustering were superimposed on the reconstructed production-log data and optimal locations to drill new wells were determined.

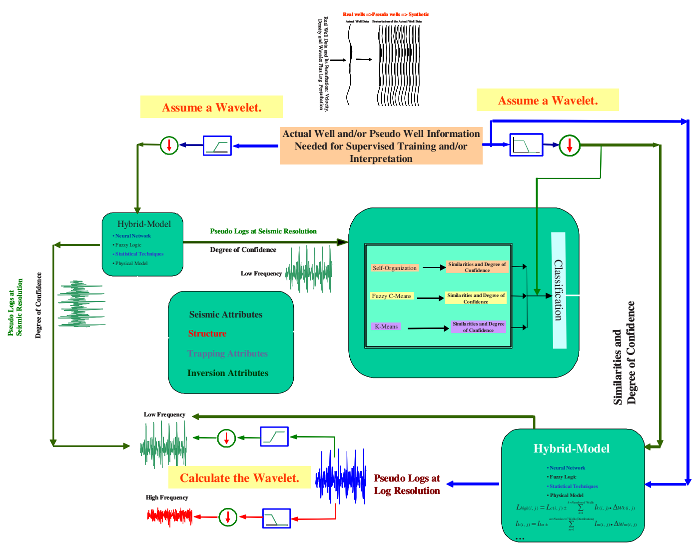


Fig. 1. Technique used in IRESC Software

2.1 Red River Reservoir

Our example is from a field that produces from the Red River Reservoir. A representative subset of the 3-D seismic cube, production log data, and an area of interest were selected in the training phase for clustering and mapping purposes. The subset (with each sample equal to 2 msec of seismic data) was designed as a section passing through all the wells. However, only a subset of data points was selected for

Table 1. List of the attributes calculated in this study

1.Amplitude envelope 2.Amplitude weighted cosine phase 3.Amplitude weighted frequency 4.Amplitude weighted phase 5.Apparent polarity 6.Average frequency 7.Cosine instantaneous phase 9.Derivative instantaneous amplitude 8.Derivative 10.Dominant Frequency 11.Instantaneous Frequency 12.Instantaneous Phase 13.Integrated absolute amplitude 14.Integrate 15.Raw seismic 16.Second derivative instantaneous amplitude 17.Second derivative	**1-17; Seismic Attributes**
	Structure and Trapping Attributes. Six horizons and with four attributes out of seven attributes.. Column A: line identifier Column B: trace or cross-line identifier Column C: easting in feet Column D: northing in feet 1 Column E: horizon time in msecs 2 Column F: time_resd, first order residual of horizon time, negative is high or above plane 3 Column G: aspect, angle of updip direction at horizon (present day) Column H: next deeper horizon time (used for calculation of iso values) 4 Column I: iso, incremental time to next horizon 5 Column J: iso_resd, first order residual of iso time, negative is thinner (faster) than plane 6 Column K: iso_aspect, angle of updip direction (at time of burial) 7 Column L: cum_iso_resd, cumulative iso_resd from Winnipeg to this horizon
18. Acoustic Impedance 19. Low Frequency of 18. 20. Reflectivity Coefficients 21. Velocity 22. Density	**18-22; Inversion Attributes**
23. computed_Neutron_Porosity 24.computed_Density_Porosity 25. computed_Pwave 26. computed_Density 27. computed_True_Resistivity 28. computed_Gamma_Ray	**23-28; Pseudo Logs Attributes**

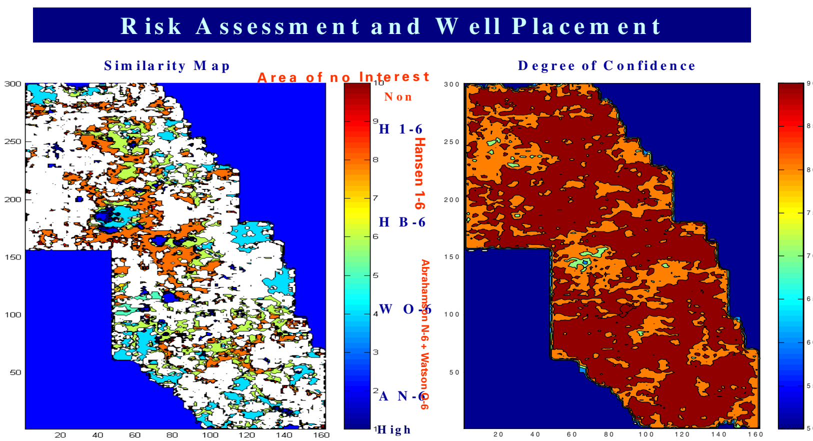
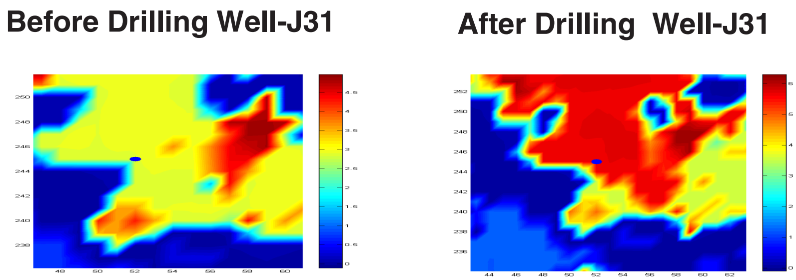


Fig. 2. Performance of IRESC technique for prediction of the high-potential and no-potential producing D-Zone based on virtual logs

clustering purposes, representing the main Red River focus area. This subset covers the horizontal and vertical boreholes of producing wells. For clustering and mapping, there are two windows that must be optimized, the seismic window and the well log window. Optimal numbers of seismic attributes and clusters need to be determined, depending on the nature of the problem. Expert knowledge regarding geological parameters has also been used to constrain the maximum number of clusters to be selected. In this study, seventeen seismic attributes, five inversion attributes, six pseudo log attributes in seismic resolution and seven structure/trapping attributes, equaling a total of 35 attributes have been used (Table 1).



Before Drilling Well-J31

Actual		Cutoff		Delta(Depth)
	8.000	10.000	12.000	
D Phi-h	6.480	5.812	4.537	58.000
D1 phi-h	1.811	1.501	1.171	15.000
D2 phi-h	4.669	4.311	3.366	38.000

Predicted		Cutoff		Delta(Depth)
	8.000	10.000	12.000	
D Phi-h	4.790	4.351	3.882	53.500
D1 Phi-h	0.944	0.724	0.588	10.500
D2 Phi-h	3.846	3.627	3.294	31.500
±	0.609	0.620	0.707	2.500
	0.129	0.084	0.111	3.500
	0.738	0.703	0.817	2.500

After Drilling Well-J31

Actual		Cutoff		Delta(Depth)
	8.000	10.000	12.000	
D Phi-h	6.480	5.812	4.537	58.000
D1 phi-h	1.811	1.501	1.171	15.000
D2 phi-h	4.669	4.311	3.366	38.000

Predicted		Cutoff		Delta(Depth)
	8.000	10.000	12.000	
D Phi-h	6.760	6.137	5.591	52.000
D1 Phi-h	1.575	1.267	1.054	14.000
D2 Phi-h	5.186	4.870	4.537	35.000
	0.119	0.111	0.105	0.354
	0.039	0.032	0.033	0.315
	0.081	0.079	0.072	0.253

Fig. 3. Qualitative and quantitative analysis and performance of IRESC technique for prediction of the high-potential and no-potential

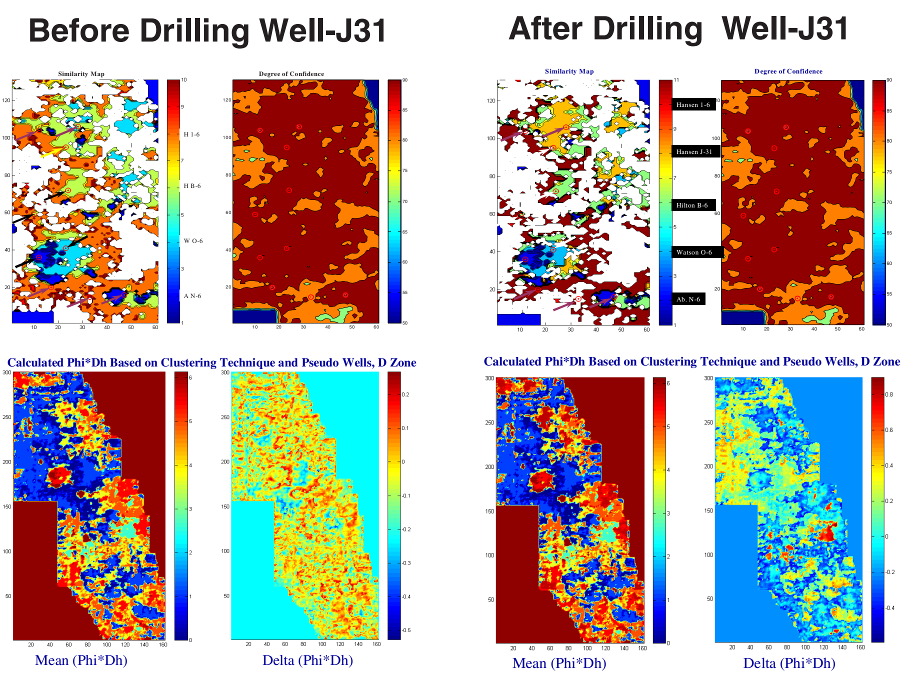


Fig. 4. Qualitative and quantitative analysis and performance of IRESC technique for prediction of D-Zone and Phi*Dh

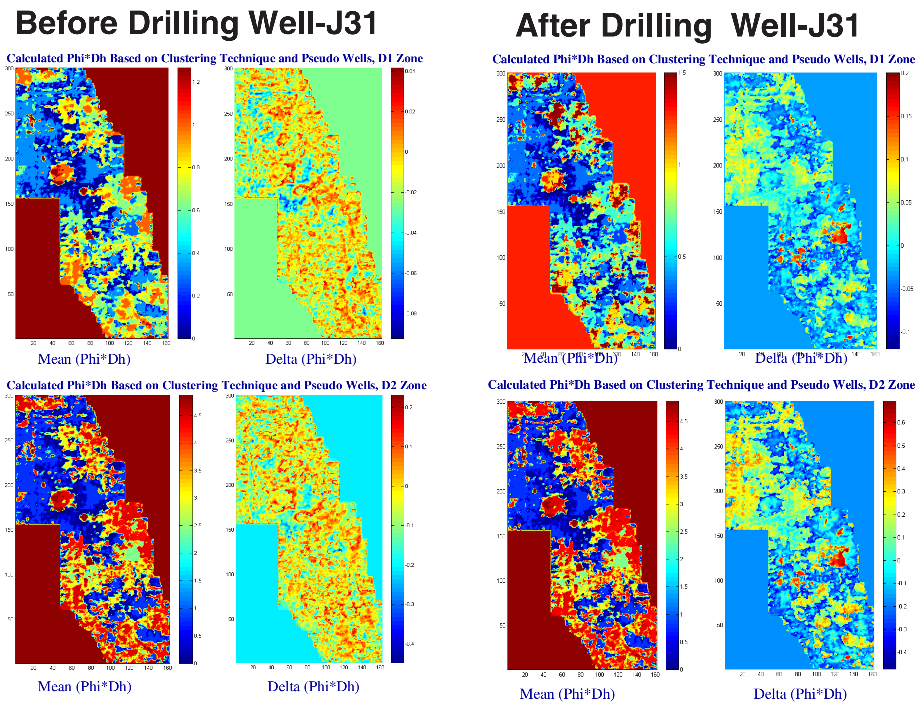


Fig. 5. Performance of IRESC technique for prediction of Phi*Dh for D1-Zone and D2-Zone

Clustering was based on three different techniques, k-means (statistical), neural network, and fuzzy c-means clustering. Different techniques recognized different cluster patterns and one can conclude that the neural network predicted a different structure and patterns than the other techniques. Finally, based on a qualitative and quantitative analysis given the prediction from high resolution data using the technique presented in Figure 1, specific clusters that have the potential to include producing zones were selected. In this sub-cluster, the relationship between production-log data and clusters has been recognized and the production-log data has been reconstructed and extrapolated away from the wellbore. Finally, the production-log data and the cluster data were superimposed at each point in the 3-D seismic cube.

Figure 2 was generated using IRESC techniques (Figure 1). Figure 3 shows both qualitative and quantitative analysis of the performance of the proposed technique. In this study, we have been able to predict the D1-Zone thickness whose its presence is very critical to production from D-Zone. D1-Zone thickness it is in the order of 14 feet or less and it is not possible to be recognized using seismic resolution information which is usually in the order of 20 feet and more in this area.

Figures 4 through 6 show the performance of the IRESC technique for the prediction of classes (potential for production of high and no potential) and also the prediction of Phi*Dh which is a representative of the production zone in Red Reviver reservoirs. We have also been able to precisely predict not only the D-zone which is

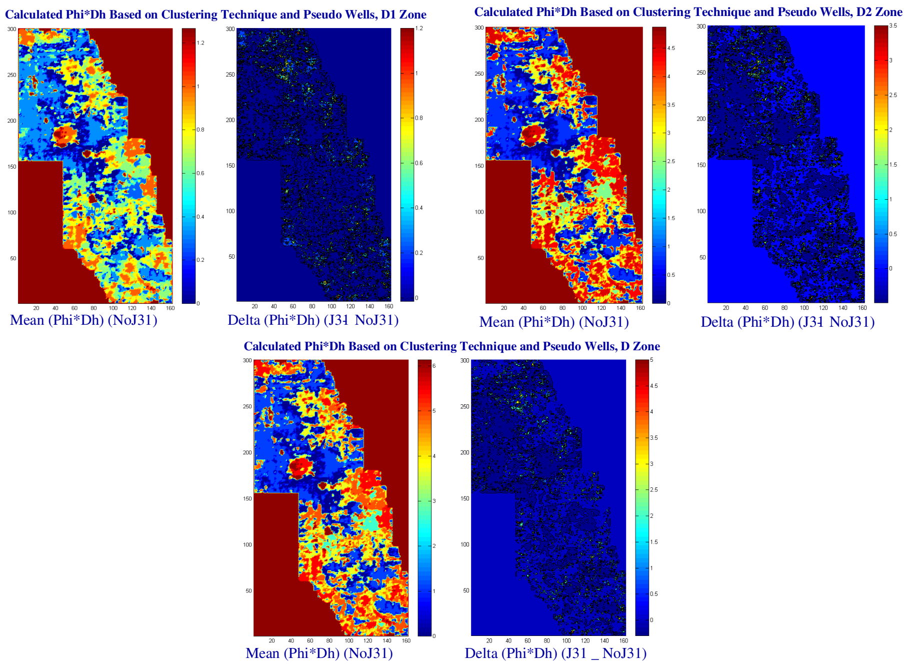


Fig. 6. Performance of IRESC technique for prediction of Phi*Dh for D-Zone, D1-Zone and D2-Zone and error bar at each point before and after drilling a new well

in the order of 50 feet, but both D1-zone which is in the order of 15 feet and D2-Zone which is in the order of 35 feet. The technique can be used for both risk assessment and analysis with high degree of confidence. To further use this information, we use three criteria to select potential locations for infill drilling or recompletion: 1. continuity of the selected cluster, 2. size and shape of the cluster, and 3. existence of high Production-Index values inside a selected cluster with high Cluster-Index values. Based on these criteria, locations of the new wells can be selected.

2.2 Ellenburger Dolomite

Figure 7 shows schematically the flow of information and techniques to be used for intelligent reservoir characterization (IRESC). The main goal will be to integrate soft data such as geological data with hard data such as 3-D seismic, production data, etc. to build a reservoir and stratigraphic model. In this study, we will only concentrate on integrating 3-D seismic data and production data to build similarity cubes based on clustering techniques. We will use 3-D seismic attributes to find similarity cubes and clusters (clusters can be interpreted as lithofacies, homogeneous classes or similar patterns that exist in the data). Then the relationship between each cluster and production log will be recognized around the wellbore and the results will be used to reconstruct and extrapolate the production log away from the wellbore. The results

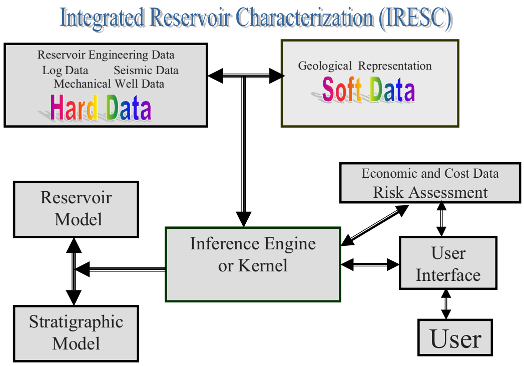


Fig. 7. IRESC

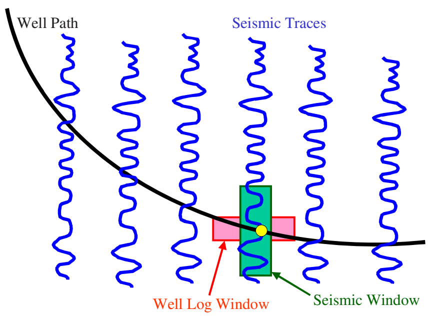


Fig. 8. Well log vs. Seismic

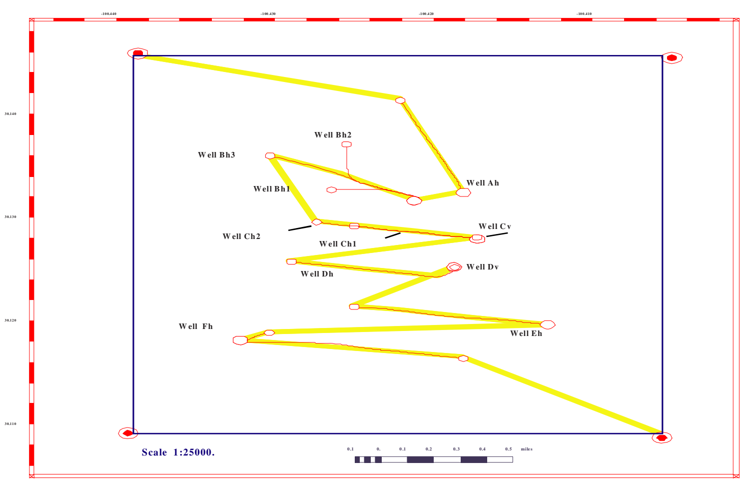


Fig. 9. Well path intersects the Seismic traces

from clustering will be superimposed on the reconstructed production log and optimal locations to drill new wells will be determined.

Implementation: Our example is from a field that produces from the Ellenburger Dolomite. The Ellenburger is one of the most prolific gas producers in the conterminous United States, with greater than 13 TCF of production from fields in west Texas. The Ellenburger Dolomite was deposited on an Early Ordovician passive margin in shallow subtidal to intertidal environments. Reservoir description indicates the study area is affected by a karst-related, collapsed paleocave system that acts as the primary reservoir in the field studied [Adams et al. 1999 and Levey et al. 1999].

The 3-D seismic volume used for this study has 3,178,500 data points. Two hundred, fifty-two data points intersect the seismic traces. Finally, 89 production log data points are available for analysis (19 production and 70 non-production). In the training phase for clustering and mapping purposes, a sub set of seismic data, which is representative of the 3-D seismic cube, production log data and area of interest, was selected. The subset was designed as a section passing through all the wells as shown in Figure 8. Section has 105,150 data points. However, 35,751 data points are selected for clustering purposes, representing the main focus area. Figure 9 shows the schematic diagram of how the well path intersects the seismic traces. For clustering and mapping, there are two windows that must be optimized, 1) the seismic window and 2) the well log window. There are over one hundred seismic attributes that exist. Therefore, an optimal number of seismic attributes needs to be recognized. An optimal number of clusters to be recognized must to be chosen depending on the nature of the problem. Figure 10 shows the iterative technique that has been used to select an optimal number of clusters, seismic attributes, and optimal processing windows for the seismic section shown in Figure 9. Knowledge of experts such as geological layering has also been used to constrain the maximum number of clusters to be selected. In this study, six attributes have been selected (Raw Seismic, Cosine Instantaneous Phase, Instantaneous Amplitude, Instantaneous Phase, Instantaneous Frequency, and Integrate Absolute Amplitude) out of 19 attributes calculated. Ten clusters were recognized and a window of 1 has been used as the optimal window size for the seismic and a window of 3 for the production log data. Finally, based on qualitative analysis, specific clusters that have the potential to be in the producing zones are selected (Figure 11). Figure 11 shows the cluster (group of clusters) selected using three different techniques, k-means (statistical), neural network, fuzzy c-means (fuzzy logic) clustering. By comparing Figures 11.a, Figure 11.b, and Figure 11.c one can conclude that (in this study) that all the techniques predicted the same cluster (or group of clusters) based on our objectives (producing zones). However, this may not always be the case. In addition, the information that can be extracted based on different techniques will be different. For example, clusters using classical techniques will have sharp boundaries whereas those generated using the fuzzy technique will have fuzzy boundaries.

Based on the clusters recognized in Figure 11 and the production log, a subset of the cluster has been selected as shown in Figure 12. In this sub-cluster, the relationship between production log and clusters has been recognized and the production log has been reconstructed and extrapolated away from the wellbore. Finally, the production log and the cluster data are superimposed at each point in the

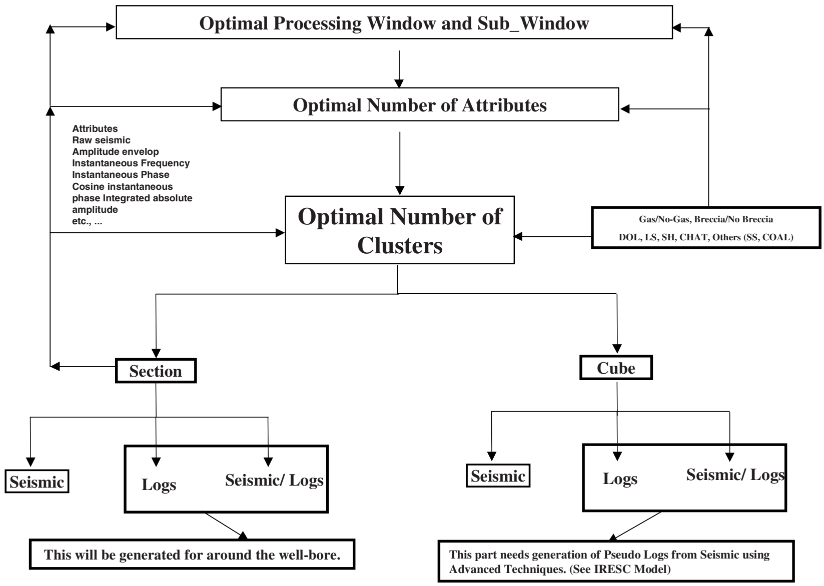


Fig. 10. Criteria for optimization of IRESC

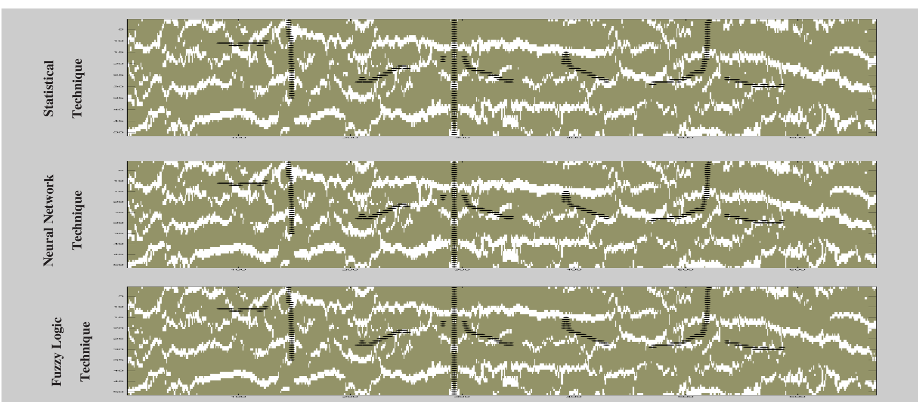


Fig. 11. Clusters recognized by three different techniques (possible oil and no-oil clusters)

3-D seismic cube. Figure 13 shows a typical time-slice of a 3-D seismic cube that has been reconstructed with the extrapolated production log and cluster data.. Three criteria have been used to select potential locations for infill drilling or recompletion, 1) continuity of the selected cluster and production log, 2) size and shape of the cluster and production log, 3) existence of the production log inside the selected cluster. Based on these criteria locations of the new wells are selected and two such locations are shown in Figure 13, one with high continuity and potential for high production and one with low continuity and potential for low production. The neighboring wells that are already in production confirm such a prediction as shown in Figure 13

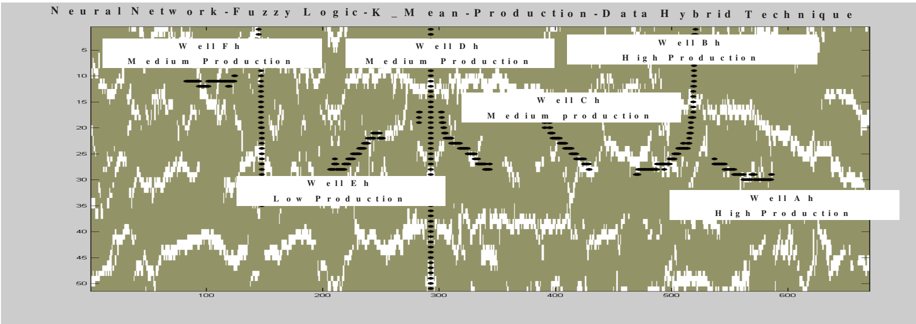


Fig. 12. Clusters recognized by integration of three different techniques (possible oil and no-oil clusters**, Figure 11**)

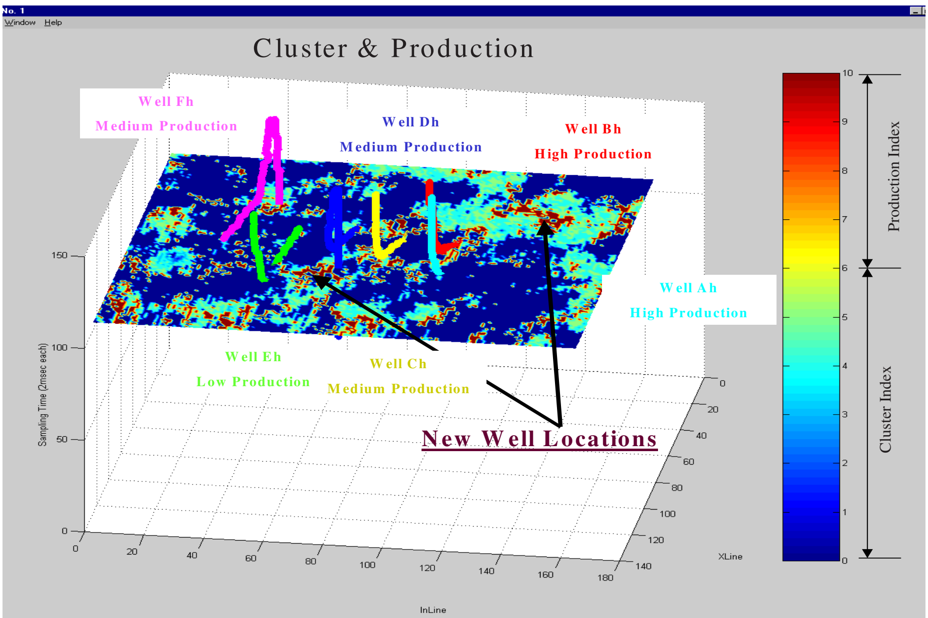


Fig. 13. Typical time-slice of 3-D Seismic cube that has been reconstructed with the extrapolated production log and cluster data

3 Conclusions

This paper addressed the key challenges associated with development of oil and gas reservoirs, given the large amount of by-passed oil and gas and the low recovery factor in many reservoirs. We are proposed the next generation of Intelligent Reservoir Characterization (IRESC) tool, based on Soft computing (as a foundation for computation with perception) which is an ensemble of intelligent computing methodologies using neuro computing, fuzzy reasoning, and evolutionary computing. The true benefit of soft computing, which is to use the intelligent techniques in combination (hybrid) rather than isolation, has not been demonstrated in a full extent. This section will address two particular areas for future research: hybrid systems and computing with words.

References

[1] Adams, R.D., J.W. Collister, D.D. Ekart, R.A. Levey, and M. Nikravesh, Evaluation of gas reservoirs in collapsed paleocave systems, Ellenburger group, Perminan basin, Texas, AAPPG Annual Meeting, San Antonio, Tx, 11-14 April, 1999.

[2] Levey, R., M. Nikravesh, R. Adams, D. Ekart, Y. Livant, S. Snelgrove, J. Cllister, Evaluation of fractured and paleocave carbonate reservoirs, AAPPG Annual Meeting, San Antonio, Tx, 11-14 April, 1999.

[3] Nikravesh, M., F. Aminzadeh, and L. A. Zadeh (2003a), *Intelligent Data Analysis for Oil Exploration,* Developments in Petroleum Science, 51; ISBN: 0-444-50685-3, Elsevier (March 2003).

[4] Nikravesh, M., L.A. Zadeh and V. Korotkikh (2003b), *Fuzzy Partial Differential Equations and Relational Equations: Reservoir Characterization and Modeling*, to be published in the Series Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer (Expected, August 2003).

[5] Nikravesh, M., F. Aminzadeh and L.A. Zadeh (2001a), Soft Computing and Earth Sciences (Part 2), Journal of Petroleum Science and Engineering, Volume 31, Issue 2-4, January 2001; Special Issue.

[6] Nikravesh, M., F. Aminzadeh and L.A. Zadeh (2001b), Soft Computing and Earth Sciences, Journal of Petroleum Science and Engineering, Volume 29, Issue 3-4, May 2001; Special Issue, 2001b.

[7] Nikravesh, M., R. D. Adams and R. A. Levey (2001c), Soft computing: tools for intelligent reservoir characterization (IRESC) and optimum well placement (OWP), *Journal of Petroleum Science and Engineering, Volume 29, Issues 3-4, May 2001, Pages 239-262.*

[8] Nikravesh, M. and F. Aminzadeh (2001), Mining and fusion of petroleum data with fuzzy logic and neural network agents, *Journal of Petroleum Science and Engineering, Volume 29, Issues 3-4, May 2001, Pages 221-238.*

[9] Wong, P. M., F Aminzadeh, and M. Nikravesh (2001), Soft *Computing for Reservoir Characterization*, in Studies in Fuzziness, Physica Verlag, Germany

[10] P.M. Wong and M. Nikravesh (2001), A thematic issue on “Field Applications of Intelligent Computing Techniques,” Journal of Petroleum Geology, 24(4), 379-476; Special Issue.

A Fuzzy Approach to the Study of Human Reliability in the Petroleum Industry

J. Domech More, R. Tanscheit, M.M. Vellasco, M.A. Pacheco, and D.M. Swarcman

Dept. of Electrical Engineering, Catholic University of Rio de Janeiro

Abstract. This work presents a methodology for characterization of human reliability based on fuzzy sets concepts, aiming at reducing the possibility of human errors in oil refineries. It is suited for operation, maintenance and inspection activities in oil production and distribution units and is based on the API-770 guide for reduction of human errors, which identifies 64 performance factors. In order to assess the possibility of a human fault, the model analyses the elements that interact with each operator. It is possible, by this methodology, to obtain a human reliability index, to find out the problems that may constitute causes of human errors and to devise strategies for the control of potentially adverse impacts of interactions that add uncertainty and complexity to processes.

1 Introduction

It is estimated that 60-80% of causes of technological accidents are related, among others, to faults in equipment design, instrument calibration, execution or interpretation of written and oral procedures and organizational factors [1]. In general, resources are applied mostly in equipment reliability and process optimization through real time automation; thereafter the human being must adapt to the process. However, priority should be given to the study of human reliability and then, if possible, adapt the equipment and the working environment to the capabilities, limitations and necessities of the human being. In the petrochemical industry, human errors have contributed, directly or indirectly, to many accidents. The American Petroleum Institute and its members, for example, recognize the importance of reducing human errors so that safety, productivity and quality of production processes are enhanced. The high risk factors in oil industry, and the fact that many working procedures are carried out in a hostile environment justify the need for a rigorous approach to human reliability. Human error usually arises from inadequacies of the system design, such as task complexity and error-likely situations. Humans have limited capacities for perceiving, attending, remembering, calculating, etc. Errors are likely to occur when the task requirements exceed these capacity limitations. Some general situational characteristics may predispose operators to make errors. For example: inadequate workspace and training procedures, poor supervision. Errors may also reflect individual differences. These differences are human attributes of the worker, such as abilities and attitudes. Important individual factors are susceptibility to stress and inexperience, which can produce a tenfold increase in human error possibility. Human reliability depends on physical and mental factors. Some of the physical elements to consider include motor skills (eye-hand coordination, dexterity, flexibility, etc.),

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 415–424, 2007.
springerlink.com

vision capabilities (color discrimination, near and far field visual acuity and field of vision), general physical condition and stamina to work for the required periods in given environment (to climb, kneel, bend, etc). The mental or cognitive process in a typical task includes sensation, perception, short-term and long-term memory, decision making and a resulting action. Studies in human reliability date back to the fifties [2] and are generally divided into two generations [3]. The first-generation methods are characterized by comparing human performance to that of a machine, associating probabilities of success or of a fault to the operators' actions. The second-generation methods extend the analysis of human reliability to cognitive systems, by considering decision levels, diagnosis processes, dexterity, knowledge and organizational factors. The following methods for the analysis of human reliability can be singled out: THERP (Technique for Human Error Rate Prediction) and ATHEANA (A Technique for Human Error ANAlysis), from first and second generations, respectively. These are probability-based methods, which makes it difficult to establish a precise model for human fault prediction, since a large quantity of data is needed for mapping all the uncertainties inherent to human behavior. Most of the traditional methodologies show a disregard for system complexities, and assume that the formal properties of mathematics correspond to some existing relationships characteristic to the system under investigation. For example, an uncertainty due to vagueness is often modeled (if not disregarded) as being of stochastic nature. The probabilistic analysis is used in order to analyze system reliability rationally, i.e., objectively. It is based on the assumption that an equipment or human failure occurs at random. A failure of a single component may occur at random; a human error, however, does not necessarily occur in that way, since a human factor is composed of a large number of components (attributes or performance shaping factors) and its functional structure is very complex. By using the probabilistic approach, where the equipment and procedure are qualified, it is assumed that the operator correctly implements all the procedure's provisions and thus isolates the human factor elements. Human centered systems are very complex and therefore difficult to analyze. At least three different types of uncertainty are inherent to such systems: inaccuracy, randomness, and vagueness. Traditional scientific thinking, based primarily on the Aristotelian logic, is oriented towards exact quantitative methods of analysis. Such methods (and corresponding models) equate uncertainty with randomness only and fail to recognize the human and system based uncertainties due to vagueness. According to Zadeh's principle of incompatibility, at a high level of complexity, precision and significance (of the statements about the system's behavior) become almost mutually exclusive characteristics. Therefore, an attempt to make precise and yet significant statements about the complex relationships between people, machines and environments may be an illusive task, and the traditional modeling methods may not have much relevance. Zadeh points out that, although conventional mathematical techniques have been and will continue to be applied to the analysis of humanistic systems, it is clear that the great complexity of such systems call for approaches that are significantly different in spirit as well as in substance from the traditional methods – methods which are highly effective when applied to mechanistic systems, but are far too precise in relation to systems in which human behavior plays an important role. Furthermore, in order to be able to make significant assertions about the behavior of humanistic systems, it may be necessary to abandon the high standards of rigor and precision expected from mathematical analyses of well structured mechanistic systems, and become more tolerant of approaches which are approximate in nature. The theory of fuzzy sets attempts at constructing a conceptual framework for a systematic treatment of vagueness and uncertainty due to fuzziness in both quantitative and qualitative ways. Such framework is very appropriate for dealing with human reliability, since in man-machine systems there is always a degree of fuzziness, due to the inability to acquire and process an adequate amount of information about systems, vagueness of the relationship between people and working environments and vagueness of the human thought process. This work deals with a fuzzy-based system for the analysis of human reliability in operation, maintenance and inspection activities in industrial and production processes where the human error may have a great impact on safety and on the environment. The analysis of human reliability is carried out through the evaluation of several factors that affect human performance. The importance of each of these factors (called Performance Shaping Factors – PSFs) varies in accordance with the operator's activity [4]. The degree of importance of each factor is obtained through interviews with experts, which is one justification for employing fuzzy sets concepts. Fuzzy Set Theory and Fuzzy Logic provide the necessary tools for building approximate models of the real world and for dealing with linguistic variables. Thus, it is possible to represent mathematically subjective measures, affected by uncertainties and expressing the personal opinions of experts. A set of linguistic rules can be obtained which can be analyzed in the search for configurations with the least error possibility. The methodology, together with results of its application to a unit of Petrobras (Brazilian Company) is described in the next section. Section 3 concludes the work.

2 Methodology and Case Study

The characterization of human reliability in a process has the objective of establishing the degree of attendance of Performance Shaping Factors (PSFs), which can be of human, technical or environmental types. Such characterization may be carried

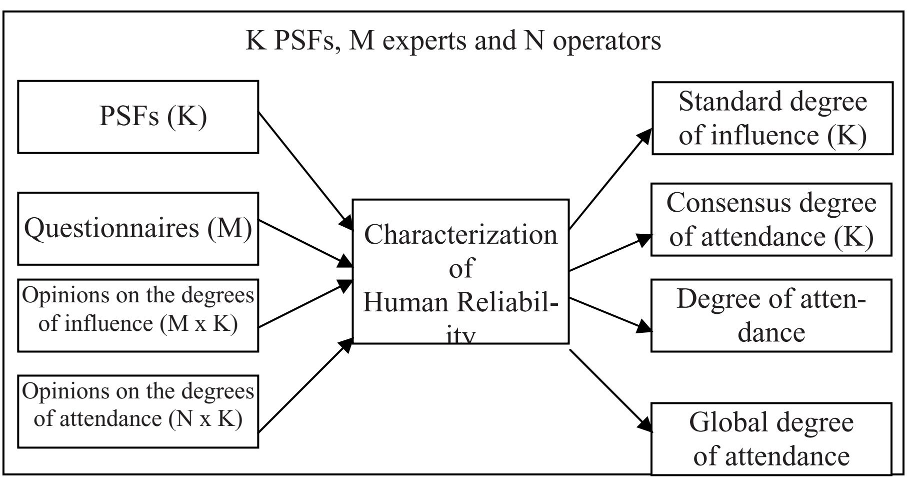


Fig. 1. Characterization of Human Reliability in a Process

out in any category: operation, maintenance or inspection. The method for determination of the degree of attendance of a set of attributes begins with the selection of PSFs that affect the human being. After that, experts specify, through a questionnaire, the influence of each of these factors on human reliability. By aggregating the experts' opinions about the PSFs' influences, weighed by the degree of importance of each expert, a standard degree of influence is established for each PSF. In a second stage, operators' opinions about the attendance of each of the PSFs are taken, also through a questionnaire. After normalization of those opinions (with respect to the maximum values), a degree of attendance of each PSF by each operator is determined. A global degree of attendance may also be determined. This procedure is depicted in the diagram shown in Fig. 1.

The methodology consists of the following stages.

Identification of the object of evaluation and of the set of PSFs

Human behavior is affected by many PSFs, e.g., task environment, stress, motivation, etc. Some of them are external to the person and some are internal. The external PSFs include the entire working environment, especially equipment design and written procedures or oral instructions. The internal PSFs represent the person's individual characteristics – skills, motivation and expectations that influence the performance. Psychological and physiological stresses result from a working environment in which the demands placed on the operator by the system do not conform to his capabilities and limitations. In the petroleum industry, the API-770 standard lists 64 attributes (PSFs) that affect human behavior [5]. These are employed in this work.

Establishment of a committee of decision-makers

This is one of the most important steps in the methodology, since the quality of information will depend on the experts' hierarchical levels. In this work, seventeen experts of high hierarchical level have been selected. They are known for their experience, knowledge and practice in ultrasonic nondestructive examinations.

Establishment of the relative importance of each expert

This step is accomplished through an Expert Profile Identification Questionnaire (EPIQ), which consists of a set of questions with the objective of evaluating each expert's importance and thus assigning him a weight [6]. This is always relative to the other experts' weights and will have an influence on the final result.

Choice of linguistic values for the evaluation of human reliability attributes

This step consists of choosing linguistic terms, or values, for the evaluation of the importance (by experts) and of the attendance (by operators) of each PSF. Terms used in this work for importance: *critical, very important, important, of little importance and not important.* Terms for attendance: *excellent, very good, good, regular, bad.* All terms are associated to triangular fuzzy sets, defined by three parameters.

Degrees of importance and attendance of each PSF

This step consists of obtaining from the selected experts, through questionnaires, their opinions on the importance of each PSF, as well as obtaining from operators their judgment on the degree of attendance of each attribute. From the operational viewpoint, the experts' and operators' opinions are expressed by numbers associated to the previously defined linguistic terms.

Fuzzy treatment of the data provided by the experts and by the operators in the evaluation of each PSF considered

In this step, the individual prognoses from each expert for the PSF are aggregated, generating a consensus for each evaluated attribute. The Hsi-Mei-Hsu and Chen-Tung-Chen´s model [7] is used to pool the expert's opinions: an aggregation procedure, called similarity aggregation method (SAM), is used for combining the opinions of each expert. The opinion of an expert i is expressed by a fuzzy set denoted by $\tilde{A}i$. The agreement degree (or similarity measure) $S\,(\tilde{A}i, \tilde{A}j)$ between two experts i and j can be determined by the proportion of the consistent area to the total area:

$$S(\tilde{A}i,\ \tilde{A}j) = \frac{\int_x \min(\mu_{\tilde{A}_i}(x), \mu_{\tilde{A}_j}(x))dx}{\int_x \max(\mu_{\tilde{A}_i}(x), \mu_{\tilde{A}_j}(x))dx} \tag{1}$$

Once all e agreement degrees between experts are measured, an agreement matrix (AM) can be built, giving an insight into the agreement between the experts.

$$AM = \begin{bmatrix} 1 & S_{12} & S_{13} & \ldots\ldots & S_{1n} \\ S_{21} & 1 & S_{23} & \ldots\ldots & S_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ S_{n1} & S_{n2} & S_{n3} & \ldots\ldots & 1 \end{bmatrix} \tag{2}$$

The average agreement degree $AADi$ of expert Ei (i = 1, ..., n) is given by:

$$AAD_i = \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} S_{ij} \tag{3}$$

The relative agreement degree $RADi$ of expert Ei (i = 1, ..., n) is given by:

$$RAD_i = \frac{AAD_i}{\sum_{i=1}^{n} AAD_i} \tag{4}$$

Finally, by weighing the relative agreement degree of each expert k by the degree of importance Gi_k (obtained through a questionnaire), the consensus coefficient for that expert can be calculated.

$$CC_k = \frac{RAD_k * Gi_k}{\sum_{i=1}^{n}(RAD_i * Gi_i)} \tag{5}$$

The aggregated opinions are given by:

$$\tilde{N} = \sum_{i=1}^{n} CC_i \times \tilde{A}_i \tag{6}$$

Defuzzification (by the Center of Gravity method) [8] gives as a result a single value that represents the importance of each PSF. After normalization, and considering all PSFs, a Quality Standard is obtained. The normalized values, shown in the column identified by QS in Table 1, constitute the quality standard of operation activities. The degrees of attendance of three selected operators (OP1, OP2, OP3) are also shown in Table 1.

Table 1. Quality Standard (QS) and degrees of attendance for three operators

Factor	Description	QS	OP1	OP2	OP3
1	Oxygen insufficiency	1.00	1.00	0.80	0.80
2	Threats (of failure, loss of job)	0.98	1.00	1.00	1.00
3	Radiation (physiological effect)	0.98	0.60	0.60	0.40
4	High risk of chemical exposition	0.97	0.80	0.80	0.80
5	Chemical exposition	0.97	0.60	0.60	0.60
6	Pain or discomfort	0.96	1.00	1.00	0.40
7	Duration of psychological stress	0.95	1.00	1.00	0.80
8	Fatigue	0.93	1.00	1.00	1.00
9	Sensibility privation	0.93	1.00	1.00	1.00
10	Stress	0.91	0.60	0.40	0.40
11	State of current practice or skill	0.89	0.80	0.60	0.40
12	Temperature extreme	0.89	1.00	0.80	0.60
13	Previous training/experience	0.89	1.00	1.00	1.00
14	Vibration	0.88	0.80	0.40	0.20
15	Interruption of circadian rhythm	0.87	1.00	0.80	0.60
16	Physical conditions and health	0.86	1.00	1.00	1.00
17	Movement constriction	0.86	1.00	1.00	1.00
18	Motivation and attitudes	0.85	0.80	0.60	0.40
19	Interpretation (decision making)	0.85	0.60	0.80	0.80
20	Conflicts at work	0.83	1.00	0.60	0.60
21	Availability and adequacy of equipment	0.83	1.00	1.00	1.00
22	Emotional state	0.83	1.00	0.80	0.80
23	Monotonous or meaningless work	0.83	1.00	1.00	1.00
24	Procedures required	0.82	0.80	0.60	0.60
25	Task speed	0.81	1.00	1.00	1.00
26	Negative reinforcement	0.81	0.80	0.60	0.40
27	Man-machine interface factors	0.80	0.80	0.80	0.80
28	Rewards, recognition, benefits	0.80	0.80	0.80	0.80
29	Written or oral communications	0.80	0.80	0.80	0.60

Table 1. (*Continued*)

30	Carefulness	0.79	1.00	0.80	0.40
31	Team structure and communication	0.79	0.60	0.60	0.60
32	Shift Rotation	0.79	1.00	1.00	1.00
33	Hunger or thirst	0.78	1.00	1.00	1.00
34	Critical capacity of task;	0.77	0.80	0.80	0.80
35	Number of people	0.76	0.80	0.60	0.40
36	Architectural features	0.74	0.80	0.60	0.60
37	Personality	0.74	0.80	0.60	0.40
38	Culture	0.73	0.80	0.80	0.60
39	Influence of external agents	0.73	1.00	1.00	1.00
40	Perception needs	0.73	0.80	0.60	0.20
41	Supervisors and co-workers' actions	0.71	1.00	1.00	1.00
42	Anticipation needs	0.70	0.60	0.60	0.40
43	Quality of the working environment	0.70	1.00	1.00	1.00
44	Lack of physical exercise	0.69	0.20	0.40	0.60
45	Frequency and repetitiveness	0.69	0.80	0.80	0.80
46	Long, uneventful vigilance	0.69	0.80	0.60	0.40
47	Working hours and breaks	0.69	0.80	0.80	0.80
48	Sample-control relations	0.68	1.00	1.00	1.00
49	Plant policies	0.68	0.80	0.80	0.80
50	Complexity of information	0.68	0.80	0.60	0.60
51	Working methods	0.68	1.00	1.00	1.00
52	Feedback of results	0.68	1.00	0.80	0.40
53	Distractions (noise, glare, movement, color)	0.67	0.80	1.00	0.80
54	Heavy task load	0.67	0.80	0.80	0.80
55	Inconsistent suggestions	0.66	1.00	1.00	1.00
56	Repetition of movements	0.65	1.00	1.00	1.00
57	Suddenness of psychological stress	0.64	1.00	1.00	1.00
58	Memory requirements (long and short term)	0.63	0.60	0.60	0.40
59	Organizational structure (authority, communication channels)	0.62	1.00	1.00	1.00
60	Identification with the group	0.60	0.60	0.20	0.20
61	Knowledge of performance factors	0.58	0.80	0.80	0.80
62	Intelligence level	0.41	0.60	0.60	0.60
63	Calculating requirements	0.40	0.80	0.80	0.80
64	Physical needs (speed, force)	0.30	0.20	0.60	0.40
Human reliability index			**0.86**	**0.79**	**0.71**

Some of PSFs are psychological and behavioural. In this case, it is necessary to make use of psychometric tests, whose results are expressed in linguistic terms (*inferior, medium inferior, medium, medium superior, superior* and *highly superior*). Finally a degree of attendance of the operators to the Quality Standard is obtained.

Defuzzification is performed through the weighted height method [8], where the degree of attendance (OP_k) to each PSF (column identified by "Factor" in Table 1) is multiplied by its quality standard (QS_k) and a weighted average R is obtained:

$$R = \frac{\sum_{k=1}^{n} QS_k * OP_k}{\sum_{k=1}^{n} QS_k} \tag{7}$$

The result of 0.86 above means that there is a distance of 0.14 to the optimum. Therefore, it may be expected that operator 1 will perform better than operator 3, who is 0.29 units away from the optimum.

In Figs 2, 3 and 4 below, the horizontal axes show the 64 factors that contribute to human reliability, while the vertical axes show the degree of attendance of each of those factors. These are arranged in a decreasing degree of importance. In other words, for each factor there is a corresponding degree of importance (Quality Standard) and a degrre of attendance to this standard. It is possible to identify, for each operator, those factors for which the performance is below the Quality Standard.

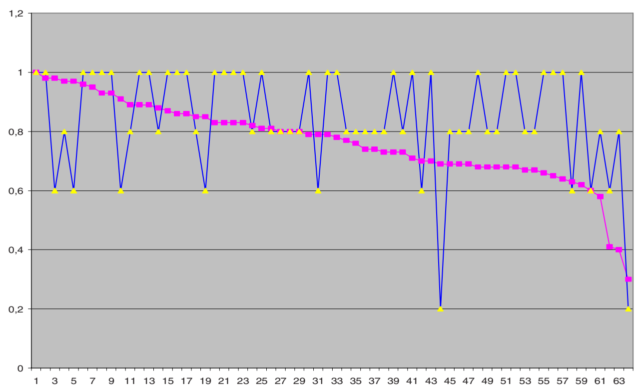


Fig. 2. Operator 1's attendance of the Quality Standard

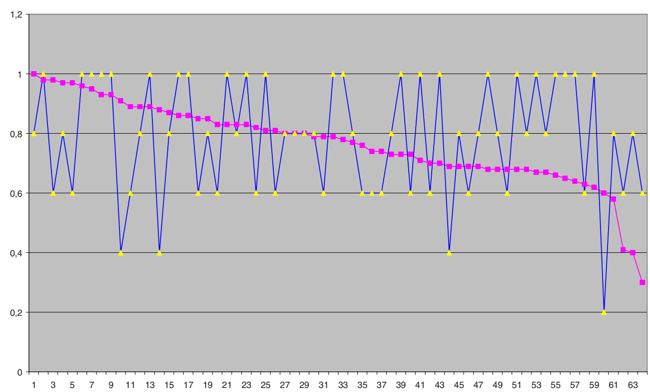


Fig. 3. Operator 2's attendance of the Quality Standard

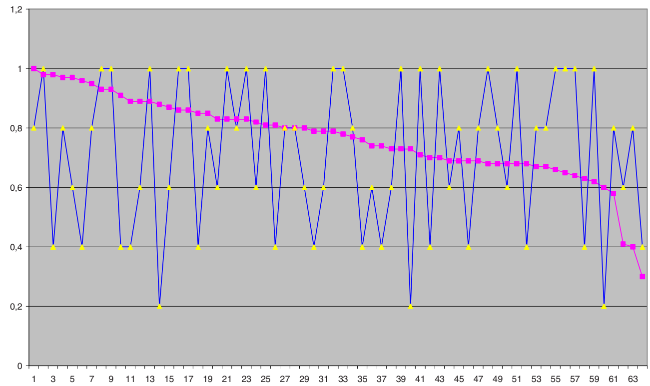


Fig. 4. Operator 3's attendance of the Quality Standard

It can be noticed, for example, that operator 1 performs above the Quality Standard in more factors than operator 3. This is in accordance with the overall reliability index shown in Table 1.

3 Conclusions

A new procedure for a qualitative evaluation of human reliability in the Petroleum Industry has been presented. This procedure makes use of fuzzy sets so that, uncertainties involved in the evaluation procedure can be taken into account.

The findings of this study indicate that it is possible to determine the Quality Standard (QS) for operation, maintenance and inspection activities in oil production and distribution units by using a fuzzy approach, so that a degree of attendance to this QS and an expected value of human reliability in a given physical environment can be obtained.

By knowing the Quality Standard and the degree of attendance of each of the performance shaping factors, resources can be allocated to those aspects which contribute most to human reliability.

References

1. Swain AD (1986) Design techniques for improving human performance in production. 712 Sundown Place SE, Albuquerque, New Mexico, USA
2. Fujita Y (1992) Human reliability analysis: a human point of view. Reliability Engineering and System Safety 38
3. Kim IS (2001) Human reliability analysis in the man-machine interface design review. Annals of Nuclear Energy 28: 1069-1081
4. Swain AD, Guttmann HE (1983) Handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG/CR-1278, USNRC

5. Lorenzo DK (2001) API 770: A guide for the reduction of human errors, EQE International
6. Domech JM (2004) A fuzzy approach to evaluation of human reliability in ultrasonic non-destructive examinations, Ph.D. Thesis, COPPE/UFRJ, Brazil
7. Hsu HM, Chen CT (1996) Aggregation of fuzzy opinions under group decision making. Fuzzy Sets and Systems 79: 279-285
8. Ross TJ (2004). Fuzzy Logic with Engineering Applications 2nd Ed., John Wiley & Sons

Evolutionary Computation for Valves Control Optimization in Intelligent Wells Under Uncertainties

Luciana Faletti Almeida, Yván J. Túpac Valdivia, Juan G. Lazo Lazo, Marco A. C. Pacheco, and Marley M.B.R. Vellasco

ICA: Applied Computational Intelligence Laboratory,
Department of Electrical Engineering, PUC-Rio, Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro CEP 22453-900, RJ – Brazil
{faletti, yvantv, juan, marco, marley}@ele.puc-rio.br

Abstract. This work presents a new decision support system for intelligent wells control considering technical uncertainties. The intelligent control of valves operation tends to become a competitive advantage for reservoirs development. Such control refers to the opening and shutting of the valves that distinguish the intelligent wells. The strategy consists in identifying a valve configuration that maximizes the net present value. The developed system uses Genetic Algorithms, reservoir simulation, Monte Carlo simulation, techniques of sampling variance reduction and uncertainties representation by probability distribution and geologic sceneries. The theoretical concepts applied and the implementation of a system capable of supporting, managing and developing the intelligent fields, constitute an advance to petroleum exploration area. The obtained results demonstrate that the approach given to the problem and the used methodologies deal with the control valves in an efficient and practical way.

Keywords: Intelligent Wells, Intelligent Fields, Evolutionary computation, Uncertainties.

1 Introduction

In projects of the petroliferous exploration area [1], the optimization of the exploitation of a field involves the search for production strategies that are more economically attractive. Following this idea, the engineer intervenes in the wells production by performing operations such as: isolating producer intervals, opening of new intervals, acidifications, fracturing, tests of formation for data collection and other restoring operations. The high costs of these operations, however, especially those in offshore fields with wet completion, can make some of these operations unfeasible, and as a consequence, the field management will not be optimal.

The concept of wells with intelligent completion arises as a technological alternative. This concept is proposed to reduce the cost of the most common restoring operations, as the isolation or the opening of producer intervals. In addition, the monitoring of the production data in real time – flows, pressures and temperature - allows a better field management.

An intelligent completion can be defined as a system capable of collecting, transmitting and analyzing data, which enables the monitoring and the remote drive of

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 425–434, 2007.
springerlink.com

flow control devices. As a consequence, the control of reservoir production is made possible. These technologies, however, are associated with generally high costs, due to the fact that they are considered something new and with fewer field information related to reliability and ways of use. This fact makes the assets managers feel a little fearful in approving the implantation of these technologies, especially because there is not a standard methodology to calculate the benefits of this technology.

Considering the different possible combinations of flow control devices operation, several profiles of production can be generated suggesting the application of an efficient optimization method that allows the discovery of a profile that optimize the production under some criteria.

Thus, the goal of this work is to present a new optimization system, based on genetic algorithms [2], for development and management of petroleum intelligent fields. The system obtains optimal production strategies for the control of the valves in the intelligent wells considering technical uncertainties, such as the risk of valve failure. The developed methodology in this study can be a tool for decision support that helps technicians and managers in the implementation of systems with intelligent completions in their fields. The proposed optimization criterion is the maximization of the Net Present Value (NPV) that causes the increasing of oil production and a maximum delay of water flow into the wells.

Three more sections follow: section 2 shows the optimization system, section 3 shows the case studies, and the section 4 shows conclusions and bibliographic references.

2 Optimization System for the Control of Valves in Intelligent Wells

This section introduces in details the components of the proposed system of valves control optimization, based on Genetic Algorithms. Despite of works addressing some optimization problems for intelligent wells be found in the literature, these works employ classic optimization methods as gradient descent, conjugated gradient or non linear conjugate gradient [3]. These methods, however, have some limitations in the number of variables to be optimized. Thus, if the number of valves is increased or other wells are considered in the optimization, the problem becomes complex and hard to approach by classic methods. In this project, a genetic algorithm is used as the system optimization method, due to the fact that this kind of technique can easily deal with many variables, different and multiple constraints problems. The proposed system consists on three main modules: Optimization Module, Objective Function Computation and Uncertainties Treatment Module (see figure 1). The following sections detail each of these modules separately.

2.1 Optimization Module for Valves Control

This module implements a genetic algorithm model for valve controls optimization. This optimization consists in finding the best configuration of a certain type of valves set, for each predefined time interval, along the total production time. The optimization criterion is the Net Present Value (NPV) computed for the total production time.

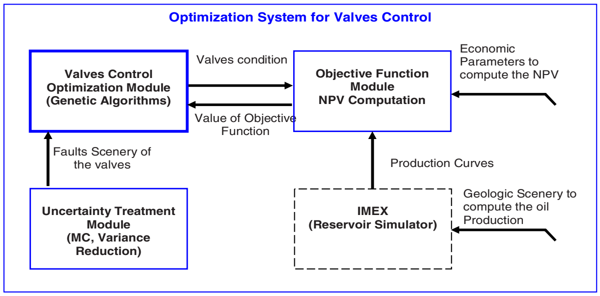


Fig. 1. Main Modules of the proposed system

A set of values that determines a certain configuration of valves is generated by the genetic algorithm as a chromosome. These values must be sent to the objective function module that returns the NPV for the configuration. In the following sub-sections, the modeling of a genetic algorithm for the optimization of valves control is explained.

2.1.1 Chromosome Representation

The genetic algorithm has a population of chromosomes, which represent solutions of the problem. Specifically in this work, a solution refers to a configuration of the existing valves in an alternative, in all time intervals, along the total field production time. Considering that alterations of the opening conditions of the valves can be made at two years intervals, for a total production time of twenty years, the chromosome must represent the total set of valves alterations, for all time intervals in the twenty year period. Existing valves in intelligent completions can be defined as on/off or multi-positions. The chromosome must be able to represent all valve types.
Thus, the chromosome encoding is made in the following way:

1. Each chromosome represents a configuration of valves for every interval being considered.
2. The fixed length of the chromosome is defined by the number of intervals needed to complete the total production time. In the case of valves being altered every 2 years, during twenty years, the chromosome has 10 genes, where each gene represents an interval of 2 years;
3. Each gene holds information about valves conditions in the interval represented by the gene;
4. For on/off valves, the gene represents the valve state as a binary string;
5. For multi-positions valves, the valve state is represented as a real number in the range (0, 1), where 0 represents a totally closed valve and 1 represents a totally opened valve;

Figure 2 shows a chromosome that encodes the on/off valves. In this example, the chromosome represents an alternative with four valves, to be modified every two years, during 20 years of production.

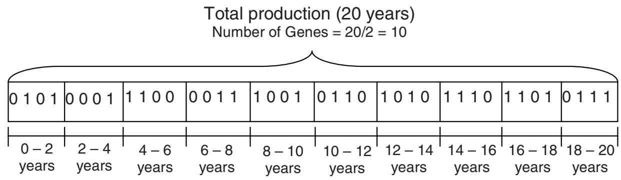


Fig. 2. Chromosome representation of the valve controlling optimization

In figure 2, the first gene of the chromosome indicates, for the first two years of production, that the first and third valves must be closed and the second and fourth valves must be opened.

In order to verify the impact of the configuration mapped by the chromosome, the proposed system uses a reservoir simulator (IMEX) [4]. Thus, the chromosome evaluation occurs in the following way:

Using the chromosome, a file defining the valves configuration for the determined time intervals is written for IMEX. The file is used by the reservoir simulator along with the other files that define the alternative. The simulator, then, provides the water and oil production profile for this configuration. Using the oil production and water affluence, the net present value of the reservoir is calculated as described in section 2.2.

2.1.2 Genetic Operators

In the case of on/off valves, the following crossover and mutation operators were used in the genetic algorithm [5]:

Mutation ⇒ binary; Crossovers ⇒ one point, two points;

In the case of valves with multi-positions, where the chromosome representation uses real numbers, the following mutation and crossover operators were employed:

Mutation ⇒ uniform; Crossovers ⇒ arithmetic, single;

2.2 Objective Function Module

This module calculates a measure for each solution generated by the genetic algorithm in the optimization module. To perform the calculation, this module uses the IMEX reservoir simulator to obtain the oil and water production, and a discounted cash flow model and initial investment to estimate the Net Present Value (NPV).

The NPV is a way of evaluating long term investments. The NPV represents the difference between the incomes and costs expectations (shifted to the present value using the exponential discount parameterized by taxes rates) and the initial investment applied at the initial time.

In the case where the initial investment is related to a previous investment decision, this initial investment can be disregarded, remaining only the need for accounting the discounted cash flow.

The net present value is therefore composed by the difference between the total income present value (VP_R) and the operational cost present value (VP_{Cop}), to which the taxes aliquot I is applied. The present value is computed as the equation (1) shows.

$$VP = \left(VP_R - VP_{Cop}\right)\left(1 - I\right) \tag{1}$$

The income value depends on the oil production $Q(t)$ and on the oil price P_{oil}(t) during the production time. For this analysis, certain market conditions are considered. Therefore, the oil price is a constant P_{oil}. For each time t, the income value can be obtained as shown in equation (2).

$$R(t) = Q(t)P_{oil} \tag{2}$$

In order to obtain the Income Present Value, the exponential discounted sum with discount rate ρ is applied as the equation (3) shows.

$$VP_R = \sum_{i=1}^{T} R(t)e^{-\rho t_i} \tag{3}$$

Where, the value T represents the total production time for the alternative. Each t_i value is the sample which, by simulation, the oil production is obtained.

The operational in this project, only the water handling cost is considered. Thus, the operational cost for time t is given by the following equation (4).

$$C_{OP}(t) = C_w W(t) \tag{4}$$

Where: $C_{OP}(t)$ is the operational cost for t and $C_w W(t)$ is the water handling cost multiplied by the water affluence in each time t.

Finally, the present value of operational cost is obtained by applying, again, the exponential discount sum with discount rate ρ, as the equation (5) shows.

$$VP_{COP} = \sum_{i=1}^{T} C_{OP}(t)e^{-\rho t_i} \tag{5}$$

2.3 Uncertainties Treatment Module

The valves in the intelligent wells are not totally reliable, failures might occur during the operation. This fact decreases the expected benefit in using these types of wells. Thus, in order to obtain a more appropriated decision about the use of intelligent completions the possibility of valve failures must be considered. Next, the probabilistic model and the computing of failures probability, used in this study for valves reliability, are described.

2.3.1 Probabilistic Model of Valves Reliability

Valve failure can affect the performance of intelligent completions, and the impact depends on when the failure occurs and on the valve state after failure. The use of intelligent wells, however, is relatively new, so there is not enough information available about the reliability of the valves they employ.

In order to considering uncertainties about the occurrence of valves failure for a certain alternative of production, the Monte Carlo Simulation (MCS) is normally used [7-8]. The most convenient distribution function to represent the valves reliability, which the failures rate varies in time, is the Weibull distribution [6].

In this work, Quasi Monte Carlo methods [9-12] were used. Using few samples, an adequate sampling of the probability distribution is obtained, causing the reduction of the sampling variance and speeds up the convergence. Therefore, the computational efficiency of the simulation is improved.

2.3.2 Computation of Faults Probabilities

For each defined time interval for the valves adjustments (figure 3), the failure probability p_{ij} is computed for the time interval (t_i, t_j) (assuming that the valve is working at t_i time). As stated in section above, in order to make the calculation of failures probability, a Weibull cumulative failures distribution of sampled values $F(t)$ is used. The equations (6) and (7) show the Weibull cumulative distribution and the failures probability calculation respectively.

$$F(t) = 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\} \tag{6}$$

$$p_{ij} = 1 - \left(\frac{1 - F(t_j)}{1 - F(t_i)}\right) \tag{7}$$

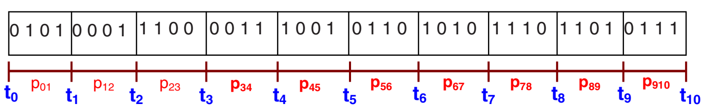


Fig. 3. Chromosome representation for optimization of valves with failures uncertainty

After obtaining the failure probabilities for each time interval, the Monte Carlo simulations begin. For each Monte Carlo iteration (i):

1. From a uniform distribution, a random value R_{ij} is chosen for each valve. This value determines the failure realization at the interval (t_i, t_j);
2. The R_{ij} value is compared with correspondent interval failure probability computed by equation (7). If $R_{ij} \leq p_{ij}$, the valve has failed; otherwise, the valve remains working. Thus, a scenery of failures is created, which is passed to the genetic algorithm (GA);
3. The valves that failed are kept totally opened from the failure time until the end of the production;
4. The GA is initialized and the chromosome is evaluated considering only the intervals with working valves;
5. At the end of the GA execution, the NPV of the best individual is stored;
6. If k < number of Monte Carlo iterations, then go to step 1;
7. If k > number of Monte Carlo iterations, then the best NPVs average found by the GAs is computed.

This way, an average NPV is obtained and this value will show if the usage of intelligent wells, in an alternative is viable even considering the existence of failures of these mechanisms.

3 Case Studies

The suggested system was evaluated using a synthetic reservoir model. This reservoir model consists in a 40x11x3 grid, with blocks length of approximately 50.0x50.0x10.0 meters. The main feature of this model is the existence of three layers with different permeability that are isolated by shale barriers. The geological values of this reservoir are: Porosity: 0,20;

Permeability: 500,0 (md) in i, j directions of the upper layer;
800,0 (md) in i, j directions of the middle layer;
1200,0 (md) in i, j direction of the lower layer;
50,0 (md) in k1 direction, 70,0 (md) in k2 direction and, 120,0 (md) in k3 direction;

The wells alternative used in the tests is composed by a vertical injector with well-head located in (1,6,3) and a vertical producer in (40,6,3). Considering one valve in each completed layer of the injector well, there are three valves in this alternative. Figure 4 shows the alternative described above.

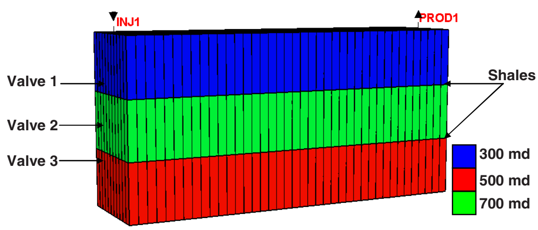


Fig. 4. 40x11x3 Model of the reservoir and alternative used in the experiments of the reservoir 40x11x3

Table 1. Main characteristics of the experiments

Experiment	Exp 1	Exp 2	Exp 3
Generations	120	120	100
Population	75	75	65
Genes	10	10	10
Steady State	0.4	0.4	0.4
Valves type	On/Off	AC	AC
Production time	20 years	20 years	20 years
Time interval	2 years	2 years	2 years
Uncertainties (MC)	Not	Not	500

Table 2. Parameters used for the NPV computation

Parameter	Value
Petroleum price (US$/bbl)	20.00
Royalties	0.10
Taxes	0.34
Discount rate	0.10
Water handling cost (US$/m3)	1.00

Table 1 shows the main characteristics of the experiments. In all tests, the time interval of valves controlling was two years, and twenty years as the total production time. In consequence, there are ten intervals for alter the valves operation. Table 2 describes the parameters used for NPV computation.

Table 3 shows a comparison of the NPV, accumulated oil and water between the base case and controlled case for all experiments. All experiments in this table has a NPV increment, as a consequence, there is an increase of the cumulative oil and a decreasing of the cumulative water. In the experiment 4, a comparison of the mean NPV of the controlled case obtained by GA is shown by considering uncertainties and the NPV of basis case. Also, in this experiment there is an increase of the NPV in the controlled case by showing the viability of using the intelligent technology.

Table 3. Outcomes of the three experiments

Experiment 1	**Basis**	**Controlled**	**Increase %**
NPV (US$)	128 724 826.30	134 374 722.30	4.40%
Cumul. Oil (m^3)	2 025 550.00	2 130 100.00	5.10%
Cumul. Water (m^3)	740 325.00	622 055.00	-15.90%
Experiment 2	**Basis**	**Controlled**	**Increase %**
NPV (US$)	128 724 826.30	134 650 141.00	4.60%
Cumul. Oil (m^3)	2 025 550.00	2 139 630.00	5.60%
Cumul. Water (m^3)	740 325.00	609 469.00	17.00%
Experiment 3	**Basis**	**Mean of Controlled**	**Increase %**
NPV (US$)	128 724 826.30	132 503 151.80	3.00%

From the experts' point of view, the multi-positions valves are the most accurate. This fact is verified in Table 3; for this valves type was able to obtain more NPV increasing, besides, the reality of modeled valves was better represented. Next, some details of experiment 2 are shown:

Table 4 shows the best valves configuration found by GA of experiment 2. The time is expressed in years.

Figure 5 illustrates the GA performance for the experiment 2. It is possible to note that at end of iterations; the genetic algorithm still not converged, indicating the possibility of achieving better results.

Table 4. Best valves configuration found by experiment 2

Time	Valve 1	Valve 2	Valve 3	Time	Valve 1	Valve 2	Valve 3
0-2	0.9	0.6	0.0	10-12	0.6	0.0	0.2
2-4	0.8	0.1	0.2	12-14	0.1	0.1	0.0
4-6	0.7	0.1	0.0	14-16	0.8	0.5	0.1
6-8	0.3	0.5	0.5	16-18	0.4	0.8	0.4
8-10	0.7	0.1	0.2	18-20	0.3	0.2	0.1

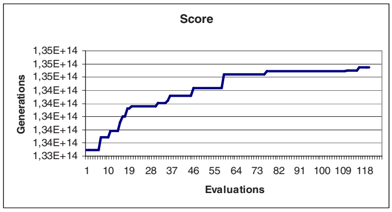


Fig. 5. GA Performance of experiment 2

In figures 6 and 7, the water saturations after twenty years of production are illustrated for basis and controlled case. It can be observed that, for the controlled case, the water saturation is more equalized into three layers, leading to higher oil recuperation in the field.

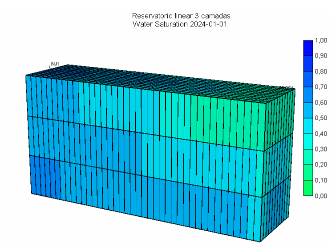


Fig. 6. Field condition – basis case

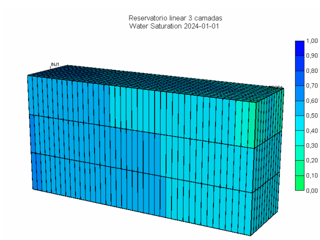


Fig. 7. Field conditions controlled case

4 Conclusions

Although the outcome of the results were obtained from a single reservoir model, it expressively indicates that the evolutionary computation techniques can become useful tools for decision support to find control strategies of the valves in the intelligent wells proving the objective of this study. The obtained values in experiments indicate significant profits in using the intelligent completions in question, with increase recuperation factor of field, reduction of the water produced volumes and the increase of the well life time.

The proposed chromosome representation obtains a formulation of a control strategy for all valves present in an alternative, for any desired time interval. For all valves representation used, profits were obtained if compared with the basis case. The introduction of the intelligent completions into injector well brings expressive profits in the reduction of water production. In experiment 2, the water volume was reduced in 17%. By observing the performance curve, in the experiment 2, using valves with multi-positions, it is noticeable that the optimization will still be able to find better results, by modifying the GA parameters.

In experiment 4, where technical uncertainties are considered, a better average of NPV compared with the NPV of basis case is reached, this fact shows that the effort in investing in intelligent wells technology is still valid, even having the possibility of imperfections in the valves, there is an increase in the NPV alternative. Thus, the obtained results with this type of analysis involve a more qualitative value, indicating if the field has potential of significant profits using the intelligent completion. More experiments on real cases has been done, but the results are not finished yet.

References

1. Crichlow, H. B.: Modern Reservoir Engineering - A Simulation Approach, Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632, 1977.
2. Goldberg, D. E.:Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, 1989.
3. Yeten, B.: Optimum Deployment of Nonconvencional Wells – PhD Tesis, Stanford 2003.
4. CMG. IMEX Advanced Oil/Gas Reservoir Simulator Version 2000 User's Guide. Computer Modelling Group LTD., Calgary, Alberta, Canada, 746pp. 2000:a.
5. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996).
6. Abernathy. R. B.: The New Weibull Handbook. Published by the author. ISBN 0-9653062-1-6, fourth edition, 2000.
7. Boyle, P., Broadie, M., Glasserman, P.:Monte Carlo Methods for Security Pricing, Journal of Economic Dynamics and Control, Vol.21, pp.1267-1321, 1997.
8. Hammersley, J. M. e Handscomb,D.C. :Monte Carlo Methods. Methuen & Co.,Londres, 1964.
9. Wang, Xiaoqun, Variance Reduction Techniques and Quasi-Monte Carlo Methods, Journal of Computational and Applied Mathematics, Nº 132, 2001, pp.309-318.
10. Sobol, I.M., On the Distribution of Points in Cube and the Approximate Evaluation of Integrals. U.S.S.R. Computational Mathematics and Mathematical Physics, Vol. 7(4), 1967, pp. 86-112.
11. Halton, J.H., On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multidimensional Integrals, Numerische Mathematik, Nº 2, 1960, pp. 84-90.
12. Faure, H., Discrepance de Suites Associees a un Systeme de Numeration (en dimension s) Acta Arithmetica, Nº 41, 1982, pp.337-351.

A Genetic Algorithm for the Pickup and Delivery Problem: An Application to the Helicopter Offshore Transportation

Martín Romero, Leonid Sheremetov, and Angel Soriano

Mexican Petroleum Institute, Av. Lázaro Cárdenas, 152,
Col. San Bartolo Atepehuacan, Mexico D.F., CP 07730, México
{mromerom, sher, vsoriano}@imp.mx

Abstract. This paper is a result of the application of soft computing technologies to solve the pick up and delivery problem (PDP). In this paper, we consider a practical PDP that is frequently encountered in the real-world logistics operations, such as Helicopter Offshore Crew Transportation of Oil & Gas Company. We consider a typical scenario of relatively large number of participants, about 70 persons and 5 helicopters. Logistics planning turns to be a combinatorial problem, and that makes it very difficult to find reasonable solutions within a short computational time. We present an algorithm based on two optimization techniques, genetic algorithms and heuristic optimization. Our solution is tested on an example with a known optimal solution, and on actual data provided by PEMEX, Mexican Oil Company. Currently, the algorithm is implemented as part of the system for simulation and optimization of offshore logistics called SMART-Logistics and it is at a field-testing phase.

Keywords: Genetic algorithm, logistics, pickup and delivery problem.

1 Introduction

Our interest on decision support for the operational activities in Oil & Gas (O&G) industry lays in the fact that at petroleum companies, it is common that senior personnel have to solve logistic problems daily. For instance, the pickup and delivery problem, which is about how to organize the transportation of n goods using m vehicles, and each good has to be transported from an origin to a destination. According to our experience, the staff of O&G companies has to develop solutions based on their experience gained through years of field work; for instance, at PEMEX, PDP problems are currently solved by a group of experienced planners. In this paper, we consider the pickup and delivery problem (PDP) as a model for helicopter platform crew transportation [14]. The PDP is a logistic problem that has attracted the attention of the research community because it is a frequent and practical [12,18,19]. In application scenarios, vehicle routing and scheduling has been studied extensively in the context of truck transport, rail transport and ship routing and scheduling [9,10,11]. But relatively little work has been done on helicopter transportation [16], despite that helicopter transportation represents a costly activity for a O&G company with offshore business as we explain next.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 435–444, 2007.
springerlink.com

Many important oil fields are already located in the offshore zone and the deepwater production technology is going to increase their number. This is the case for Brazil (Campos Basin with the most important oil fields), US (Northern Gulf of Mexico) and Mexico (Campeche Zone of the Golf of Mexico). For instance, at Campos Basin [17], 40,000 people lives in offshore habitation; so, each day, hundreds of employees need to be transported across the platforms before and after their shift. The current mode of transportation for platforms crew is primarily based on helicopters provided by a third party logistic company (3PL). We got interest on this problem because the planning of air transportation of personnel by helicopters has to be done daily, and most of the oil companies do not transport themselves due to the high cost of assets. The smallest helicopters cost approximately $41,000/month and $440 per flight hour; whereas, the largest helicopter, which holds 18 people, costs up to $185,000/month and $2400 per flight hour. We collected actual data from PEMEX E&P division that showed that there can be as much as 900 crew members that needs to be transported across 60 platforms everyday. And, current payable charges to the 3PL is in the range of ~USD160,000 per day. Thus the PDP problem might be relevant to other PEMEX E&P divisions [15] because an important piece of the oil production chain is how to transport people, luggage, and other kinds of cargo.

The paper is organized as follows. First, section 1 describes our approach to the PDP problem. Then, section 2 tests our approach on a synthetic problem which has a known optimal solution. Finally, Section 3 describes some implementation details followed by the Conclusions.

2 Our Approach to the PDP Problem

We define the PDP problem. Let us assume that a delivery company has to deliver n packages using m helicopters. The information on the packages is as follows:

1. Let σ be the set of places where packages have to be picked up or delivered; usually, elements of σ belong to R^2
2. Package x has to be picked up at place $o(x)$ and delivered at place $d(x)$, where $o(x)$ and $d(x)$ belong to σ. We call $o(x)$ the origin and $d(x)$ the destination
3. $\lambda(s_1, s_2)$ is the distance between places s_1 and s_2
4. $w(x)$ is the weight of package x

Regarding the information on a helicopter y, $\kappa(y)$ is the weight capacity of helicopter y and helicopters are restricted to fly at most a τ distance. Additionally, helicopters are restricted to start and finis its trip at $s(y)$, the helicopters capacity is large enough to transport any package. We look for a solution that minimizes the cost of picking up and delivering the n packages. For illustrative purposes, we define the cost as the total distance that is flew by the helicopters. This problem description omits some details that happen in practice; for instance, packages might have to be picked up only in specific time intervals. We decided to present only the most relevant features of the problem because we wanted to know the performance of our algorithm under circumstances where an optimal solution is easy to identify.

PDP problem turns to be a combinatorial problem because we have to make two allocations: allocate packages to helicopters, and to specify the order on which packages have to be picked up. Several techniques have been used to assist the logistics planning ranging from techniques based on mathematical programming, [2], to meta heuristics such as tabu search, simulated annealing, and genetic algorithms [5,6,7,18]. Genetic algorithms (GA) are a well developed field in the optimization literature; for instance, see [6]. Briefly, GA starts with a population of *N* starting solutions. Next, GA generates new *M* solutions by *breeding* solutions in the starting population,. Then, GA selects the best *N* solutions among the *M* starting and new *N* solutions. So, GA gets a new population of *N* solutions that are not worse than the starting solutions. Then GA can do *breeding* again, and get a new population of solutions that is not worse that the actual population. GA breeding is the key for GA performance, and it depends on how solutions are represented; so we will not make a general discussion on GA. Thus we describe only the specific GA used to approach the PDP problem. We decided to use GA because as pointed by Bodin [1] "In my opinion, many of the problems described in the literature oversimplify the ones that occur in practice ...", and also pointed by Fisher[3], "Real vehicle routing problems usually include complications beyond the basic model ...". we thought that practical issues are easier to consider using heuristic/genetic approaches. Although, heuristics usually lack robustness and their performance is very much problem dependent [3, 8].

Our approach to the PDP problem is to solve it by a two stage formulation. At the first stage, we consider a planning problem. Let $T(y)=\{ x_1(y),\ldots,x_{n(y)}(y) \}$ be a sequence of packages that has to be picked up and delivered by helicopter *y*. The planning problem is about how to find the best journey for $T(y)$, where a journey for $T(y)$ consists of a sequence of places that helicopter *y* has to visit in order to pickup and deliver the packages in $T(y)$. Journeys are restricted by the order in $T(y)$; this is, $x_i(y)$ is picked up before $x_{i+1}(y)$, or equivalently, the origin of $x_i(y)$ is visited before the origin of $x_{i+1}(y)$. We propose a heuristic algorithm to the planning problem, algorithm that is described in Subsection 1.2

At the second stage, we consider an allocation problem. Our allocation problem looks for the best allocation of packages to helicopters. This is, we look for the best definition of $T(y)$ for $y=1,\ldots,m$, by minimizing the objective function $Z = \Sigma_{y=1,..,m} \Lambda(T(y))$. Where $\Lambda(T(y))$ is the minimum distance that helicopter *y* has to fly in order to pickup and deliver the packages in $T(y)$. This is, $\Lambda(T(y))$ is defined by solving the planning problem. Let Γ be a function that returns the distance that a helicopter flies during a journey. If a journey is not feasible either by the distance restriction $(\leq\tau)$ or the capacity restriction $(\leq\kappa(y))$, then Γ returns a positive infinite value. So Λ can be defined as $\Lambda (T(y)) = min\{ \Gamma(J) : J$ is feasible journey of $T(y) \}$. Next we approach our allocation problem by minimizing *Z* using a genetic algorithm which find the best definition of $T(y)$ for $y=1,\ldots,m$.

2.1 The Allocation Problem

Let $T(y_1)$ and $T(y_2)$ be two sequences. $T(y_1)\oplus T(y_2)$ is a sequence composed of the elements of $T(y_1)$ placed before the elements of $T(y_2)$; this is, $\{1,2,3\} \oplus \{4,5\} = \{1,2,3,4,5\}$. In contrast, $T(y_1)+T(y_2)$ is a set composed of the elements of $T(y_1)$ and $T(y_2)$; similarly, $T(y_1)- T(y_2)$ is a set composed of the elements in $T(y_1)$ but not in $T(y_2)$. Lastly, $\#(T(y))$ is the number packages in $T(y)$. Next, we describe how a genetic algorithm is used to solve our allocation problem. Start by defining initial values for $T(y)$ for $y=1,\ldots,m$, as follows.

1. Find the package x whose origin, $o(x)$, has the shortest distance to $s(y)$, the place at which helicopter y starts its journey. Make $T(y) = \{ x \}$.
2. Find the package u^* that minimizes the distance between the destination of x and the origin of u^*; where x is the last package inserted into $T(y)$; this is $\lambda(d(x),o(u))$ is minimized with respect to u .
3. If $\Lambda(T(y)\oplus\{u^*\}\oplus\{s(y)\}) \le \tau$, then make $T(y)=T(y) \oplus \{u^*\}$. Otherwise finish, and make $T(y)=T(y)\oplus\{s(y)\}$.
4. If there are free packages to allocate go to 2). Otherwise finish, and make $T(y)=T(y)\oplus\{s(y)\}$.

At the end of the initializing algorithm, we have a solution that satisfies the distance restriction; this is $\Lambda(T(y)) \le \tau$, for $y=1,\ldots,m$. In addition, we have a set of packages that were not allocated to any car. Let $L = \{1,\ldots,n\} - T(1) - \ldots -T(m)$ be the packages not allocated. Once we have an initial solution for the allocation problem. We improve the solution as follows.

1. Select two helicopters randomly; let it be y_i and y_j.
2. Improve the journey of helicopter y_i . Find the best ordering of packages in $T(y_i) + T(y_j) + L$ that satisfy the distance restriction. The best ordering is found by a genetic algorithm described in Appendix A. Denote the best ordering by T^*_i . Since T^*_i has to satisfy the distance restriction $(\Lambda(T^*_i)\le\tau)$, it may happens that some packages of $T(y_i) +T(y_j) + L$ were not assigned
3. Improve the journey for helicopter y_j . Find the best ordering of packages in $T(y_i) +T(y_j) + L - T^*_i$, and denote the best ordering by T^*_j. T^*_j is also found by the genetic algorithm of Appendix A.
4. Update the allocation, make $L = T(y_i) + T(y_i) + L - T^*_i - T^*_j$, $T(y_i) = T^*_i$ and $T(y_j) = T^*_j$ when any of the two following conditions holds. First, when $\Lambda(T(y_i)) - \Lambda(T^*_i) + \Lambda(T(y_j)) - \Lambda(T^*_j) >0$ and $\#(T(y_i)) - \#(T^*_i) + \#(T(y_j)) - \#(T^*_j) =0$; this, is when the number of allocated packages does not change and the distance flew by helicopters decreases. Also, update the allocation when $\#(T(y_i)) - \#(T^*_i) + \#(T(y_j)) - \#(T^*_j) <0$; namely, when the number of allocated packages increases.
5. Repeat from 1) as long as the current solution can be updated

Thus, we got a solution that satisfies the distance restriction. Although, the GA performance depends on the ability to solve the planning problem.

2.2 The Planning Problem

Let $T(y)=\{x_1,\ldots,x_{n(y)}\}$ be a sequence of packages that have to be picked up by helicopter y, which starts its journey at s. The condition that package x_i is picked up before x_{i+1} does not define a journey for helicopter y. For instance, let $T=\{x_1,x_2\}$ be a sequence of packages, where $o(x_1)=o_1$, $d(x_1)=d_1$, $o(x_2)=o_2$, and $d(x_2)=d_2$. Table 1 show that three journeys are feasible to make x_1 to be picked up before x_2, where each row of the Table 1 can be translated into a detailed journey if we know the average helicopter speed, w. For instance, the journey o_1,d_1,o_2,d_2 can be translated into the

Table 1. Journeys associated to the sequence of packages { x1=(o1, d1,), x2= (o2,d2)}

Journey	Distance of the Journey
o_1, d_1, o_2, d_2	$\lambda(s, o_1) + \lambda(o_1, d_1) + \lambda(d_1, o_2) + \lambda(o_2, d_2) + \lambda(d_2, s)$
o_1, o_2, d_1, d_2	$\lambda(s, o_1) + \lambda(o_1, o_2) + \lambda(o_2, d_1) + \lambda(d_1, d_2) + \lambda(d_2, s)$
o_1, o_2, d_2, d_1	$\lambda(s, o_1) + \lambda(o_1, o_2) + \lambda(o_2, d_2) + \lambda(d_2\ d_1) + \lambda(d_2, s)$

following schedule. First, leave from s; then arrive to place o_1 at time $\lambda(s, o_1)/w$ and pick up package 1; and so on.

The number of packages associated to $T(y)=\{ x_1,\ldots, x_{n(y)} \}$ might be greater that $n(y)!$ because $n(y)!$ is the number of journeys where packages are delivered only after all the $n(y)$ packages are first picked up. Since the number of journeys that can be generated from a sequence $T(y)$ is at least $n(y)!$. Then, we used a heuristic algorithm in the sense that we explore only a part of the space of feasible journeys. Our heuristic algorithm is as.

1. Start with the journey $J_1^* = \{ o(x_1), d(x_1) \}$. If $\Gamma(J_1^*) \geq \tau$ then stop, there is not feasible journey; otherwise, let $k=2$ and we have at least a feasible journey.
2. Consider $J^{**}=J_{k-1}^* \oplus \{ o(x_k) \} \oplus \{d(x_k)\}$.
3. Shift $o(x_k)$ to the left in J^{**}, name the result as $J^{\#}$. If $J^{\#}$ satisfies that x_i is picked up before x_{i+1} for all i, then continue. Otherwise go to (9).
4. If $\Gamma(J^{\#}) < \Gamma(J^{**})$ and $\Gamma(J^{\#}) \leq \tau$ then make $J^{**}=J^{\#}$.
5. Shift $d(x_k)$ to the left in $J^{\#}$, name the result as $J^{\#\#}$. If $J^{\#\#}$ satisfies that x_i is picked up before x_{i+1} for all i, then continue. Otherwise go to (8)
6. If $\Gamma(J^{\#\#}) < \Gamma(J^{**})$ and $\Gamma(J^{\#\#}) \leq \tau$ then make $J^{**}=J^{\#\#}$.
7. Repeat from (5)
8. Repeat from (3)
9. If $\Gamma(J^{**}) \geq \tau$ we did not find a feasible journey for k packages; our best journey is J_{k-1}^*. Otherwise, $J_k^* = J^{**}$ is our best journey for k packages
10. If $k=n$, then stop, and our best journey is J_k^*. Otherwise, go to (2)

The best way to test a heuristic solution is to evaluate its performance on a situation where we know the optimal solution, and next we do so.

3 Case study

Figure 1 presents 10 arrows that represent the origin and destination of 10 packages that have to be moved by two helicopters. In order to identify the optimal solution we assume that origins and destinations are over a grid. The complete specification of our problem is as follows. There exists a helicopters and ab packages have to be delivered. τ is set to $2(b+1)$ and helicopter y starts and finishes from $(0, \alpha^*(y-1))$. In addition, each package has a weight of 1 kg, and capacity of each helicopter was set to 2kg. Also,

1. If $(x)mod(b) > 0$, package x starts from $((x)mod(b)-1, \alpha^*(x-1)div(b))$ and finishes at $((x)mod(b), \alpha^* (x-1)div(b))$

2. If $(x)mod(b) = 0$, package x starts from $(b\text{-}1, \alpha^*(x\text{-}1)div(b))$ and finishes at $(0, \alpha^* (x\text{-}1)div(b))$

Figure 1 suggests that optimal solution consists of $T(1) =\{1,\ldots,b \}$, $T(2)=\{b+1,\ldots,2b\}$, … , $T(a)=\{ b(a\text{-}1)+1,\ldots, ab \}$. Thus, $\Lambda(T(i))=2(b\text{-}1)$ and the minimum value of Z is $2a(b\text{-}1)$.

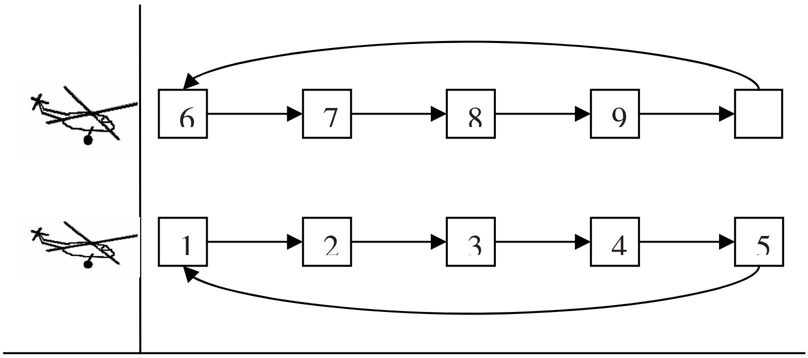


Fig. 1. Graphical representation of the example

We used the optimization algorithm for the cases where b in $\{10,20,30\}$, a in $\{ 2, 4\}$, and α in $\{0.25,0.5,1\}$. The performance of our algorithm was measured by two indices. The distance of packages that are not allocated. $\phi =100^*(2a(b\text{-}1))^{-1} \Sigma_{not\ allocated\ x} \lambda(o(x),d(x))$. And, the distance that helicopters fly compared against the minimum $\varphi = (2a(b\text{-}1))^{-1} \Sigma_{y=1,\ldots,n} \Lambda(o(y),d(y))$. Thus, a solution is better when ϕ is close to zero and φ is close to 100.

Results of Table 2 show that if the number of packages by helicopter is small $(b\leq 10)$, then the genetic/heuristic algorithm finds a solution close to the optimal solution. The solutions, for $(b\leq 10)$, are close to be optimal because all packages are allocated and

Table 2. Results of the proposed genetic/heuristic algorithm. *ab* is the number of packages. *a* is the number of helicopter. α is the vertical separation of the grid points.

Distance that helicopters fly $\varphi=100(2a(b\text{-}1))^{-1}\Sigma_{y=1,\ldots,n} \Lambda(o(y),d(y))$						Distance of not allocated packages $\varphi=100^*(2a(b\text{-}1))^{-1}\Sigma_{not\text{-}allocatedx}\lambda(o(x),d(x))$				
α	*a*	*b* 10	*b* 20	*b* 30		α	*a*	*b* 10	*b* 20	*b* 30
1.00	2	109	108	106		1.00	2	0	57	60
0.50		106	106	101		0.50		0	29	65
0.25		101	106	104		0.25		0	0	32
1.00		110	108	104		1.00		0	62	71
0.50	4	110	109	105		0.50	4	0	47	67
0.25		107	106	105		0.25		0	47	66

the distance traveled by the helicopters is about 10% greater than the optimum. When *b*=20 or 30, the genetic/heuristic algorithm fails to allocate all the packages after 3000 iterations, although we note that this example is difficult because distances that have to fly helicopters are quite heterogeneous. For instance, when *a=2* and *b=30*, there are two packages that have to fly 30 distance units and 58 packages that have to travel only one unit distance. Thus, results of the example suggest that the heuristic algorithm for the selection can be improved.

4 Evaluation Results

The approach presented in this paper has been applied in a system for simulation and optimization of offshore logistics called SMART-Logistics. The application is programmed in C# and makes use of the connectivity infrastructure of .NET of Microsoft, mainly for the communication between distributed components. Figure 2 illustrates a screenshot of the SMART-Logistics interface visualizing helicopter routs generated by the GA for a selected zone of the Zonde of Campeche.

During the first experiments, while testing the GA with PEMEX real data, the flight plans generated by the GA, though providing a viable solution meeting constraints, were outperformed in using helicopters capacity by the manual solution in about 18%. At the second stage, while using complementary simulated annealing instead of heuristic optimization, the manual results were outperformed in about 10% in total cost savings. On the other hand, it took about 5 minutes for a GA to get a solution, while the manual one required the work of 4 people during 4 hours. Although, the algorithm has been tested on the documented evidences from the PEMEX oil company, further work is needed to implement this program as an industrial solution.

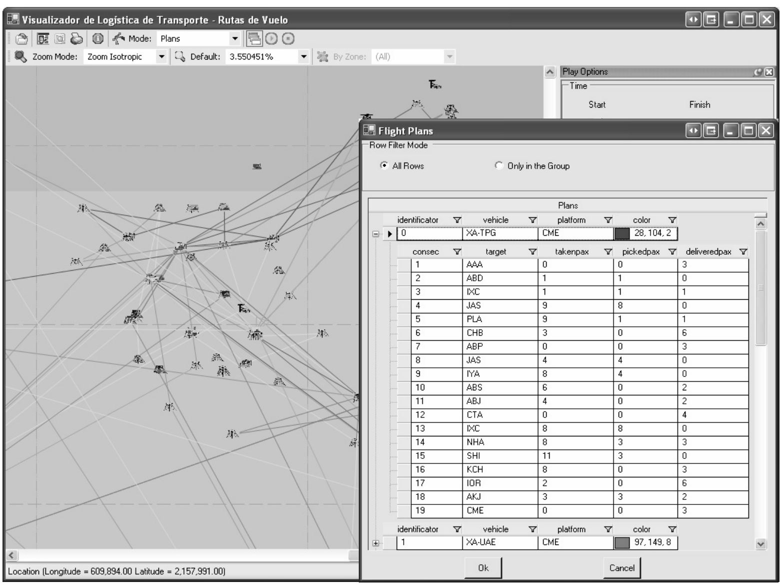


Fig. 2. SMART-Logistics: visualization of the generated routs and helicopter flight plans

The validation process would be benefited by the constant work with real data, which could indicate corrections to improve the system. For instance, it is needed to use data that resembles actual values of the objective functions in order to explore convergence properties of the algorithm. Although this process has already begun, further validation and update of the heuristics on the basis of the analysis of the real problems is carried out.

5 Conclusions

Generally speaking, offshore and onshore O&G exploration and production is driven by the availability of capacity through logistics and transportation in operations. The support of marine vessels of various capacities along with helicopters are needed both in the construction phase of the development of an oil field, and during the production phase when the continual need for the transportation of food, stores, personnel and maintenance equipment to the platforms arises. With growing competition and the continuous installation of offshore facilities the need for traditional logistics knowledge is a need for a better management. This paper illustrates a successful application of soft computing technologies to the pickup and delivery problem arising while transporting the personnel and materials between onshore offices and offshore oil platforms. A solution based on two optimization techniques, genetic algorithms and heuristic optimization is described. As shown, the algorithm guarantees an optimal solution for particular ranges of parameters. Another advantage of the proposed solution is that it also generates a schedule which is not available in a manual fashion.

In future work, we consider that significant savings could be achieved by rerouting. More should be attained with collaboration among helicopters. We also consider the implementation of the NLMI algorithm in order to compare its behavior with the GA and the development of the algorithm of multi-agent optimization inspired in the TRACONET (*Transport Cooperation Network*) system and NEO, probably the best-known example of the use of agent technology, where agents are used for cooperative planning and executing of contingency plans [4,13]. Then, SMART-Logistics application will enable the selection of the most efficient algorithm depending upon the problem specification.

Acknowledgments

Partial support was provided by the Project D.0006 of the IMP. Special thanks to all our colleagues that participated in the development of SMART-Logistics. We would also like to thank our colleagues from the National University of Singapore, especially a Dr. Yew Seng, for their contribution with the simulated annealing algorithm.

References

1. Bodin L.D., Twenty years of routing and scheduling, Operations Research, (1990) 38: 571-579.
2. Conte S.D. and de Boor C. Elementary Numerical Analysis. An Algorithmic Approach. McGraw-Hill Kogakusha: Tokyo. 1972.
3. M. Fisher, K.O. Jornsten, and O.B.G. Madsen, Vehicle routing with time windows: Two optimization algorithms, Operations Research, 1997, 45: 488-492.

4. Giampapa, J., Paolucci, M. and Sycara, K. Agent Interoperation Across Multiagent System Boundaries. In Proc. of Agents2000, Barcelona, Spain, June 3-7, 2000.
5. Gendreau M., Laporte G., and Potvin J.-Y., Vehicle routing: modern heuristics", In: Aarts, E. and J.K. Lenstra (Eds.), Local Search in Combinatorial Optimization, John Wiley & Sons, (1997) 311-336.
6. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Ontario. (1989).
7. Golden B.L., Wasil E.A., Kelly J.P., and Chao I.-M., Metaheuristics in vehicle routing, In: T.G. Crainic and G. Laporte (Eds.), Fleet Management and Logistics, (1998) 33-56.
8. Hang X., Zhi-Long C, Srinivasan R., Sundar A. Solving a Practical Pickup and Delivery Problem. Transportation Science, 2003, 37(3): 347 – 364.
9. P. H. Koo, W. S. Lee, and D. W. Jang. Fleet sizing and vehicle routing for container transportation in a static environment. OR Spectrum, 2004, 26: 193–209
10. S. Kunnathur, U. Nandkeolyar, D. Li. Shipment partitioning and routing to minimize makespan in a transportation network, Computers & Industrial Engineering, 2005, **48:** 237–250.
11. C.-L. Li, G. Vairaktarakis, Ch.-Y. Lee, Machine scheduling with deliveries to multiple customer locations. European Journal of Operational Research, 2005, 164: 39–51.
12. Pankratz G. A Grouping Genetic Algorithm for the Pickup and Delivery Problem with Time Windows. OR Spectrum, Springer Berlin/Heidelberg, Volume 27, Number 1, January, (2005) 21-41.
13. Sandholm, T. An implementation of the contract net protocol based on marginal cost calculations. In Proc. of the National Conference on Artificial Intelligence. AAAI Press. (1993).
14. Savelsbergh M.W.P., Sol M. The General Pickup and Delivery Problem. Transportation Science, (1995) 29: 17-29.
15. Sheremetov L., Contreras M., and Valencia C., Intelligent Multi-Agent Support for the Contingency Management System. Int. J. of Expert Systems with Applications, Pergamon Press, (2004) 26(1): 57-71,
16. Sierksma G. and Tijssen G. A. Routing helicopters for crew exchanges on off-shore locations. Annals of Operations Research, (1998) 76: 261 – 286.
17. da Silva L., Galvao de Medeiros P. R., Orlando Gomes J., Simenc M. C., Joseph J.-Y., Patterson E., Woods D. Helicopter Offshore Safety in the Brazilian Oil and Gas Industry, In Ellen J. Bass, (ed.) Proc. of the 2005 Systems and Information Engineering Design Symposium, (2005).
18. Tam V., Tseng L. C.Y., Effective Heuristics to Solve Pickup and Delivery Problems with Time Windows, In Proc. 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03), (2003) p. 184.
19. Wan-Rong J., Cheng-Yen K., Yung-Jen H. J., Using family competition genetic algorithm in pickup and delivery problem with time window constraints. In Proc. IEEE international symposium on intelligent control, Vancouver BC, Canada, (2002) 496-501.

Appendix A. A Genetic Algorithm

Let $X=\{x_1,\ldots,x_K\}$ be a set of packages that have to be picked up and delivered. We represent a solution for the PDP by a sequence of integers. First, we describe procedures used for the *breeding and mutation* of solutions.

The SWITCH algorithm is used for mutation. The switch algorithm switches the elements r_1 and r_2 of $X=(x(1),x(2),\ldots, x(L))$ where r_1 and r_2 are randomly selected.

The SHIFT algorithm is also used for mutation. The shift algorithm shifts the elements to the left, SWITCH($x(1),\ldots,x(r_1),\ldots, x(L)$) = $(x(1),\ldots, x(r_1-1), x(r_1+1),\ldots, x(L-1), x(r_1)$)) where r_1 is randomly selected.

The MIX Algorithm is used for breeding. Let χ and χ^* be two random permutations of *X*. Initialize *MIX* with an empty sequence.

1. Randomly select *x* from χ.
2. Add *x* to *MIX* and remove *x* from χ and χ^*. Identify the successors of *x* in χ and χ^* call them as $S(x)$ and $S^*(x)$.
3. If *x* has only one successor, say $S(x)$, make $x=S(x)$. Then go to (b). Otherwise, If *x* has two successors; make $x=S(x)$ with probability 0.5, and make $x=S^*(x)$ with probability 0.5. Then go to (b).
4. If *x* has no successor because *x* was the only remaining in χ then stop. Otherwise, go to a).

Finally, we describe the GA used for optimization. Let $\chi_1,\ldots, \chi_K$ be *K* random permutations of *X*. Assume without loss of generality that $F(\chi_1) \geq\ldots\geq F(\chi_{k.})$. Then,

1. Select χ_1 and make $\chi^*_1 = \chi_1$
2. Get $\chi^*_{j+1} = SWITCH(\chi_j),\ 1 \leq j \leq K/4,\ \chi^*_{j+1+K/4} = SHIFT(\chi_j),\ 1 \leq j \leq K/4,\ \chi^*_{j+1+K/2} = MIX(\chi_1, \chi_j),\ 2 \leq j \leq K/4$.
3. Select two permutations from $\chi_1,\ldots, \chi_{K/2}$ and mix them in order to obtain $\chi^*_{j+K/4*3,}$ for $j=1,..K/4$
4. Order $\chi^*_1,\ldots, \chi^*_K , \chi_1,\ldots, \chi_K$ according to $F(.)$; then, replace $\chi_1,\ldots, \chi_K$ with the *K* permutations among $\chi^*_1,\ldots, \chi^*_K , \chi_1,\ldots, \chi_K$ that provide the largest values of $F(.)$. We take χ_1 as the permutation with the largest value of $F(.)$, and we take χ_2 as the permutation with the second largest value of $F(.)$, and so on.

Real Options and Genetic Algorithms to Approach of the Optimal Decision Rule for Oil Field Development Under Uncertainties

Juan G. Lazo Lazo, Marco Aurélio C. Pacheco, and Marley M.B.R. Vellasco

ICA : Applied Computational Intelligence Laboratory, Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, PUC-Rio, Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro, CEP 22453-900, RJ – Brazil
{juan, marco, marley}@ele.puc-rio.br

Abstract. A decision to invest in the development of an oil reserve requires an in-depth analysis of several uncertainty factors. Such uncertainties may involve either technical uncertainties related to the size and economic quality of the reserve, or market uncertainties. When a great number of alternatives or options of investment are involved, the task of selecting the best alternative or a decision rule is very important and complex due to the considerable number of possibilities and parameters that must be taken into account. This paper proposes a new model, based on Real Option Theory, Genetic Algorithms and on Monte Carlo simulation to find an optimal decision rule for alternatives of investment regarding the development of an oil field under market uncertainty that may help decision-making in the following situation: immediate development of a field or wait until market conditions are more favorable. This optimal decision rule is formed by three mutually exclusive alternatives, which describe three exercise regions through time, up to the expiration of the concession of the field. The Monte Carlo simulation is employed within the genetic algorithm to simulate the possible paths of oil prices up to the expiration date. The Geometric Brownian Motion is assumed as stochastic process for represents the oil price. A technique of variance reduction was also used to improve the computational efficiency of the Monte Carlo simulation.

Keywords: Real Options, Genetic Algorithms, Monte Carlo Simulation, Latin Hypercube Sampling.

1 Introduction

A decision to invest in the development of an oil reserve requires an in-depth analysis of several uncertainty factors. Such factors may involve either technical uncertainties related to the size and economic quality of the reserve, or market uncertainties (e.g., oil price). Real option theory is a suitable tool, since it deals with uncertainties and the flexibility of management [1] [2], by maximizing the value of the investment opportunity. Considering that the technical uncertainties are known, the analysis of market uncertainties will help decision-making with regard to investing in a field immediately or wait until market conditions are more favorable. When a great number of alternatives of investment are involved, the task of selecting the best alternative, that is,

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 445–454, 2007.
springerlink.com

defining a decision rule, is very important and also very complex, due to the considerable number of possibilities and parameters that must be taken into account.

This paper extends the works of Dias [3] and Lazo [4] by investigating the optimization of a decision rule, with a great number of alternatives of investment (real options), that it maximizes the value of the real option using a hybrid model that combines a Genetic Algorithm with Monte Carlo Simulations and stochastic processes. This paper presents the proposed Genetic Algorithm (GA) [5] [6] model that obtains the optimal investment decision rule for the development of an oil reserve under market uncertainty, particularly with regard to the oil price. This optimal decision rule is formed by three mutually exclusive alternatives which describe three exercise regions through time, up to the expiration of the concession of the field. The Monte Carlo simulation is employed within the genetic algorithm to simulate the possible paths of oil prices up to the expiration date. The market uncertainty is modeled with one stochastic processes: the Geometric Brownian Motion (GBM), which is used in most financial and real options models. A variance reduction technique, called Latin Hypercube Sampling (LHS), is used to improve the computational efficiency of the Monte Carlo simulation.

This paper is organized as follows: Section 2 describes the problem of the optimal exercise of the development option. Section 3 describes how the problem was modeled with the use of a GA, Real Options theory and the Monte Carlo simulation. Section 4 presents the results obtained with the proposed model. Finally, section 5 contains the conclusions of this study.

2 Description of the Problem

The analysis of the option to develop a delimited oil field requires investments, of which the sum and benefits depend on the chosen alternative. Some alternatives have more wells than others; some have a different geometric distribution for the wells. There are also different types of wells (vertical, directional, horizontal, multilateral, etc.) with wide range of investments and benefits.

The combination with other aspects, such as types of platform, ducts, ship, method of maintenance of pressure in the reservoir, etc, makes this a problem of complex optimization.

This work tries to obtain an optimal decision rule to invest in an oil reserve under market uncertainty, particularly with regard to the price of oil. This decision rule is formed by three mutually exclusive alternatives which describe three exercise regions through time, up to the expiration of the concession of the field. Each alternative presents a threshold curve, which is the critical value for optimal exercise of the real option, any value above it determines the optimal exercise of the real option. All the threshold curves together represent the decision rule that maximizes the value of these alternatives or investment options. Traditionally, in order to evaluate each alternative for investment in the oil field, an attempt to maximize the Net Present Value (NPV) [7] [8] is done, where the best alternative represents the one that provides the highest NPV.

In order to get a simple and adequate equation for the NPV, let v be the market value of one barrel of reserve (that is, v is the price of the barrel of reserve). If this reserve price v is directly related with the long-run oil prices, let be q the factor of proportionality [9], so that $v = q P$.

For developed reserve transactions, as higher is the price per barrel of a specific reserve, as higher is the economic quality for that reserve. For a fixed reserve size and fixed oil price, as higher is the factor q as higher is the value of this reserve. So, let q be the economic quality of the reserve, defined as $q = \partial v / \partial P$. The value of q depends of several factors: the permo-porosity properties of the reservoir-rock; the quality and properties of the oil and/or gas; reservoir inflow mechanism; operational cost; country taxes; cost of capital; etc. For details about q see: http://www.puc-rio.br/marco.ind/quality.html. In this case, the NPV equation for the business model may be written as (1):

$$NPV_t = qP_tB - D \tag{1}$$

Where q is the economic quality of the reserve; P is the petroleum price; the current value is supposed to be US$ 20/bbl and the future value is uncertain; B is the estimated size of the reserve; D is the investment for development of the reserve.

For the fiscal regime of concessions (USA, UK, Brazil, and others), the linear equation for the NPV with the oil prices is a very good approximation [10] [11] [12].

In this paper it is assumed that there is only market uncertainty, i.e., that the oil price is the only source of uncertainty.

2.1 Modeling the Uncertainty of Oil Prices with Two Stochastic Processes

This paper considers one stochastic processes for modeling the oil prices: The Geometric Brownian Motion (GBM). Assume that the oil prices follow the popular Geometric Brownian Motion (GBM), in the format of a risk-neutralized stochastic process, that is, using a risk-neutral drift (r –δ) instead the real drift α, is (2):

$$dP = (r - \delta)Pdt + \sigma P dz \tag{2}$$

The GBM for simulating the future oil price P(t), given the current price P(t-1), is then given in (3):

$$P(t) = P_{t-1} \exp\left[\left(r - \delta - 0.5\sigma^2\right)\Delta t + \sigma\varepsilon\sqrt{\Delta t}\right] \tag{3}$$

Where r is the interest rate free of risk, δ is the convenience yield rate of the oil field, σ is the volatility of the oil price, Δt is the length of the time step, $\varepsilon\sqrt{\Delta t}$ is the Wiener increment, where ε is a normally distributed random variable whit a mean of zero and a variance of one.

The model that has been used is an extension of the one presented by Dixit [8], which was adapted to oil projects (for details see Dias [3] [10]).

2.2 Monte Carlo Simulation and Sampling Variance Reduction

Monte Carlo Simulation (MCS) is the appropriate method for problems of higher dimension and/or stochastic parameters, often used to evaluate the expectation of a variable that is function of various stochastic variables, which is not analytically tractable. Therefore, samples are generated from some target probability distribution to create the diverse scenes to be evaluated. To reduce the error of the variable estimation provided by the simulation, the number of samples must be very large to achieve the desired precision. However, the bigger the number of samples, the greater the computational cost. The reduction of the error estimatives is also possible if the deviation standard is reduced.

Latin Hypercube Sampling [13] [14] was suggested as a variance reduction technique, but can also be seen as a screening technique, in which the selection of sample values is highly controlled. The basis of LHS is a full stratification of the sampled distribution with a random selection inside each stratum. The stratification consists of the division of the probability distribution in intervals with equal probability of occurrence. Samples of each interval are selected, in accordance with its probability. LHS is used in this work as a screening technique, it provides the necessary random numbers for the stochastic process that represents the oil price.

Using LHS, an input sample is also generated based on the inverse transform method, given by (4):

$$xh_{i,j} = F^{-1}\left[\left(i + U_i - 1\right)/n_j\right] \quad i = 1, \cdots, n_j, \quad j = 1,, \cdots, k \tag{4}$$

Where: $xh_{i,j}$ is a sample of LHS; i, j are the dimensions of LHS; n, k states of each dimension of LHS, U_i stands for an independent random uniform distribution on [0,1], $i = 1, \cdots, n_j$, and $F^{-1}(U)$, U ∈ (0,1), is the inverse transform for the particular input distribution (in this application, the approximation for the inverse of the cumulative Normal distribution is used)

2.3 The Optimal Decision Rule

The threshold curve, or optimal exercise curve of the option, represent the decision rule for the development of the field. With the simulated oil price P(t) it is possible to estimate the value of developed of a reserve, $V(t) = q * B * P(t)$. The threshold is the critical level that makes optimal the immediate investment to develop the oilfield. This threshold curve is the decision rule to exercise the option (exercise at or above the threshold), which maximizes the real options value. This optimal exercise curve is a function of the time. In this work, an approach of the threshold curve is obtained using a Genetic Algorithmic model.

Figure 1 shows the threshold curve and two paths of the oil price until the expiration of the option. One path reaches the threshold line at the point "W" (at t = 1.2 in this example), so the decision rule is to exercise the investment option at this time (the option is "deep in the money"), the option value at this moment is F(1.2) = NPV = V – D = qP_tB – D. This value is a future value (t = 1.2 year), so to calculate the present

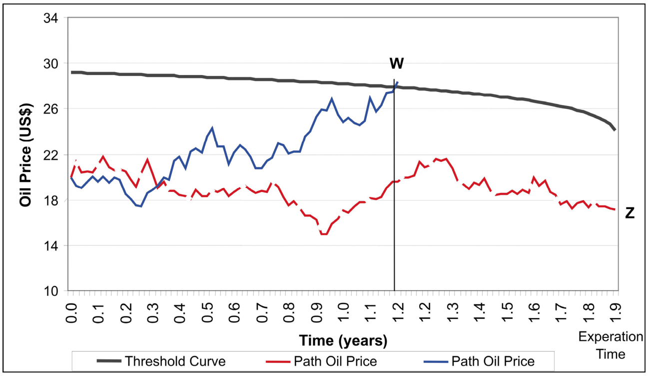


Fig. 1. The Threshold Curve and Exercise of the Real Option Simulated Value

value, it must be multiplied by the discount factor $\exp(-rt)$. The other path pass all the option period without reaching the threshold curve (point "Z"), in this case, the value of the option is zero (expires worthless).

When the number of alternatives increases, the decision rule is formed by the intersection of the threshold curves of each alternative. In this case, the creation of waiting regions between the alternatives is possible [15].

3 Modeling of the Problem

This section describes the proposed model, which integrates the Monte Carlo simulation and the Real Options Theory into a Genetic Algorithm to obtain an optimal decision rule for three alternatives of investment in an oil reserve, considering that the price of oil is uncertain. In order to simulate the oil price, this paper has considered the following parameters:

Expiration Time (T): 2 years;
Discretization Time (Δt): 7 days;
Interest rate free of risk (r): 8 % per year;
Convenience yield rate of the oil field (δ): 8 % per year;
Price volatility (σ): 25 % per year;
Initial oil price (P_0): 20 US$/bbl
Risk-adjusted discount rate(ρ): 0.12 % per year;
Estimated size of reserve (B): 400 MM barrels.
The three alternatives that have been considered present the following parameters:

Table 1. Parameters of the alternatives

	Alternative 1	Alternative 2	Alternative 3
Quality of reserve (q)	8 %	16 %	22 %
Investment for development (D)	400 MM US$	1000 MM US$	1700 MM US$

It is observed that it only makes sense to consider higher investment alternatives if an increase in the economic quality of the reserve is generated, that is, if the investment in more wells, for further drainage of the same reservoir, enhances the economic quality of the reserve that is being developed. In other words, such investment represents the means to extract oil more quickly. For this reason, this type of alternative worths more than the other one with few wells for draining the same reserve.

The problem of determining the optimal decision rule for these three alternatives considering the uncertainty of oil prices is difficult to compute because of its nonlinearity. As a result, the genetic algorithms represent a good choice for finding the optimal decision rule for the oil field development.

The threshold curve may be approximated by means of logarithmic functions in the form of: $a + b\ln(\tau)$ and a free point, which is situated near the expiration time of the option, T, where $\tau = T - t$ and t is an instant of time [3] [4] [16]. The logarithmic function is chosen because it represents a good approximation to the threshold curve obtained by finite differences [3] [16]. For the various alternatives, there are several mutually exclusive threshold curves that determine the exercise regions. This exercise regions are delimited by the intersections between threshold curves. The possible existence of waiting regions [15] between the regions formed by the alternatives is also considered. The waiting regions are approximated by a logarithmic function $a_W - b_W \ln(\tau)$ and also a free point. The value of the coefficients of the above functions (a, b, a_W, b_W) and the free points are determined by the genetic algorithm. Thus, for the case of three alternatives, five regions may be formed (two waiting regions and three exercise regions, one for each alternative) [15].

3.1 Representation of the Chromosome

The chromosome is composed by 5 genes, each gene is formed by three alleles. The three alleles of each gene of the chromosome are real numbers that represent the parameters of the threshold curve of each alternative (variables a e b of logarithm curve and the free point) [3] [4] [16], as well as the parameters of waiting regions (the free point and variables a_W, b_W), as illustrated in Figure 2.

These threshold and waiting curves are subject to a set of constraints that must be satisfied in the process of generating each chromosome, to guarantee the formation of the exercise regions and to reduce the search space of optimal solution.

The domain restrictions are defined from the critical oil price in the expiration, so that in the expiration, the exercise of the option is attractive. The NPV of the alternative of lesser investment must be as minimum zero.

The free points are chosen in each alternative for the same instant, corresponding to 0.1 year. The logarithmic curve begins at instant 0.1 + Δt, where Δt corresponds to the time interval. Therefore, the linear restrictions for each threshold and waiting curve are:

$$a + b\ln(0.1 + \Delta t) \geq \text{FreePoint}$$

$$a_W - b_W \ln(0.1 + \Delta t) \leq \text{FreePoint}$$

$$a + b\ln(0.1 + \Delta t) \geq 0$$

$$a_W - b_W \ln(0.1 + \Delta t) \geq 0$$

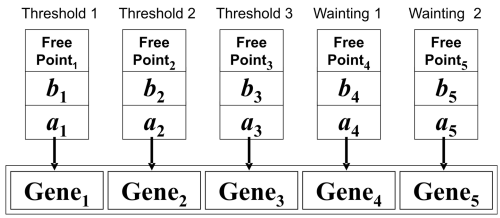


Fig. 2. Chromosome

The genetic algorithm was executed with a population of 1000 individuals, with 100 generations and percentage of population restored (Gap) of 0.25%.

3.2 Evaluation of the Chromosome

The objective of the genetic algorithm is to maximize the Net Present Value of the real option (NPV of the oil reserve, Equation 1). The Monte Carlo simulation is employed with 10000 iterations and at each iteration (i) the price of the oil is estimated for each 2-day interval (t) until expiration (2 years), based on the assumption that the oil price follows the stochastic process called Geometric Brownian Motion. The Latin Hypercube Sampling was used to provide the necessary random numbers for the stochastic process. Then, for each iteration, a "path" of the oil price, which has been named $Path_i$, is formed.

The following describes the steps to evaluate each chromosome:

1. The evaluation of chromosome j (for j = 1,2,…population size) begins with the first iteration of the Monte Carlo simulation (i = 1).
2. From the parameters for the logarithmic function and free point contained in the chromosome (j), the decision rule is constructed for the three alternatives, defining the waiting and exercise regions.
3. One $Path_i$ of the oil price is created and for each time interval(t) it is verified if the oil price reaches one of the exercise regions.
   - If the oil price reaches an exercise region, the option is exercised, the NPV (F_i) for this oil price is calculated, and then the algorithm goes on to the next iteration (stept 1).
   - If Pathi has been completed, i.e., it is at expiration, and none of the exercise regions has been reached, then the NPV is zero and the algorithm goes on to the next iteration (stept 1).
4. The process described in steps 1, 2 and 3 is repeated for each iteration of the Monte Carlo simulation (i).
5. Once the Monte Carlo simulation is finished, the evaluation value (fitness) for this chromosome (j) has been determined by the mean value of the NPVs (F_i), which was found in each iteration (5).

$$F_j = \frac{\sum_{i=1}^{10000} F_i}{Number \quad of \quad Iterations} \tag{5}$$

The best chromosome obtained by the genetic algorithm will be the one that maximizes the value F_j .

4 Results

The table 2 presents the comparative results of the net present value (NPV) achieved by the Genetic Algorithms and by the Partial Differential Equations (PDE), supposing that the oil price follows a Geometric Brownian Motion (GBM). As it can be observed, the results are very close, proving the good performance of the Genetic Algorithm.

Table 2. Table Comparative of the NPV by Genetic Algorithm and D.P.E

	NPV (MM US$)	
	Genetic Algorithm	P.D.E.
GBM	325.063	323.34

The Figure 3 presents the optimal decision rule obtained by the genetic algorithm and by the partial differential equations method (PDE), considered that oil price follows a Geometric Brownian Motion, for the three investment alternatives that have been considered for an option period of two years, with 10000 iterations of the Monte Carlo simulations for each chromosome.

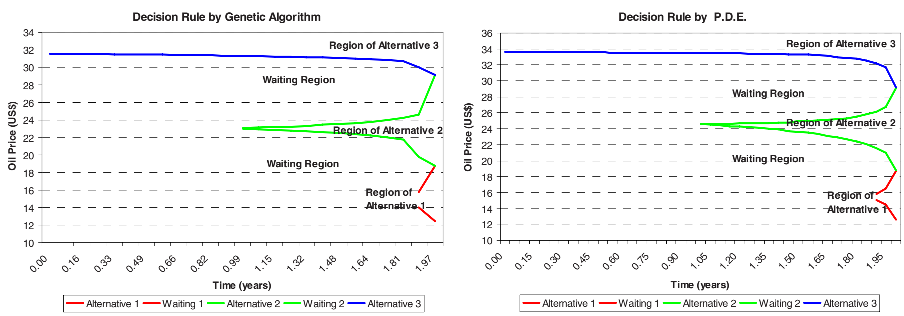


Fig. 3. Compare the Optimal Decision Rule obtained by GA and PDE: Oil Price follows a Geometric Brownian Motion

The results of table 2 and figure 3 show as the proposed model obtains a good approximation of the solution obtained by PDE. This approximation of the solution still can be improved using more free points with the logarithmic function or executing more generations of the genetic algorithm

5 Conclusions

In this paper was proposed a model using genetic algorithms and the Monte Carlo simulation, which was projected to find an optimal decision rule of investment and to determine the value of the option for a project of oil exploration that has some options of investment under market uncertainty. This decision rule is formed by three mutually exclusive alternatives that describe three regions of exercise through time up to the expiration of the option. The MCS is used with a genetic algorithm to simulate the possible paths for the price until the expiration. The Geometric Brownian Motion had been used to represent the oil price, using a technique of variance reduction (LHS), with the intention to improve the computational efficiency and the precision of the results of the Monte Carlo simulation.

The result obtained by the genetic algorithm proved to be similar to the result obtained in the analysis by partial differential equations.

The advantage of using the model with genetic algorithms in the analysis of development alternatives is that it is more flexible. This enables the introduction of a greater number of investment alternatives, to change the stochastic process or to introduce other uncertainties with minor modifications. Such aspects represent one of the most important limitations in the case of analytical methods, where the increasingly higher number of random variables and of alternatives makes it practically impossible to solve the partial differential equations.

In addition, changing the stochastic process involves changing all the partial differential equations.

Another advantage is that the GA makes it possible to obtain optimal or suboptimal decision rules and avoids the need to solve partial differential equations (PDE).

The model proposed and verified in this research presents itself as a promising alternative for systems of support to the decision in the area of real options in E&P.

References

1. Dixit, Avinash K. and Pindyck, Robert S.: Investment under Uncertainty. Princeton University Press, Princeton, N.J. (1994)
2. Dixit, A.K.: Choosing Among Alternative Discrete Investment Projects Under Uncertainty. In: Economic Letters, Vol.41. (1993) 265-288
3. Dias, Marco Antonio G.: Investment in Information for Oil Field Development Using Evolutionary Approach with Monte Carlo Simulation. 5th Annual International Conference on Real Options – Theory Meets Practice, UCLA, Los Angeles, USA, July 13-14, (2001)
4. Lazo L. Juan G., Pacheco Marco A. C., Vellasco, Marley M. R., Dias, Marco A. G.: real option Decision Rules for Oil Field Development Under Market Uncertainty Using Genetic Algorithms and Monte Carlo Simulation. In: Proceedings 7th Annual International Conference on Real Options - Theory Meets Practice, Washington DC,USA, July 10-12, (2003)
5. Goldberg, David E.: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, Inc. (1989)
6. Michalewicz, Zbigniew: Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag, USA (1996)
7. Brealey and Myers: Principles of Corporate Finance. McGraw-Hill, Inc., sixth ed.. (1999)

8. Copeland, Tom and Antikarov, Vladimir: Real Options: A Practitioner's Guide, Texere, first edition (2001)
9. Paddock, J.L., Siegel, D. R. and Smith, J. L.: Option Valuation of Claims on Real Assets: The Case of Offshore Petroleum Leases. In: Quarterly Journal of Economics, August (1988) 479-508
10. http://www.puc-rio.br/marco.ind/quality.html
11. Adelman, M.A., Koehn, M.F. and De Silva, H.: The Valuation of Oil Reserves. In: SPE Hydrocarbon Economics and Evaluation Symposium, SPE paper no 18906, Dallas, Texas, March (1989) 45-52
12. Adelman, M.A. and Watkins, G.C.: The Value of United States Oil and Gas Reserves. MIT Center for Energy and Environmental Policy Research, May (1996) 92.
13. McKay, M. D., Conover, W. J. and Beckman, R. J.: A Comparison of Three Methods for Selecting Values of Input Variables in Analysis of Output from a Computer Code. In: Technometrics, No 21 (1979) 239-245
14. Tang, B.: Ortogonal Array-based Lain Hypercube. In: Journal of the American Statistical Association, Vol 88 (1993) 1392-1397
15. Dias, Marco A.G., Rocha, Katia and Teixeira, José P.: The Optimal Investment Scale and Timing: A Real Option Approach to Oilfield Development. 8th Annual International Conference on Real Options – Theory Meets Practice, Montrela, Canada, June 17-19, (2004)
16. Dias, Marco Antonio G.: Real Option Evaluation: Optimization under Uncertainty with Genetic Algorithms and Monte Carlo Simulation. In: Working paper, Department of Electrical Engineering, PUC-Rio, Brazil, July (2000)

Documenting Visual Quality Controls on the Evaluation of Petroleum Reservoir-Rocks Through Ontology-Based Image Annotation

Felipe I. Victoreti[1], Mara Abel[1], Luiz F. De Ros[2], and Manuel M. Oliveira[1]

[1] Instituto de Informática, UFRGS; Av. Bento Gonçalves, 9500 – Bloco IV – Campus do Vale; Bairro Agronomia; Porto Alegre – RS – Brazil - CEP: 91501-970
{ftoreti, marabel, oliveira}@inf.ufrgs.br
[2] Instituto de Geociências, UFRGS; Av. Bento Gonçalves, 9500 – Campus do Vale; Bairro Agronomia; Porto Alegre – RS – Brazil - CEP: 91501-970
lfderos@inf.ufrgs.br

Abstract. Depositional and post-depositional (diagenetic) processes control the distribution of porosity and permeability within petroleum reservoir rocks. The understanding of these controls is essential for the construction of models for the systematic characterization and prediction of the quality (porosity, permeability) of petroleum reservoirs during their exploration and production. The description and documentation of key petrographic features is an important tool for the evaluation of reservoir quality that try to minimize the uncertainty associated to visual recognition of the features. This paper describes the role of visual controls on the petrographic analysis of reservoir rocks, and presents a knowledge-based tool that supports a workflow for the collection and documentation of visual information. This tool allows the spatial referencing of significant features in thin sections of reservoir rocks and the association of these features to a complete ontology of description. The whole process allows the preservation of original information that would support reservoir evaluation and guarantees further analysis even when the original rock sample is not available.

Keywords: Visual knowledge, image annotation, reservoir quality evaluation.

1 Introduction

The most important intrinsic properties of petroleum reservoir rocks are their porosity – the percentage of their total volume occupied by fluids, i.e., oil, natural gas or water – and their permeability – the amount of such fluids that can flow through a rock section in a time unit. The values and distribution of porosity and permeability within reservoir rocks are conditioned by depositional and post-depositional (diagenetic) aspects, such as the depositional structures, grain size and selection, the types, textures and location of diagenetic processes and constituents. These parameters are described during the systematic petrographic analysis of reservoir rocks, in order to provide the essential information for the creation of models for the characterization of the quality and heterogeneity of reservoirs under production, or for the prediction of quality of new reservoirs during exploration. Therefore, the acquisition and documentation of the key textural and

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 455–464, 2007.
springerlink.com

compositional petrographic features has an enormous importance for the evaluation of effective or potential reservoir rocks. However, capturing information from images is a natural uncertain process. Image recognition involves previous knowledge and hypothesis about what is being seen, the progressive adjustment of the viewer in order to fit the understanding to the seen features during image scanning and, also, judgments about the significance of this features related to geological interpretation.

This paper describes an approach to overcome the difficulty in applying visual knowledge in reservoir evaluation. We formalize a workflow for the systematic description and storage of the key visual petrographic features, as seen in petrographic microscopes, and provide computer support by a system composed by a piece of software and hardware.

The treatment of uncertainty provided by the system was conceived to deal with the incomplete collection of information and the partial confidence of the expert interpretation rules. The system tries to offer the best support to overcome the information loose in image recognition. It provides a petrographic ontology to orient the feature recognition and, when it is done, saves the spatial locations of the key features and associates them both to ontological terms as well as to hyperlinks for other objects supporting the description, such as pictures, audio and video files, websites, etc. As a result, it produces a complete documentation of the features, in the form of a *virtual map* of the thin section, that preserves the evidences for the reservoir evaluation. The reasoning method searches over this description for the required features to prove one or other interpretation hypotheses. The approach was implemented in the *PetroGrapher* system, an intelligent database application designed to support the detailed petrographic description and interpretation of oil reservoir rocks, and the management of relevant data using resources from both relational databases and knowledge-based expert systems. Systematic description through the system is facilitated by the use of flexible menus with standardized nomenclature and parameters, what radically reduces description time and errors. An integrated electromechanical microscope stage, the *StageLedge*, allows an optimized quantification, as well as the referencing of every point identified and/or photographed in the thin sections. The collection of qualitative features is driven by a domain ontology [1] that formalize the knowledge related to petrographic description and diagenetic interpretation. The main approaches of *PetroGrapher* system are described in [2-4]. In this paper, we present the basis for the automation of the process of quantitative petrographic analysis with the support of an electromechanical appliance [5] attached to the microscope and interfaced to the software system. This approach reduces the uncertainty associated to visual interpretation by improving the semantic capturing and allowing a better register for recovering the evidences that support inference. This information is used by the problem-solving method implemented in the system to propose the probable diagenetic environment where the rock where formed. The petrographic analysis is accomplished through two stages. First, the petrographer analyzes the fabric, texture and structure of the rock, and identifies the main rock-forming constituents, pores and diagenetic relations. After that, he/she performs a quantitative analysis, through the systematic scanning of the rock, following a virtual grid of evenly-spaced points over the thin section. During this scanning, the petrographer identifies and annotates the constituents and

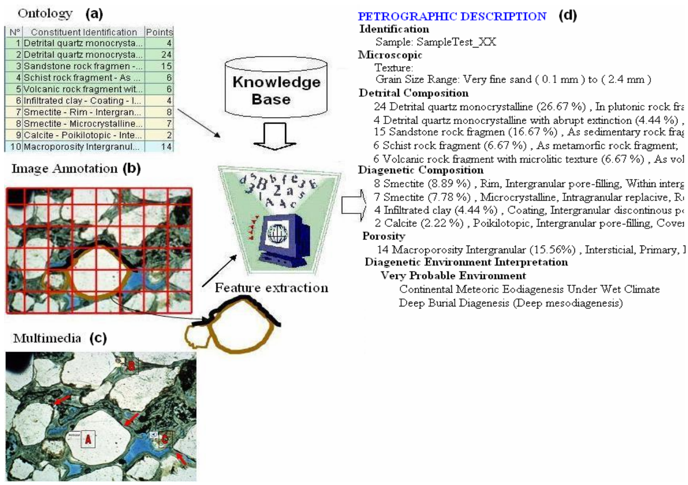


Fig. 1. Thin section of rock sample description process. The system uses: (a) qualitative and quantitative description using a domain ontology; (b) virtual map of thin section allowing annotation to identify objects on image (feature extraction); (c) hyperlinks and observations are used for documentation. The system provides a textual report with all information (d).

the diagnostic features that may suggest interpretations of reservoir quality. In addition, important features may be captured through photographs, which can be annotated later on. All process is presented in Fig.1.

Although the qualitative and quantitative description is a systematic process, it is strongly affected by different styles of descriptions and subjective factors. Petrographers with distinct levels of expertise can produce quite different descriptions, mainly because the spatial relationships among the constituents, or *paragenetic relations*, as well as uncommon minerals are not properly recorded. Even the same petrographer when recalling a previous description may find it difficult to check some of its specific aspects, because the original position and described points over the thin section cannot be found again with precision. Our approach helps in minimizing the information loose typical of the visual-based interpretation process by providing a standard description based on a complete formal vocabulary and petrographic support tables. It also supports the capture of all features along with the original spatial position in the rock, and allows the re-utilization of information, even in the absence of the original object.

2 Preliminary Concepts

The structuring and the representation of the knowledge for problem solving in domains that require image pattern matching and then high level of abstract inference is a challenge for knowledge engineering although is essential for many critical tasks, such as automatic surveillance in intensive care units [6], recognition of biological organisms [7], and biochemistry [8]. The uncertainty in the task are related mainly to the follow factors: (1) the collection of information is not complete because the observer does not posses the knowledge to recognize what is being seen providing incomplete or wrong descriptions; (2) even with the adequate capture of the diagnostic features, the level of significance of these features in indicating a particular interpretation may not provide a fully trustful solution.

Our system helps in making available the knowledge by supporting the feature recognition through a domain ontology and providing a problem-solving method to deal with significance and confidence factors of the interpretation task.

2.1 Ontology and Image Annotation

Ontologies that formalize the visual concepts of the domain are shown as the alternative to reduce the gap between the knowledge representation and image features, as presented in [9] and also in this paper. The most common definition of ontology asserts that an ontology is a formal, explicit specification of a shared conceptualization [10]. Formal refers to the fact that the ontology should be machine-readable, and shared reflects the notion that an ontology captures consensual knowledge, accepted by a group. Ontologies express the descriptive knowledge and also the way in which the evidences support the conclusions. Usually, they are combined with problem-solving methods [11] to provide reasoning features to expert systems, like we done in our system.

There are many approaches to ontology-based content annotation of images, based on a standardized vocabulary defined by a domain ontology that allows further processing [11] [6] [12]. Images must be annotated with keywords and content-based combined queries and refinements [13]. During annotation, a lot of information can be expressed, helping to improve the understanding and to reduce ambiguity.

The images can also be annotated with links over the image to any kind of multimedia content like it was done in [14]. It is possible to associate data under analysis to different resources, in order to provide a complete documentation.

2.2 Spatial Referencing

A spatial reference system provides an association of the object with its location related to a specific coordinate system. A coordinate system can reference a particular point in an n-dimensional space defined by an origin, directions of the axis and a distance scale. The quantitative analysis of a thin section in an optical microscope is usually done referring each identified feature to an imaginary net of points along a sequence of evenly-distanced steps, corresponding to successive positions marked by the cross lines of the microscope eyepiece. In some cases, this net is scanned with the help of some equipment, such as performed in [5] and in [15]. However, the complete association of the net of points to a spatial coordinate system with an *anchor* linked to a real reference of a physical object, as developed in this work, is a new proposition.

3 Petrographic Analysis Using Software Control and the *StageLedge* Device

In order to support the systematic quantitative analysis in the *PetroGrapher* system, we have developed an electromechanical device, called *StageLedge* [5], to be attached to the optical microscope and connected to a software control. The quantification of the constituent elements requires a uniform scanning of the thin section, so we have defined an automatic method. This approach not only guarantees that the thin section can be completely explored, but most important, that the entire process can be repeated, preserving the spatial coordinates of each point of the virtual map.

3.1 Virtual Thin Section

Our approach is based on the creation of a digital version of the thin section, on top of which virtual feature maps can be created. Although it does not completely eliminate the need for the real thin section, it has some significant advantages over the traditional approach that uses the original section both for description and for later verifications. This virtual thin section allows the recording of reference features through an existent ontology. For instance, our electronic version of the section can be sent over the Internet to an arbitrary number of petrographers, who can independently go over the reasoning process previously applied in the original documentation. Besides reducing the risks associated with a possible loss of the physical section, it also eliminates the costs associated to shipping the real thin section.

Moreover, the virtual thin section has the potential for unlimited documentation via hyperlinks to images, video, audio, and text provided by expert petrographers, as well as to other resources available on the Internet and related to the contents of the section. This sets the stage for a new level of rich documentation, turning the virtual thin sections into ideal training tools. This situation is illustrated in Fig. 2, which shows an image corresponding to a portion of a thin section containing links to different media formats: audio, video, other images, websites, and other observations. These resources intent to complement the information that is captured by the user, minimizing the lack of comprehension about what is seen in the description time. The images should have less than two megabytes to not slow down the system execution.

3.2 Complete Description Process

The analysis of a thin section starts with its positioning in the petrographic microscope, which has a rotating stage that allows the examination of the optical properties of minerals when examined with polarized light. The direction of the transect lines across which the points are quantified during analysis is defined transversally to the structures and fabric of the rock, and the modal size of grains indicates the adequate size of the step to be used with the StageLedge.

In the process of creating a digital map of the rock, we digitize a physical thin section using a regular flatbed scanner and use the resulting image as a base map on top of which the additional documental information will be placed. This stage requires the careful correction of the scale, tilt and coordinates of origin of the scanned image, in order to provide a correct association with the origin and scale provided by StageLedge. Once the image has been captured and associated to the actual position in the equipment, the documentation can be referenced to the real spatial coordinates. Since flatbed scanners can capture images at different resolutions, it is necessary to specify the selected resolution in points per inch (ppi), thus relating pixels to real distances in the thin section. According to our experience, the use of 600 ppi provides

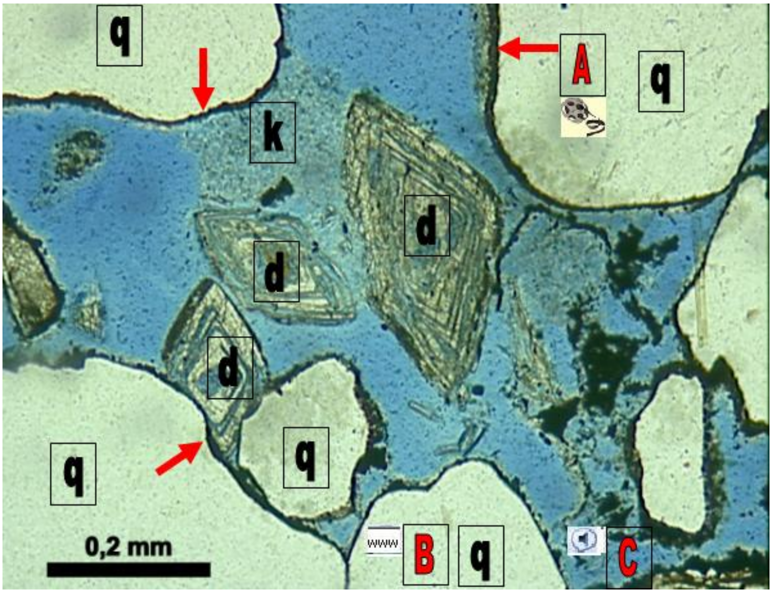


Fig. 2. A rock thin-section image with links and observations: A: Video file C:\video.mov; B: WEB link; C: Sound file C:\quartzExplain.mp3; Observations: Grains of quartz (q) covered by coatings of iron oxide (arrows); zoned and partially dissolved crystals of dolomite (d); microcrystalline kaolinite (k) partially filling the intergranular pores (impregnated by blue epoxy resin); uncrossed polarizers.

satisfactory results. In order to support the petrographic description, the *PetroGrapher* system controls the steps of the *StageLedge* and allows the user to select constituents and features described in the domain ontology and associates them to the current position under analysis in the thin section. Fig. 3 illustrates the interface of the *PetroGrapher* system, showing the constituents of a given rock sample and a partial menu providing specific ontology terms.

The quantitative petrographic analysis identifies and saves the location of every constituent positioned in each of the coordinates in the virtual net, controlled by *StageLedge*. The *PetroGrapher* interface depicts different minerals using colors, as presented in Fig. 4. Thus, with just a quick glance, the geologist can have a good idea of the spatial distribution of the constituents and pores identified in the thin section.

The possible constituents and pores that can be found in a thin section are fully described in a domain ontology, as well as the attributes and domain of values of them. The ontology also describes in which way the instances of qualified constituents can indicates the rock-formation environment. This is expressed by *knowledge graphs*, a one-level tree where the root node represents the interpretation hypothesis and the leaf nodes represent visual chunks identified by the experts in the image of rock as pieces of evidence necessary to support the interpretation. The uncertainty of interpretation is represented in the knowledge-graph by a threshold value that represents the minimum amount of evidence needed to indicate it. Also, the chunks have an influence

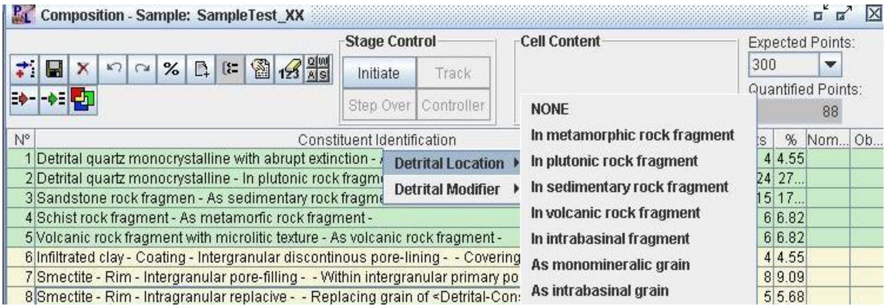


Fig. 3. Rock sample composition interface with the description ontology for constituents

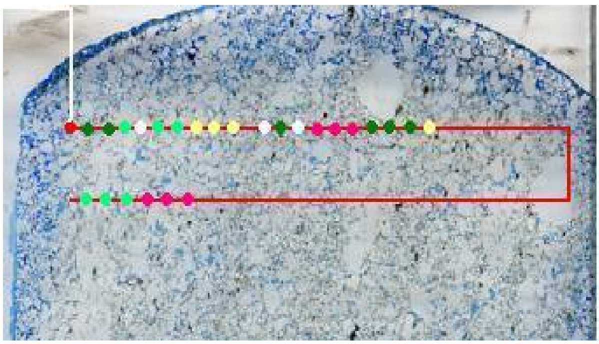


Fig. 4. Thin section with points of different constituents, identified by different colors

factor and are combined to increase the influence and the certainty of the interpretation stated. By their side, the chunks represents in an AND-OR tree the several ways in which way an evidence can be recognized in the rock, such as, possible minerals, possible habits, locations, etc. The knowledge model and problem-solving method of the system is extensively described in [3].

For each significant feature identified in the thin section, the user can capture a photograph and associate it to the coordinates of the described point or annotate the captured image itself, describing the special characteristics that must be considered. Thus, the quantitative analytical process generates a map containing the documentation of the most important diagnostic features for reservoir evaluation. According to the user interest, the system can selectively show the location of the special features, as exemplified in the left window of Fig. 5, where the segments indicate the trajectory of the analysis and the white dots shows the position of the selected constituent. As the user moves the mouse over a point for which a picture has been taken, it is automatically shown in right side window (Fig. 5). This documentation will provide further validation to the reasoning process or may show possible errors in feature identification.

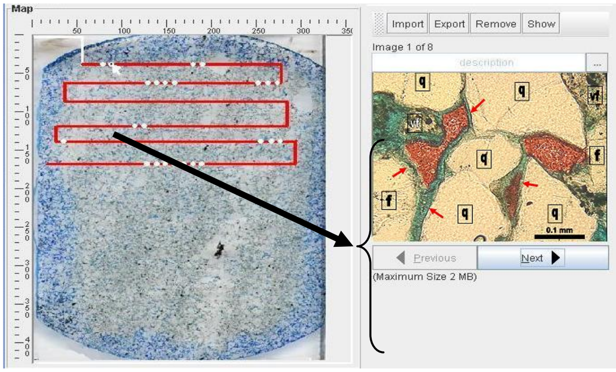


Fig. 5. Interface showing the virtual thin section with a superimposed map of the analytical pathways and points (left). As the mouse cursor passes over a feature point (white dot) for which a picture has been taken, the photograph is shown on the right. Note that the photograph itself contains a series of annotations in the form of hyperlinks.

The digital map of the rock uses the digital thin section image as a base map on top of which the additional documental information will be placed. This stage requires the careful correction of the scale, tilt and coordinates of origin of the scanned image, in order to provide a correct association with the origin and scale provided by StageLedge. Once the image has been captured and associated to the actual position in the equipment, the documentation can be referenced to the real spatial coordinates.

At the end of the process, an extensive documentation of the thin section is provided. For example, it is possible to locate where the 10th detrital quartz is located and then visualizing it. Moreover, the system guarantees that all descriptions would be performed based on a formal and complete petrographic vocabulary, defined in the domain ontology. This feature will provide extra capabilities by allowing the automatic geological interpretation and correlation with the captured information.

3.3 The Interpretation Process

The features described are stored in a database, along with the spatial coordinates of their position in the virtual map. The reasoning method loads each knowledge graph and match the chunk representation from the knowledge base against the user descriptions on the database. When the set of features that describes a chunk is found, the reasoning accumulates the related influence factor. When the sum reach the threshold of the graph, the conclusion is stored in the database in the description record and presented to the user. More than a conclusion can be associated to one rock, since more than one environment can act in rock consolidation.

4 Conclusions

The tool described in this paper has been tested by a group of six geologists from the Geosciences Institute at the Federal University of Rio Grande do Sul (Brazil), who works usually with petrographic analysis. The group has long time experience with the manual method and has migrated recently to *PetroGrapher* system. The goal was measuring the time expending in the task description and the amount of information that was collected in each method in normal work condition. Each participant received six different rock samples to be described. Each sample was described by four geologists, two using the *PetroGrapher* system, two using an electronic spreadsheet and a mechanical stage. According to these experiments, the use of the *PetroGrapher* system with the StageLedge has led to a reduction of 25% in the time required for a full petrographic description. Otherwise, the descriptions are longer and contain more information than those made by manual method.

The spontaneous comments from *PetroGrapher´*s users include: (1) the system guarantees a standard documentation without loosing semantic of feature description; (2) the possibility of recovering the original position of some specific feature provides a better framework for reservoir understanding; and (3) the information can be easily queried and reused. As a result, one can draw further correlations between petrographic data and well logs, seismic profiles and core descriptions. These capabilities are essential for a powerful tool used for advanced reservoir characterization.

Acknowledgements. This work is supported by CNPq Brazil (process 475597/2006-0 and 477848/2006-0) and by ENDEEPER Co. (www.endeeper.com) that retails PETROLEDGE, the commercial version of the described system.

References

1. Abel, M., *Estudo da perícia em petrografia sedimentar e sua importância para a engenharia de conhecimento*, in *Programa de PG em Computação*. 2001, UFRGS: Porto Alegre. p. 239 (in Portuguese).
2. Abel, M., Silva, L.A.L., De Ros, L.F., Mastella, L.S., Campbell, J.A., and Novello, T., *PetroGrapher: Managing petrographic data and knowledge using an intelligent database application.* Expert Systems with Applications, 2004. **26**(1 SPECISS): p. 9-18.
3. Abel, M., Silva, L.A.L., Campbell, J.A., and De Ros, L.F., *Knowledge acquisition and interpretation problem-solving methods for visual expertise: a study of petroleum-reservoir evaluation.* Journal of Petroleum Science and Engineering, 2005: p. 51-69.
4. Mastella, L.S., Abel, M., Lamb, L.C., and De Ros, L.F. *Cognitive Modelling of Event Ordering Reasoning in Imagistic Domains*. in *19th International Joint Conference in Artificial Intelligence - IJCAI´05*. 2005. Edinburgh, UK.
5. Victoreti, F.I., *Uma Platina Eletrônica de Microscópio Controlada por um Computador*, in *Instituto de Informática*. 2004, UFRGS: Porto Alegre. p. 40 (in Portuguese).
6. Hu, B., Dasmahapatra, S., Lewis, P., and Shadbolt, N. *Ontology-based Medical Image Annotation with Description Logics*. in *Proceedings: 15th IEEE International Conference on Tools with artificial Intelligence, Nov 3-5 2003*. 2003. Sacramento, CA, United States: Institute of Electrical and Electronics Engineers Inc.
7. Hudelot, C., *Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms*, in *École Doctorale STIC,*. 2005, Université de Nice Sophia Antipolis.
8. Sofou, A. and Maragos, P. *Soil Image Segmentation and Texture Analysis: A computer Vision Approach*. in *IEEE Geoscience and Remote Sensing Letters*. 2005.

9. Maillot, N., *Ontology Based Object Learning and Recognition*, in *École Doctorale STIC*. 2005, Université de Nice: Nice. p. 166.
10. Studer, R., Benjamins, V.R., and Fensel, D., *Knowledge Engineering: Principles and Methods*, in *Data & Knowledge Engeneering*. 1998. p. 161-197.
11. Gertz, M., Sattler, K.-U., Gorin, F., Hogarth, M., and Stone, J. *Annotating Scientific Images: A Concept-based Approach*. in *14th International Conference on Scientific and Statistical Database Management*. 2002. Edinburgh, UK.
12. Schreiber, A.T., Dubbeldam, B., Wielemaker, J., and Wielinga, B.J., *Ontology-based photo annotation.* IEEE Intelligent Systems and Their Applications, 2001. **16**: p. 66-74.
13. Sychay, G., Chang, E., and Goh, K. *Effective image annotation via active learning*. in *IEEE International Conference on Multimedia and Expo (ICME02)*. 2002. Lausanne, Switzerland.
14. Dong, A. and Li, H. *Multi-ontology Based Multimedia Annotation for Domain-specific Information Retrieval*. in *International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC´06)*. 2006. Taichung, Taiwan.
15. Hatiboglu, C.U. and Akin, S., *A new computerized moving stage for optical microscopes.* Computers & geosciences, 2003. **30**: p. 471-481.

Event Ordering Reasoning Ontology Applied to Petrology and Geological Modelling

Laura S. Mastella[1,3], Mara Abel[1], Luiz F. De Ros[2], Michel Perrin[3], and Jean-François Rainaud[4]

Universidade Federal do Rio Grande do Sul,
[1] Instituto de Informática,
[2] Instituto de Geociências, Porto Alegre, Brazil,
[3] École des Mines de Paris, Paris, France
[4] Institut Français du Pétrole, Rueil-Malmaison, France
{mastella, marabel, lfderos}@inf.ufrgs.br,
michel.perrin@ensmp.fr, j-francois.rainaud@ifp.fr

Abstract. The inference of temporal information from past event occurrences is relevant in several applications for geological domains. In such applications, the order in which events have happened is imprinted in the domain as visual-spatial relations among its elements. The interpretation of the relative ordering in which events have occurred is essential for understanding the geological evolution in different scales of observation and for various kinds of objects, as in Petrology and Geological Modelling. From the analysis of the cognitive abilities of experts in these domains we propose an ontology for event ordering reasoning within domains whose elements have been modified by past events. We show that the *Event Ontology* can work as a pattern for domain conceptualization to be applied in distinct domains. It can be used to specify the sequence order of diagenetic paragenesis. It can also be operative for automatic reconstruction of geological surface assemblages.

Keywords: Knowledge Engineering, Ontology, Sedimentary Petrology, Geological Modelling.

1 Introduction

The inference of temporal information from past event occurrences [1] is particularly relevant in domains such as law, medicine, archaeology, geology and many others. A geologist, for instance, identifies visual-spatial relations among objects (rock constituents, geological surfaces) as does a physician when analyzing medical images to identify pathologies. In both cases, the visual-spatial relationships that are observed are the result of a sequence of past events. Late minerals grow over pre-existing ones like tumours grow over healthy tissues.

In this work, we deal with two kinds of geological interpretations that are both involved in reasoning on temporal events. First, we examine how one can reconstruct the succession of diagenetic events, which affected siliciclastic rocks and consequently modified their porosities and permeabilities. Secondly, we identify the events related to the deposition and to the further evolution of sedimentary formations in order to identify the position of the geological surfaces (horizons, faults), which

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 465–475, 2007.
springerlink.com

limit hydrocarbon reservoirs. We are concerned by representing relative time, i.e. by the mere order in which the events happened. In addition, we aim at *deriving relative temporal information from another dimension* (the visual-spatial relations between the elements of the domain). In both cases, images are the starting point of the analysis. Petrologists observe thin section under an optical microscope while geophysicists and petroleum geologists identify geological surfaces on seismic images.

In order to propose representation primitives and an inference mechanism, a long process of knowledge acquisition techniques in the p*etrology* domain was carried out. The analysis of the cognitive abilities of the experts led to the development of a cognitive model picturing the geologist's reasoning concerning an imagistic domain (rock thin sections) [2]. The *Event Ontology*, which is part of this cognitive model, was shown to be capable of modelling the expert's reasoning when deriving the sequence of events which led to the visual-spatial organisation of the domain under analysis. Here we additionally present an application of Event Ontology for *Geological Modelling*. Moreover, we compare the petrology and geological modelling domains and map the ontology already proposed to both domains, in order to demonstrate that this model can be considered as a template of domain conceptualization to be applied in evolving domains.

Section 2 presents some Knowledge Engineering (KE) approaches for modelling temporal and spatial information and the basics of ontologies. Section 3 describes the geological domains on which this work is applied. Section 4 presents the cognitive model for event reasoning. Section 5 describes the application of the developed model to the domains in study, and finally Section 7 presents some preliminary conclusions.

2 KE Theoretical Foundations

In this section we introduce the main approaches of Knowledge Engineering for temporal and spatial representation and the basics of ontologies.

Relative and absolute notions of time. In the *absolute* notion, time consists of a sequence of discrete points (dates, hours, etc.). In the *relativistic* view of time, on the contrary, events and temporal relationships between them precede the notion of time. When is possible to define absolute time stamps associated to events, developing inference about ordering becomes a relatively simple task. However, according to [3], in most real domains, timing information is also conveyed by time relationships, such as "before" and "after" (referred to above as *relative time*).

Ontologies. According to most definitions, *an ontology is a formal, explicit specification of a shared conceptualization* [4]. Using ontological constructs, it is possible to describe *static knowledge*, specifying which are the objects that compose the domain and according to which structure they are organized. Ontologies are also used as means of semantic integration. According to [5] very general ontology formalizing notions such as processes and events, time and space, physical objects, and so on, can be developed with the explicit goal of providing a ground vocabulary to domain-specific ontologies. Recently, some authors have aimed at augmenting the expressive power of ontologies by including temporal information [6]. Most proposals considers absolute time stamp associated with objects of the ontology. However, in several application domains, events are not to be interpreted by putting time stamps over them.

3 The Domains of Study in Geology

Let us describe the geological domains concerned by this work.

Sedimentary Petrology: in the case of hydrocarbon exploration, this science aims at evaluating the economic prospects of oil fields and reservoirs by interpreting observations related to rock thin sections.

Several kinds of visual-spatial relations between rock constituents can be observed such as "*A covering B*", "*A engulfing B*". They are called paragenetic relations. These relations reflect the changes undergone by the rock in the course of the geological history, which are a result of a sequence of *diagenetic events*. Diagenetic events are physical-chemical processes, which acted over the sediments transforming them into solid rocks and, consequently, modifying the porosities and permeabilities of potential oil-reservoirs. Using his extensive previous knowledge, a qualified petrologist is able to point out the ordering of events by observing how the constituents are spatially and visually related to each other. Using a simple example: *If one mineral appears to be on top of other mineral, it means that the former was generated in the rock later than the latter.* The sequence of events is an important criterion to determine the quality of a reservoir. In Fig. 1, we show an example of rock sample and the visual-spatial relations between the minerals that were identified.

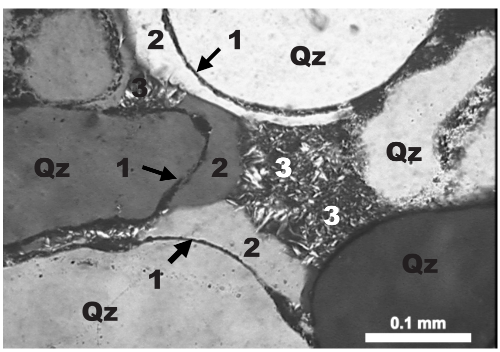


Fig. 1. Vision of a rock sample: (1) Hematite is covering grains of Quartz (Qz); (2) Quartz growings are covering hematite; (3) Quartz is being covered by Illite

Some interpretations techniques used for the evaluation of oil reservoirs were already modelled in the *PetroGrapher* system, an intelligent database application to support the description and interpretation of sedimentary rock samples [7]. The vocabulary of Petrology was elicited as a result of previous works on the domain [8] and modelled as a domain ontology.

Geological Modelling: 3D geological models are conventional representations of a definite portion of underground corresponding to hydrocarbon reservoirs or to sedimentary basin models. The blocks of geological matter are limited by surfaces such as sedimentary interfaces or faults, and geological modelling aims at reconstructing geological surface assemblages in order to obtain models that can further be populated by petrophysical properties.

Each defined surface of the model is the record of one defined geological event, which can be considered as having been instantaneous with respect to the geological time scale. Geological interpretation then consists in giving a geological qualification (stratigraphic surface, fault) to the surfaces entering into the model, and in implicitly or explicitly establishing a total or partial chronological order between the geological events to which they correspond. Since an older geological event cannot modify a younger one, the chronological order defined by the geological interpretation has consequences on the geometry and on the topology of the model to be built. Considering this, a ***geological syntax*** was defined as a result of previous work on the domain [9] and modelled as a geo-ontology [10]. The process of geological interpretation can be understood considering Fig. 2.

Fig. 2 is a synthetic example of most of the features currently present in geological assemblages. Several surfaces interrupting each others can be observed: Surface **T** is an example of an erosional surface interrupting surfaces **E** and **X**. Thus, ***T is younger than E and X***. Surface **a** is an example of an on-lap surface, which interrupts surfaces **b** and **c**. Thus, ***a is older than b and x***. Full spatial and temporal relations are thus established between surfaces a, b and c as a consequence of geological interpretation.

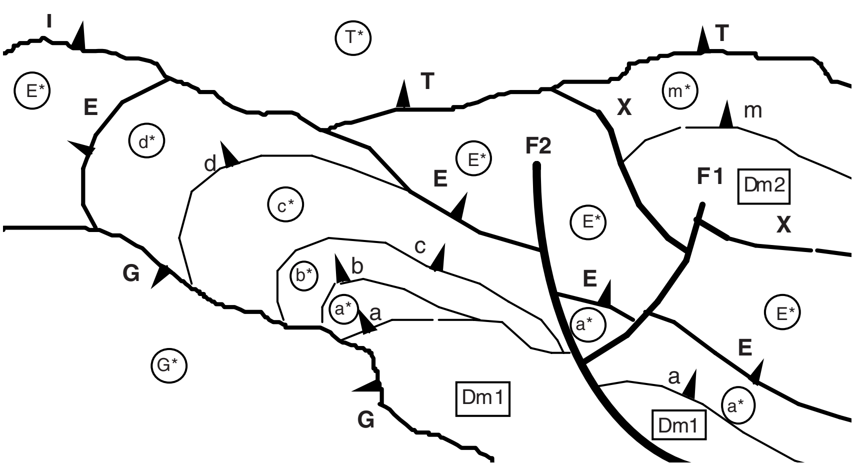


Fig. 2. Example of a geological scene

Petrography and geological modelling both consider processes and events that change spatial configurations. However, theses two domains contrast according to the scale of observation. While a petrologist observes thin sections of rocks at the microscope, structural geologists study geological assemblages whose horizontal dimensions may reach tens or even hundreds of kilometres. Even so, at both scales, the work of a geologist is comparable to that of a detective: it consists in observing spatial signatures and in trying to deduce from them the full chain of geological events that successively affected the domain. In section 5, we will make a closer comparison between the two domains, in order to identify the elements that play similar roles in petrology and in geological modelling.

4 A General Ontology of Events

The cognitive model presented here intends to model the evolution of a domain, which was submitted to various modifications resulting from events successively occurring in a non-planned order. Each of these events acted in the past as an operator transforming the domain. Their succession has induced several spatial relations among the domain elements. Considering the visual-spatial relationships that finally resulted from the full sequence of events and that can presently be observed, one can try to guess what were the events that have affected the domain and in which order they happened. This is the goal of the ontology-supported knowledge engineering approach that we propose.

4.1 The Event Ontology

We propose an extension to the classic constructs of ontological representation for evolving domains in order to capture the meaning of *events* and *temporal relations between them*. Such proposed constructs should be applied for modelling domains whose current state can be fully understood by considering the sequences of events to which they were submitted. We define the new constructs as follows (Fig. 3):

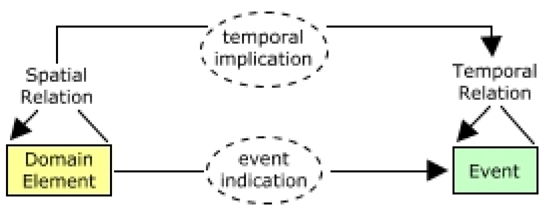


Fig. 3. The constructs of the Event Ontology

Event is a construct that acts as domain-transforming operator. It represents the phenomena that generate or modify the elements of the domain. Events are characterized by specific domain-dependent attributes, but not necessarily by a time stamp. They are also described by rules that associate them to their products. Events are, by their way, associated to each other by temporal relations.

Temporal relation is a construct proposed to represent the ordering relation between events. In order to reflect such ordering, we have defined the binary relations *before*, *after* and *during*.

Furthermore, we defined *inference rules* in order to represent the rules that the expert uses to produce the interpretation. We have two types of inference rules: *event indication rules* and *temporal implication rules*. In the *event indication rules,* the characteristics of the elements (expressed by class attributes in the ontology) are used to indicate which event originated or modified the element, as in:

```
if classA.attribute1 = value-x
then classB.attribute2 = value-y
```

The *temporal implication rules* are defined in order to allow the inference of binary *temporal relations* between events from the *visual-spatial relations* between the elements, as in:

```
if visual_relation(A,B)
then temporal_relation(A,B)
```

The main concepts that should be represented in the model are the ***domain elements***, which are the *items* of the domain that have possibly been generated or modified by the events. The relationships represented in the model are the ***visual relations*** between the domain elements (for instance, one element is *on top of* the other). Representing the visual relations is essential for the inference, because they show strong evidences of the order in which the events have occurred.

In the following section we explain how the Event Ontology was used as a base ground in order to map each of the two geological domains considered.

5 Mapping of the Cognitive Model to Geological Domains

We intend to show in this section how we identified the elements that have similar roles in the domains of Petrography and Geological Modelling and how those elements were mapped to the Event Ontology.

Sedimentary Petrology:

- *Rock Constituents* correspond to the minerals and pores that build a rock. Constituents can be minerals such as *quartz* or *illite*, and their more important properties are *habit*, *location* and *modifiers*. **They are *Domain Elements***.
- *Paragenetic Relations* describe the visual-spatial arrangements among constituents. Common paragenetic relations specify that a given mineral *covers* another mineral or *engulfs* another mineral, etc. **They are *Spatial Relations*.**
- *Diagenetic Events*. These events correspond to physical-chemical processes, which induced changes in rock mineralogy. The experts do not take into account the absolute period of time during which the various diagenetic events happened, but only the order in which they happened. Diagenetic events can be *dissolution*, *replacement*, *compaction*, *fracturing*, *deformation*, etc. **They are *Events***.
- *Ordering Relation*. Diagenetic events can have happened in a simultaneous or in a sequential way. In order to simplify the computational treatment of the sequence, we treat the ordering of events in pairs, as an expert does. The relations between pairs of events are *after*, *before*, and *during*. **They are the *Temporal Relations***.
  The ontology of Petrology resulted is as shown in Fig. 4.
- *Inference rules*. The expert is able to indicate the generating events by analyzing the characteristics of the rock constituents. For instance, when the attribute *modifier* of a constituent holds the value *deformed*, supposing that no small scale

tectonic deformation occurred, it is possible to conclude that the event that transformed the constituent is *compaction*. Hence, it was necessary to represent this knowledge as *event indication rules*. These inference rules define an association between *constituents* and *diagenetic events*, e.g. rule below:

```
if constituent.modifier = deformed
then event.event_name = compaction
```

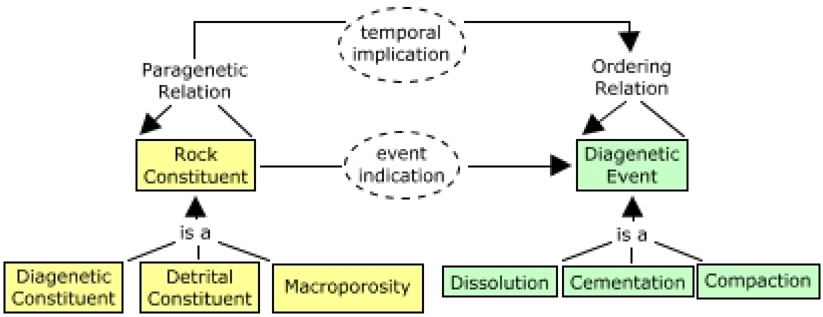


Fig. 4. A partial Petrology Ontology

After having identified the events, the expert is able to infer the order in which they occurred considering the visual-spatial (paragenetic) relations that he observes between the constituents. For instance, when a mineral appears to be *covering* (to be lying on the top of) another mineral, the expert says that the event that formed the first mineral occurred later than the event that formed the latter. The first part of this particular expert's rule is assuming a *paragenetic relation* between constituents.

The second part is defining an *ordering relation* between events. Thus, we need to represent this knowledge as *temporal implication rules*. An example of this type of rule is the following:

```
if covering(constituent1, constituent2)
  and produced_by(constituent1, event1)
  and produced_by(constituent2, event2)
then after(event1, event2)
```

Geological Modelling. The Geo-Ontology proposed by [10] deals with the broad arrangements of geological objects that are considered when building models.

- *Geo_Objects* correspond to actual physical geological objects, which are **Geological surfaces** and **Geological formations**. Geological surfaces correspond to limits of sedimentary formations (ex: horizons) or to tectonic discontinuities (ex: faults). A geological formation is a volume made of contiguous material points; it is fully limited by a set of geological surfaces. **They are the *Domain Elements***.
- *Topo_Assertions* are spatial relationships between intersecting surfaces, which can be: *interrupts* and *stops on*. **They are the *Spatial Relations.***
- *Geo_Event*, refers to a geological process occurring during a definite span of time or to a combination of such processes which correspond to *matter creation*

(sediment deposition, magma intrusion), *matter destruction* (erosion), *matter transformation* (diagenesis, metamorphism), *matter deformation* (folds, faults, thrusts). **They are the *Events*.**

- *Chrono_Assertions* represent the chronological relations that can occur between Geo_Events. They can be *younger than*, *older than*, or *contemporary to*. **They are the *Temporal Relations*.**

Current geological modelling rest on two main hypotheses [9]:

1. **The age hypothesis**: Since the events are responsible for creating or transforming surfaces, each geological surface corresponds to one defined event and has one defined age. Thus, there is a direct association between Geo_Objects and Geo_Events.
2. **The intersection topology hypothesis**: When two surfaces meet, one necessarily interrupts the other (no X-crossings).

Chrono-topological relationships between horizons can be described by providing them with attributes such as erosional meaning that they interrupt all older surfaces or on-lap meaning that younger horizons may stop on them.

The above rules are the main elements of the geological syntax that any geologist implicitly uses when interpreting crude geological data. This same syntax should also be used when building 3D models. Previous work operated in École des Mines de Paris has shown also that, in order to be geologically consistent, underground models should be built in accordance with a few chrono-spatial rules. Those rules can be expressed as, for example:

```
if stopsOn(surfaceA,surfaceB)
and (Erosional(surfaceB) or Fault(surfaceB))
then youngerThan(surfaceB,surfaceA)
if stopsOn(surfaceA,surfaceB)
and (OnLap(surfaceB))
then olderThan(surfaceB,surfaceA)
```

It means that when a geologist interprets the topological relation between the objects we can infer the temporal relations. The ontology of Geological Modelling resulted is as follows (Fig. 5):

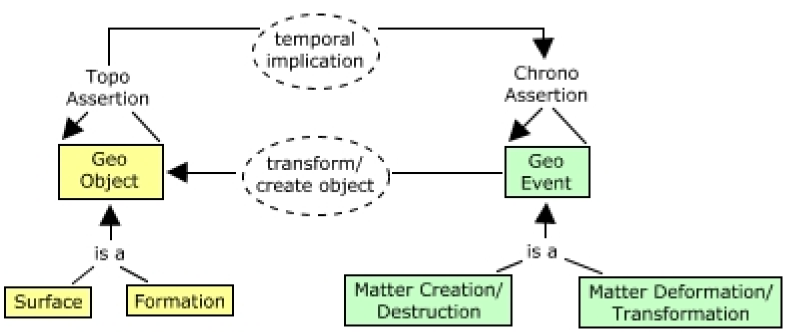


Fig. 5. Ontology for Geological Modelling

So, from the GeoOntology and from the Petrology Ontology we could identify the following equivalences, using the Event ontology as a base for the mapping (Fig. 6):

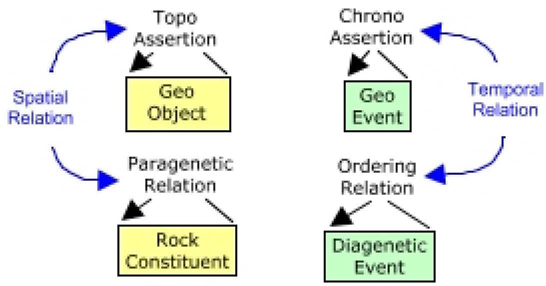


Fig. 6. Mapping of the Petrology and GeoModelling Ontologies to the Event ontology

It thus appears that the two domains have a similar event based organization in the knowledge level. Similar reasoning methods can be applied in the two domains to interpret the succession of events to which geological assemblages were submitted both at the petrology and at the geological modelling scales.

6 Validation of the Event Ontology

The proposed Event Ontology has been applied to the petrography domain, being implemented as an inference module within *the PetroGrapher* system [7]: the *diagenetic sequence interpretation module*. Real rock samples were described by the geologist in the *PetroGrapher* system and he also provided a previous interpretation of the sequence of diagenetic events, which was compared to the interpretation produced by the algorithm. The detailed experiment is described in [2]. The resulting event sequence is the same as the one inferred by an expert in most cases. For some rock samples the algorithm produces sequences of events that are not totally connected. However, in some cases, not even the expert is able to produce a complete sequence of events, because some paragenetic relations may be visible (and then described) in one sample and not in another one. Although the resulting sequence may sometimes be incomplete, it is certainly relevant to the domain, because, any sequence of events that can be inferred from a rock description is essential in understanding how the porosity and the permeability of the rock were affected, and how this influences the quality of the oil reservoir. This module is incorporated in the industrial version of *PetroGrapher* system[1].

7 Conclusion

We presented an *Event Ontology*, which allows correlating spatial and temporal relations and shows that it can work as a pattern of domain conceptualization to be

---

[1] The commercial name of the PetroGrapher system is PETROLEDGE®, which is being distributed by Endeeper (http://www.endeeper.com/).

applied in different geological disciplines. The models presented are able to describe the reasoning of an expert who observes and interprets visual-spatial relations in search for the best explanation about the sequence of events that caused them.

From the representation of the topological-temporal relation between two geological objects (Geological Evolution Schema – GES, [9]) it is possible to automatically rebuild from unsegmented geological surfaces a 3D geological model fully consistent both topologically and geologically.

Present day rocks and present day rock assemblages are the result of a complex history consisting in a succession of events related to various physical, chemical or mechanical processes. The art of the geologist consists in inferring from geological observation at different scales a geological interpretation which is nothing else that a possible or probable reconstruction of the geological history.

Considering two different scales and different types of geological objects, we have tried to show by using knowledge engineering techniques, that geological interpretation obeys to definite reasoning rules, which are similar from one geological domain to another, at least in some aspects. Although it is preliminary, this result appears to us as important since it may contribute to making fully explicit the geological interpretation procedures used during oil & gas exploration and to thus facilitating the collaboration between the various experts involved.

Acknowledgments. L. Mastella acknowledges the CAPES Foundation for financial support on her Doctoral work. M. Abel and L.F. De Ros are supported by the Brazilian Research Council - CNPq. We thank P. Verney, Doctoral student at ENSMP and IFP for discussions about the formalization of Structural Geology.

References

1. Thagard, P.S., Cameron P., *Abductive reasoning: logic, visual thinking, and coherence*, in *Logic and scientific methods*. 1997, Kluwer. p. 413-427.
2. Mastella, L., et al. *Cognitive Modelling of Event Ordering Reasoning in Imagistic Domains*. in *Proceedings of the 19th International Joint Conference on Artificial Intelligence - IJCAI'05*. 2005. Edinburgh, UK.
3. Bolour, A.D., L. J., *Abstractions in Temporal Information.* Information Systems, 1983. **8**(1): p. 41-49.
4. Studer, R.B., V.R. & Fensel, D., *Knowledge engineering: principles and methods.* Data & Knowledge Engineering, 1998. **25**(1/2): p. 161-197.
5. Noy, N.F., *Semantic integration: a survey of ontology-based approaches*, in *ACM SIGMOD Record - COLUMN: Special section on semantic integration*. 2004, ACM Press: New York, NY, USA. p. 65 - 70.
6. Bennett, B.G., Antony P., *A unifying semantics for time and events.* Artificial Intelligence, 2004. **153**: p. 13-48.
7. Abel, M.S., Luis Alvaro Lima; Ros, Luiz Fernando de; Mastella, Laura S.; Campbell, John A. & Novello, Taisa, *PetroGrapher: managing petrographic data and Knowledge using an intelligent database application.* Expert Systems with Applications, 2004. **26**(1): p. 9-18.

8. Abel, M.S., Luis Alvaro Lima; Ros, Luiz Fernando De; Campbell, John A. & Mastella, Laura Silveira. *PetroGrapher: managing petrographer data and knowledge using an intelligent database application*. in *WORKSHOP ON INTELLIGENT COMPUTING IN PETROLEUM INDUSTRY*. 2002. Mexico City, Mexico: [s.n.].
9. Perrin, M., *Geological consistency : an opportunity for safe surface assembly and quick model exploration*, in *3D Modeling of Natural Objects, A Challenge for the 2000's*. 1998: Nancy, France. p. 4-5.
10. Perrin, M., et al., *Knowledge-driven applications for geological modeling*. Journal of petroleum science & engineering, 2005. **47**(1-2): p. 89-104.

Optimization to Manage Supply Chain Disruptions Using the NSGA-II

Víctor Serrano, Matías Alvarado, and Carlos A. Coello Coello

Department of Computing, Centre of Research and Advanced Studies, CINVESTAV,
Instituto Politécnico Nacional 2508, San Pedro Zacatenco, CP 07763, México DF
vserrano@computacion.cs.cinvestav.mx,
{matias, ccoello}@cs.cinvestav.mx

Abstract. Disruption on a supply chain provokes lost that can be minimized through an alternative solution. This solution involves a strategy to manage the impact of the disruption and thus to recuperate the supply chain. Difficulty of this management is the diversity of factors such that becomes complex to provide or choice a solution among the possible ones. Depending on the objective(s) to optimize are the strategy to follow and the solution to choice. In this work the Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization NSGA-II is used as the strategy to generate and optimize (minimize) solutions (lost) in front of a disruption. The included objectives are cost, risk and the place of facilities supporting the supply chain recuperation. These objectives are combined to generate possible solutions and to choice one such that it provides a proposal to minimize the disruption impact on a delimited period of time. Advantage of NSGA-II utilization is the provision of a practical formal and computational tool to analyze different scenarios without simplifies the complexity of a standard real supply chain. The illustrative exercise presents recovery scenarios for a crude oil refinery supply chain.

Keywords: disruption management, supply chain, optimization, NSGA-II.

1 Supply Chains and Managing Disruption

A supply Chain (SC), in nature and in human societies incorporates a set of interdependent supply entities, such that the produced for one is consumed by others ones, such that they supply the consumption of thirds, and so on. This supply – consuming process, interactive and some time cyclic, applies to any individual or social activity. An alive organism is example of a supply chain of meals, oxygen and vital stimuli among organs and tissues. Departments of an organization, industrial, financial, educative, etc., chain the supply each provides, such that from this chaining it results the organization offered product: devices, financial plans, education, etc. In turns, this product (good or service) is required by suppliers in order to generate required products to other people and organizations. This supply – consume process occurs to different individual or organizational levels, with different scopes too.

In spite of the general description of before examples, it evidences the complexity of the interaction process in a SC. There are diverse suppliers and consumers, of different

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 476–485, 2007.
springerlink.com

kinds and sizes, interacting during distinct moments and through diverse circumstances, some times unexpected [17].

SC approach has shown its efficacy to modeling the supply – consume process in nature and the society to bounded scale. Several natural processes to molecular, individual or population level as well as diverse industrial, services, governmental or commercial processes can be suite modeled, to small or medium scale using this approach. Currently, challenge is to achieve such SC process modeling to large scale.

In order to model the SC interactions in a systematic way it is required do it like a process, formalized such a way to set enough flexibility to properly describe complex situation, diverse ones, and such that precise analysis and diagnosis can be made. Complementary, computational tools such that they allow trust solutions and with efficacy and simplicity. All fashioned such that the parameters variations to describe varied modalities are agile and direct.

In the interdependent supply – consume process whenever a chain entity (slave) fails to supply its product, it affects to the consumers. Eventually these affected entities leave to supply their products affecting in turns their consumers. And so on. An entity leaves to provide its products when suffer the lack of needed elements to produce it. This can be due to the fail on the supply of some element(s). Usually the fail inducing disruption is provoked by an unexpected event that affects the process.

A SC disruption has consequences of different magnitude, depending on the strategic position inside the chain of the lack supply product: fail on the gasoline supply in the Sidney airport in 2002 provoked a domino cascade of supply disruptions such that drops lost of hundred of thousand of million dollars. Fail on the blood and oxygen supply from the heart of an alive being to the rest of the organism due to an infarct, if large, get the being collapse and dead.

1.2 Disruption Management

In front of a disruption the look for emergent suppliers and routs is needed to supply the product which supply was interrupted. The look for alternative routs of provisioning sets a process combining diverse circumstances, determined by the location of replacing products. Such emergent processes to deal with disruption have a cost weighted by the urgency to incorporate to recuperate the SC operation.

In nature as well as in human societies to successfully surmount the disruption negative effects, it depends to some extent, on the capacity from the affecters to generate alternative plans and to strategically apply them. An adequate disruption management it includes to take in account, the SC disruptive scenarios, even the low probable or implausible. It means be aware about the risk situations that the entity is exposed as part of the SC. As well, it includes to be anticipated to disruption with viable solutions, of practical implantation; in the case of human SC with a costs assessment too. Complementary, the strategies to take in the interrupted resources: information on emergent providers including their location, answer capacities considering spend time and costs.

In essence, disruption management implies to minimize their negative effects by setting bounds through space, time and any kind of costs. Minimization in a dual vision implies the optimization of the plausible solutions; at least of some of them. This minimization/optimization of the disruption negative-effects/solutions, i. e, the disruption managing, let the survival of the entity being affected, either alive being or social chain; furthermore, the permanence and competence of this SC.

The analysis on optimization/minimization of the solutions/negative-effects to manage SC disruption, natural or social ones, makes evident that this is not a linear process. It combines several factors each with the relevance weight depending on the circumstances. Solutions are not exclusive but can be weighted in different manners. Furthermore, optimization/minimization is multi-objective. Thus, there are several alternatives to optimize solutions, depending on the weighting to each objective throughout the solution.

The disruption administration process in a SC involves parameters and objectives to optimize. In this circumstance the optimization must be a heuristic. The optimization process to minimize negative effects with the disruption is divided in three steps. The fist step covers identification of possible solutions, considering parameters that characterize a disruption in a SC as well as different objectives in consideration. The second step is the process that ranks the objectives, from where the solutions can be generated. The third is the choice of solutions.

Supply chains in petroleum industry is characterized by the product availability in the right place at the right time is an important matter. Supply availability in any kind of industry is necessary for its maximum development, and it is needed the knowledge of the right supplier in some fixed circumstances in order to minimize the lack of product supply. Making a finished product is a hard task, even if the distribution is decentralized because of the difficult of organizing the product delivery. Another problematic is fluctuation in the production demand, which is produced by many factors and circumstances.

The agent based systems can emulate activities performed in industry departments, as well the suppliers, logistic services, etc. Agents can model business policies and simulate different processes in industry. The supply chain dynamic is emulated by simulations of discrete events on agent based modeling.

An adequate SC disruption management is part of the whole planning in industry processes. It includes a transport network of supply raw materials and finished products, as well as risks identification and plans to control these risks once they happen. It is mandatory to have a short term demand projection and the required raw material quantity to cover this demand. It is necessary to manage and select raw materials in order to improve cost and production, and find the best way to storage raw materials.

This work shows the way that a supply chain management in petrochemical industry can be modeled to determine a viable solution to control unexpected events. Disruption management can help in the disruption detection even before they occur if their causes are well controlled, in the worst case, there would be time enough to take adequate measures to correct them. The problem causes can be detected in order to prevent and minimize looses in near future.

2 Optimization Using Multi-objective Evolutionary Algorithms

In most of the real world problems, requires information that of solutions we can obtain is by using heuristic method. A heuristic is a method that search for almost optimum solutions with a reasonable computational cost. Although this methods does not warranties the best solution, because in most of the real world scenarios we are not able of knowing how near of the optimum ours solutions are.

2.1 Basic Genetic Algorithm

Genetic algorithms were developed by John H. Holland in the early 60s [5, 6], motivated by his interest in solving problems in machine learning. Within evolutionary computation, genetic algorithms emphasize the importance of the so called sexual operator over the mutation operator, and uses a probabilistic selection. Standard construction of a Genetic Algorithm involves:

- Randomly generating an initial population.
- Computing fitness value for each individual.
- Perform selection (in a probabilistic way) based on the fitness values.
- Apply genetic operators (crossover and mutation) in order to generate the next population.
- Loop until the maximum number of iterations is reached.

Genetic algorithms do not need specific information in order to guide the search because they are a heuristic technique. A GA can be seen as a black box that can be connected to any particular application. In order to apply a genetic algorithm, the following basic components are required:

- A representation of potential solutions to the problem.
- A method to create an initial population of possible solutions (this is normally done in a random manner).
- An evaluation function that plays the role of environment, classifying the solutions depending on the fitness value.
- Genetic operators that modify the composition of the offspring produced for the following generation.
- Specify values for the genetic algorithm parameters (population size, crossover probability, mutation probability, maximum generation number, etc.)

The traditional representation used to encode a set of solutions is the binary scheme, where a chromosome is composed of a chain with the form $b_{1,} b_{2,} \ldots, b_m$ where each element is called alelo (that could be zeros or ones).

Genetic algorithms have been widely used in multi-objective optimization because of its population-based nature, which allows the generation of several elements of the Pareto optimal set with a single run [4].

The NSGA-II is an approach with $O(kN^2)$ computational complexity (where k is the number of objectives and N is the population size). The NSGA-II uses a selection operator that creates a mating pool by combining the parent and offspring populations and selecting the best (with respect to fitness and spread) N solutions from them. Because of its low computational requirements, its elitist approach, and its parameterless sharing scheme, the NSGA-II is the algorithm chosen to act as our optimizer in the research reported in this paper.

2.2 NSGA-II

The Fast Elitist Nondominated Sorting Genetic Algorithm for Multi-Objective Optimization, NSGA-II [12] successfully combines the following key elements:

1. A fast nondominated sorting approach.
2. A density estimator.
3. A crowded comparison operator.

In the fast nondominated sorting approach, each solution is compared with every other solution in the population to find if it is dominated. First, all individuals in the first nondominated front are found. In order to find the individuals in the next front, the solutions in the first front are temporarily discounted. The procedure is repeated to find all the subsequent fronts.

To get an estimate of the density of solutions surrounding a particular point in the population the average distance of the two points on either side of this point along each of the objectives is adopted. The obtained quantity serves as an estimate of the size of the largest cuboid enclosing the point of interest, without including any other point in the population (the so-called crowding distance).

The crowded comparison operator guides the selection process at the various stages of the algorithm towards an uniformly spread out Pareto-optimal front. Between two solutions with different nondominated ranks, the point with the lower rank is always preferred. Otherwise, if both points belong to the same front, then the point which is located in a region with a lower number of points is preferred (the size of the cuboid enclosing it is larger).

In the algorithm's main loop, a random parent population P_0 is initially created. The population is sorted based on nondomination. Each solution is assigned a fitness equal to its nondomination level (1 is the best level). Thus, minimization of fitness is assumed. Binary tournament selection, recombination, and mutation operators are used to create a child population Q_0 of size *N*. From the first generation onward, the procedure is different. First, a combined population $R_t = P_t \cup Q_t$ is formed. The population R_t will be of size *2N*. Then, the population R_t is sorting according to nondomination. The new parent population P_{t+1} is formed by adding solutions from the first front till the size exceeds *N*. Thereafter, the solutions of the last accepted front are sorted according to the crowded comparison operator and the first *N* points are picked. This is how the population P_{t+1} of size *N* is constructed. This population of size *N* is now used for selection, crossover and mutation to create a new population Q_{t+1} of size *N*.

In the modelling of the supply chain, profit and looses can be calculated for different scenarios. A GA can be used in order to find in which conditions the profit is bigger with the minimum of losses.

3 Modeling and Parameterization to Manage SC Disruption

There are mathematical models for supply chain, its behaviour as well as its performance indicators [14, 15]. The mathematical evaluation functions to SC parameterization are presented below.

Following the model, in this preliminary work, just to show that this approach can be used in real industrial scenarios, SC disruptions are used to generate solutions and some of them will be optimized applying the NSGA-II, using simulated binary crossover [11] in order to handle the parameters involved in SC problems. Figure 1 shows the outline of the way in which our GA works.

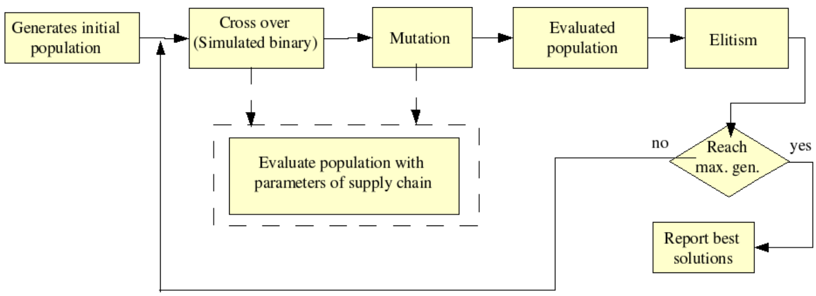


Fig. 1. Flow diagram of a genetic algorithm

In the model, the relationships between providers and clients are taken into account, such that the clients should be satisfied with an opportune and right delivery of the product. Relationships between providers and consumers in different scenarios have been modelled to represent the available scenarios to consumer satisfaction with the product delivery. Special attention is paid on the SC disruption.

As particular instances of the modelling, ad-hoc simulators oriented to a particular industry are implemented. A GA combined with a simulator or with a model of the supply chain where this model plays the role of the chromosome is used. There exist simulators that can estimate the cost and production in the chemical industry with the knowledge of the probability of the disruptions. However, it is not clear what are the best values of the parameters of this simulator to be adopted in order to increase the production while lowering the cost.

3.1 Evaluation Functions for Supply Chain

In order to model the relation between providers and consumers, a reinforcing model proposed by Lawrence et al. [14] is used, they divide SC models based on the underlying optimisation model (facility location) and the risk measure (expected cost). Their models intend to work efficiently in normal conditions as well as with disruptions.

The supply chain is analysed looking for possible disruptions in the provider-consumer network; such a network and possible disruptions is modelled on a virtual supply chain of refineries as follows:

$$\sum_{j\in J} F_j X_j + \sum_{s\in S} q_s \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ijs} \tag{1}$$

Equation 1 represents disruption cost. There are J refineries open, F is the cost during a time period (e. g. 1 day) the refinery is open, X is a Boolean variable with value 1 if refinery is open and 0 otherwise. The value q represents if in the scenario $s \in S$, a disruption occurs; such that it obligates to take a different path to deliver the product in quantity h; d is the path distance. Y is Boolean too so 1 if this path is used, or 0 if not.

One way to obtain a SC disruption model together with the associated cost is by modeling previous disruption experiences such that the obtained knowledge be applied on later similar situations. However, this is not a preventive way to deal with disruptions. A preventive manner is to set a model considering most of the every eventual disruption so that be prepared with recuperation plans, even to the worst case having maximum cost. This cost can be used as a maximum risk measure to deal with any eventual (minor) disruption. Thus, the maximum cost U should be minimized.

Minimize

$$\mathbf{U} \tag{2}$$

Subject to

$$\sum_{j\in J} F_j X_j + \sum_{i\in I}\sum_{j\in J} h_i d_{ij} Y_{ijs} \le U, \forall s \in S \tag{3}$$

$$\sum_{j\in J} Y_{ijs} = 1, \forall i \in I, s \in S \tag{4}$$

$$Y_{ijs}, X_j \in \{0,1\}, \forall i \in I, j \in J, s \in S \tag{5}$$

In this last modeling the scenarios, the probability of disruption occurrence as well as the cost to each is randomly generated. On this context, cost of recuperation to the SC disruption is optimized.

In the simplified model created each scenario represents two kinds of disruptions: 1) a refinery cannot produce, and 2) an inaccessible path to deliver or receive the product. When a refinery is not producing consumers need to be provided by other refineries, it having an extra cost directly related to the distance from the supply refinery and the amount of product required. If the amount required is low, it is preferable to ask for product from open neighbor refineries such that cost remains low. Each neighbor refinery could provide a small quantity of its production in order to not get rid of product for its own distribution. This is simple estimation illustrative to multi-objective minimization; this one will be the standard deviation of product taken from other refineries for the local distribution. An illustrative scenario with five refineries showing distances among them is in Figure 2: a path disruption occurs between R2 and R5 as well as a production disruption in R4.

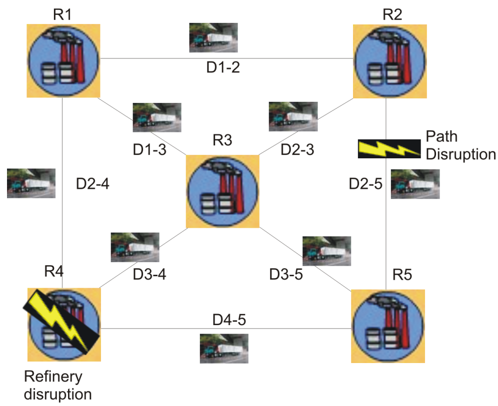


Fig. 2. An illustrative scenario

A general tip during a practical application using a multi-objective GA, is to optimize a low number of objectives simultaneously, such that the computational resources needed can remain manageable, especially when dealing with stochastic variables having a high level of unpredictability. Moreover, when the evaluation of a noisy objective function is used, more care is needed because given the same input parameters, different outputs could be obtained. Multi-objective GAs have a robust behavior because the average performance of a population acts like a noise filter [2, 7]. However, for a general multi-objective optimization, the use of a technique to avoid noisy effects to find a Pareto front with nondominated solutions is needed. To deal with difficult problems, probabilistic selection [8] and partial order [16] techniques have been used. Re-sampling methods have been found to be useful to deal with noise. An additional advantage is to avoid the loss of diversity which is induced by using probabilistic methods [3]. The price to pay is the process time increase. At the moment, because of the bounded number of elements in the SC being modeled, the evaluations time spent is low.

From our perspective, the best thing to do in this case is to fix the values of some SC variables. For that sake, we selected the more stable variables such as the number of refineries, the number of tanks, or the containers capacity. Some other variables, the two or three to be optimized, can be parameterized, e. g. the profit can be handled as a principal objective to be maximized.

4 Test of the Model

To test the model, a randomized parameterization was practiced: the number of disruptions, the probabilities for each disruption as well as the cost associated due to the distances and delivery cost to emergent suppliers is created randomly. Low probabilities to high cost disruptions are assigned. Given certain parameters, the optimization of the profit, number of refineries and product asking to any other refinery could be done.

The remarkable information is that after the algorithm execution, the output shown a consistent behavior by drawing a similar graph like the one in Figure 3. This is a front of Pareto with the set of available optimal solutions. The values do not correspond to a particular input. But the result was the same with same SC model parameterization. These evidences can support the right applicability of NSGA II to optimization practiced having an objective function evaluated in systems with low noise, this means that with the same input parameters, the resultant outputs will be near between them, where near means that the difference is not too much important in practical uses.

There are extreme cases when values change a lot with respect to previous evaluations with the same parameters. These cases are very unlikely. However when this occurs the model can be erratic and the solution could be very different to previous ones.

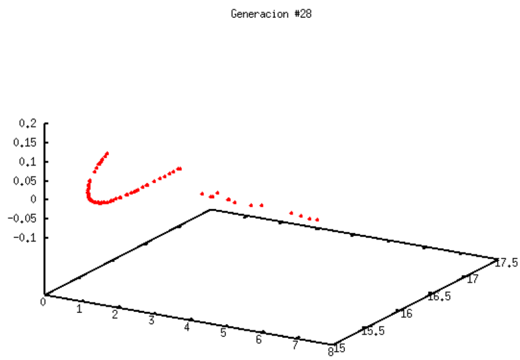


Fig. 3. Optimum values for a specific scenario

5 Conclusions

In the kind of optimization problems, an efficient and reliable algorithm is required to find a set of optimal solutions. This is the reason to use NSGA-II. A possible drawback could be the time required to run many simulations. However, this computational time investment, to assess the decision to build of a new refinery is worth spending, so that we are able to learn if our investment will be quickly recovered.

Acknowledgements

The first author acknowledges support from CONACyT to pursue graduate studies in computer science at CINVESTAV-IPN. The third author acknowledges support from CONACyT project 42435-Y.

References

[1] Bodendorf, F, Zimmermann, R. *Proactive Supply-Chain Event Management With Agent Technology*, International Journal Of Electronic Commerce, 2005.

[2] D. Bueche, P. Stoll, R. Dornberger, and P. Koumoutsakos. Preprint: *Multi-objective evolutionary algorithm for the optimization of noisy combustion processes*. IEEE Transaction on Systems, Man, and cybernetic, 32(4), 2002.

[3] Bui, L. T., Abbass, H. A., and Essam, D. 2005. *Fitness inheritance for noisy evolutionary multi-objective optimization*. In Proceedings of the 2005 Conference on Genetic and Evolutionary Computation (Washington DC, USA, June 25 - 29, 2005). H. Beyer, Ed. GECCO '05. ACM Press, New York, NY, 779-785.

[4] Coello Coello, Carlos A.; Van Veldhuizen, David A. and Lamont, Gary B. *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers, New York, 2002.

[5] John H. Holland. *Concerning efficient adaptive systems.* In M. C. Yovits, G. T. Jacobi, and G. D. Goldstein, editors, Self-Organizing Systems- 1962, pages 215-230, Spartan Books, Washington, D.C., 1962.

[6] John H. Holland. *Outline for a logical theory of adaptive systems.* Journal of the Association for Computing Machinery, 9:297-314, 1962.

[7] E. J. Hughes. *Evolutionary multi-objective ranking with uncertainty and noise.* In Zitzler et al., editor, Proceedings of the First Conference on Evolutionary Multi-Criterion Optimization, Zurich, Switzeland, 2001.

[8] Evan J. Hughes. *Multi-Objective Probabilistic Selection Evolutionary Algorithm.* Technical Report DAPS/EJH/56/2000, Department of Aerospace, Power, & Sensors, Cranfield University, RMCS, Shrivenham, UK, SN6 8LA, September 2000 .

[9] N. Julka, R. Srinivasan and I. Karimi, *Agent-based Supply Chain Management – 1: Framework*, Computers & Chemical Engineering, Vol 26(12), 1755-1769, 2002.

[10] N. Julka, I. Karimi, and R. Srinivasan, *Agent-based Supply Chain Management –2: A Refinery Application,* Computers & Chemical Engineering, Vol 26(12), 1771-1781, 2002.

[11] Kalyanmoy Deb and Ram Bhushan Agrawal. *Simulated Binary Crossover for Continuous Search Space*. Complex Systems, 9:115-148, 1995.

[12] Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, T Meyarivan, *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II (2000).* Proceedings of the Parallel Problem Solving from Nature, VI Conference, 2000.

[13] Lau, HCW, Hui, IK, Chan, FTS, Wong, CWY. *Monitoring the Supply of Products in a Supply Chain Environment: A Fuzzy Neural Approach***.** Expert Systems, 2002.

[14] Lawrence V. Snyder, M. Paola Scaparra, Mark S. Daskin, Richard L. Church. *Planning for Disruptions in Supply Chain Networks*, Tutorials in Operations Research Informs, 2006.

[15] Lee Ying Koo Yuhon Chen et al. *Evaluating Refinery Supply Chain Policies and Investment Decisions Through Simulation-Optimization*, Proceedings of the 2006 Winter Simulation Conference, 2006.

[16] Rudolph, G. *A partial order approach to noisy fitness functions.* In Proceedings of the 2001 Congress on Evolutionary Computation CEC2001 (2001), IEEE Press, Piscataway, NJ, USA, pp. 318-325.

[17] Smirnov AV, Sheremetov LB, Chilov N, et al. *Soft-computing technologies for configuration of cooperative supply chain.* Soft-computing technologies for configuration of cooperative supply chain. Applied Soft Computing 4 (1): 87-107 Feb 2004.

Part XII

Fuzzy Logic and Soft Computing in Distributed Computing

A Framework to Support Distributed Data Mining on Grid

Kun Gao, Lifeng Xi, and Jifang Li

Computer Science and Information Technology College, Zhejiang Wanli University
Ningbo 315100, Zhejiang, R.P. China
gaoyibo@gmail.com

Abstract. In many applications fields, we can obtain benefits from analyzing large distributed data sets by using the high performance computational power. The Grid provides an unrivalled technology for large scale distributed computing as it enables collaboration over the global and the use of distributed computing resources, while also facilitating access to geographically distributed data sets. In this paper, we present a framework for high performance DDM applications in Computational Grid environments called DMGrid, which is based on Grid mechanisms and implemented on top of the Globus 4.0 toolkit.

Keywords: Distributed Computing, Grid Computing, Data Mining, Task Scheduling, Resource Management.

1 Introduction

Many applications are continuing to generate massive data all the time, but because of lack of computation power and collaboration mechanism, we can't obtain fully the potential knowledge in these data that have already spent the huge cost. Without feasible method to process and analyze it, the large cost will not bring the corresponding benefit for our decision. These dataset measured in terabytes or petabytes is very large and come from various applications, including commerce, medical science, scientific experiment, bioscience, etc. They have the inherent property of distribution and heterogeneity. The users who process the data with above characteristics are geographically distributed and large at the same time. The applications that process these data are also expected to have higher performance. At present, existing maturity computation mechanism and technology of data management can't meet requirement described above.

Grid computing [1] is a new platform for distributed process, constituted of heterogeneous computers, and accessed by a general interface. Grid computing, as an important computational model, aims to solve the problem of large scale resource sharing, nontrivial application, innovative applications and high performance application. It is the most advanced scientific research that the grid computing is applied at first. With the maturity of this technology, the grid computing is already applied to coordinated resource sharing and problem solving in dynamic, virtual organizations operating in the industry and business arena.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 489–494, 2007.
springerlink.com

Grid is a natural platform for deploying a high performance application for the knowledge discovery process. On one hand, what the data mining involved is enormous distributed data; On the other hand, its algorithms have higher complexity, need higher computing capability. The grid environment provides coordinated resource sharing, collaborative processing, and high performance computation. So, the combination of these two respects will bring enormous benefit, more rational composition and higher performance.

The outline of the paper is as follows. Section 2 introduces related works and point out the limitations of some previous works. Section 3 describes the rationale of design and development the grid enabled data mining system. Section 4 describes the distributed system framework and main components of DMGrid. Section 5 presents the architecture of tasks scheduling and resource allocation and Section 6 concludes the paper.

2 Related Work

In [2], it is from San Diego Supercomputing Centre which development a middleware to store and access datasets over networks. It is the category of data replication mechanism in fact, for example [3] [4], because it does not handle application implementing in real time. In [5] [6], they are grid computing problem solving environment constructed using MPI and CORBA but is limited to that domain.

In [7], it uses a central server to receive requests and dispatch tasks based on system real time parameters. The main shortcoming of this system is the lack of dynamics. In [8] [9], the systems specialize in parameter-sweep computation, especially supporting dynamic parameters, i.e. parameters whose values are determined at runtime. However, the systems aim at optimizing user parameters and budget for computational tasks only. It has no capability to access remote dataset and optimize the data transfer.

Like [9], [10] provide deployment of parameter-sweep applications on grid. The system emphasizes on data-reuse. The system can appraise the data file that all tasks need, duplicate the data from user node to computation node. When a lot of tasks are assigned to the same resources, it has a try to reuse the data duplicated to make data transmission reduce to minimally. However, the system doesn't support the multiple repositories of data; this method is not applicable to grid.

In this paper, we present a framework for high performance DDM applications in computational grid environments called DMGrid, which is based on grid mechanisms and implemented on top of the Globus 4.0 toolkit. It integrates Grid services by supporting distributed data mining, task scheduling and resource management services that will enlarge the application scenario and the community of Grid computing users.

3 DBGrid Requirements

The rationale of design and development the grid enabled data mining system is as follows:

- DMGrid adopts the standard, common and open grid service mode, follows OGSA norm, and offers unified support to the data mining applications.
- Based on Globus Toolkit and according to the existing networks system structure, DMGrid use the grid service to realize communication, operates each other and resource management.
- DMGrid is open, supports various data mining tools and algorithms, the extensibility is good.
- DMGrid is able to realize the improvement of performance by increasing network node, high performance computing node and cluster, the scalability is strong.
- DMGrid can deal with distributed huge volumes of high dimensional dataset, support heterogeneity data source.
- The main purpose to design and develop the DMGrid system is to improve the performance.
- Users carry out the data mining tasks in a transparent way; the concrete system structure, operation and characteristic in the grid environment is to be hidden.
- In the field of data mining, the security of the data and personal secrets are a sensitive topic. DMGrid supports the choice of place that the data mining execute.

4 DMGrid Architecture

Fig. 1. describes the distributed system framework that we designed and developed. It is mainly made up by following components:

- DMGrid Client Node: In consideration of ease of use, the system adopts Browser/Server mode. Grid client exchanges information with Grid portal through Internet Explorer browser. Users submit the requirement of data mining and receive the final result at Grid client.
- DMGrid Portal Node: It provides a single access way to distributed data mining application based grid. Users can make use of the whole grid resource transparently through the grid portal. This component is responsible for translating users' demand into the RSL language (Resource Specification Language) that can be recognized by grid, is used for grid resource discovery and grid resource allocation management. The final result is returned to grid portal first, and then returned to users by the portal.
- DMGrid Resource Broker Node and DMGrid Tasks Allocation Broker Node: user's data mining requirement has driven grid resource discovery. According to users' demand condition, DMGrid resource broker looks for the resources which meet the condition in a large number of grid resources, including algorithms, computing capability and data resource. It is an important job that finds appropriate resource [11] [12]. As to any application based on grid, it is first to find appropriate resource, then allocate tasks and management them. It can be predicted that there may be many nodes which meet a condition. Resource broker is used for finding available resource among MDS (Meta Directory Service); mapping between data resource and computing resource, i.e., the task allocation broker is responsible for dispatching a certain task on a certain node.

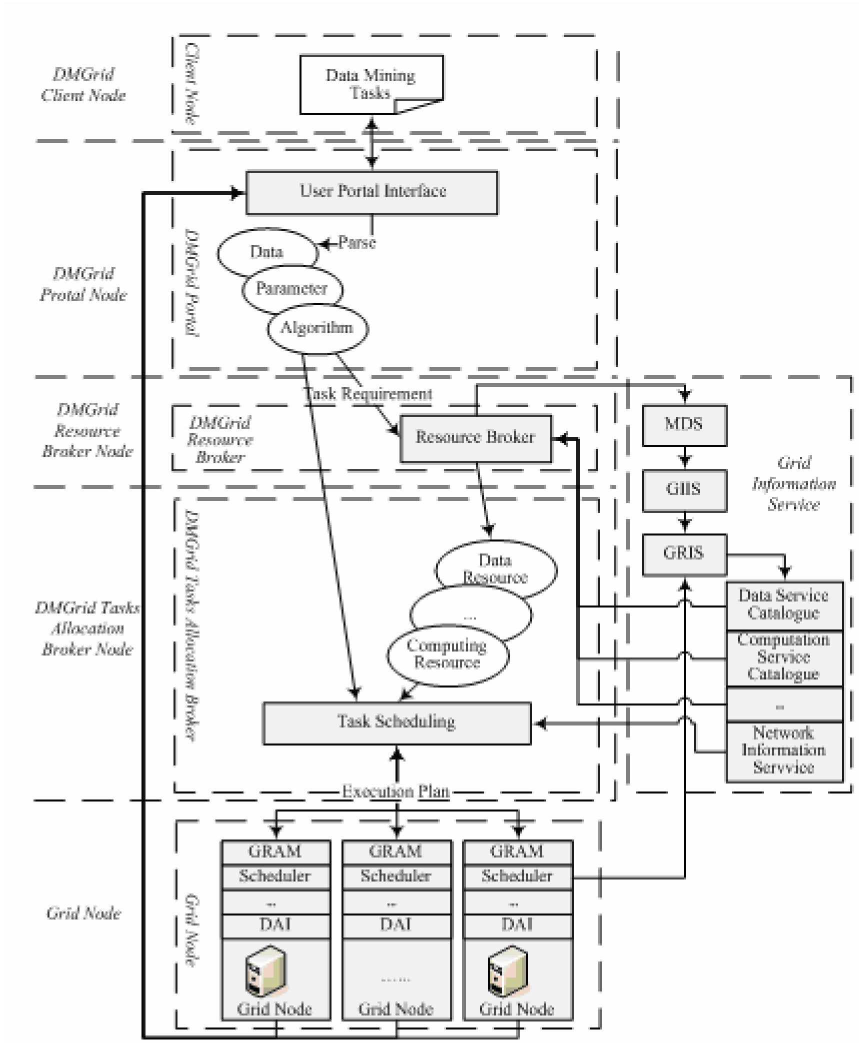


Fig. 1. The framework of distributed data mining on grid

- Grid Node: The Grid nodes are made up of personal computer, high performance computer and cluster. Each node is installed GLOBUS, as grid middleware. They are the data carrier and the computation implementation entity.

5 Tasks Scheduling and Resource Allocation

The broker is responsible for tasks scheduling and resource allocation. It is a core of the whole system. The task allocation procedure adopts the improved taboo search algorithms according to cost matrix, carry out resources mapping on the grid [13] [14].

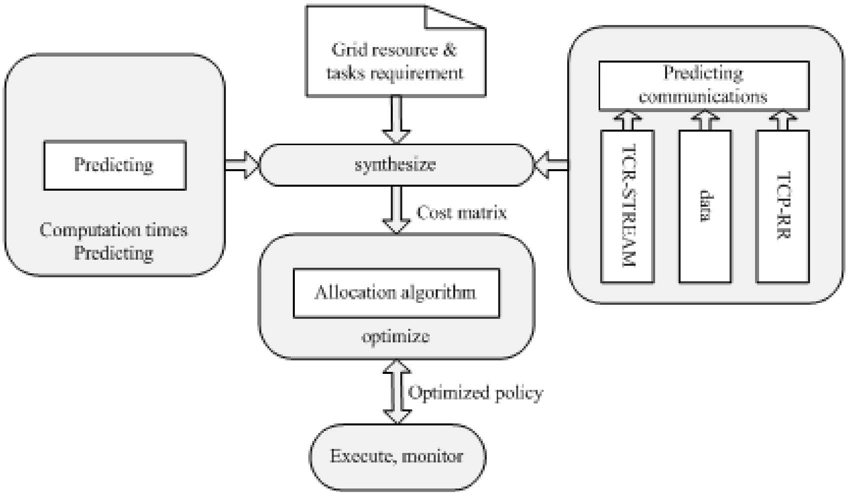


Fig. 2. Architecture of Tasks Allocation

Meanwhile, the establishment of a strategy is according to the real time state queried by the MDS. Broker exchanges information with each GRAM in the grid node to execute a task in coordination. If the grid node is the cluster or the high performance parallel machine, these grid node may have the autonomous local scheduling and allocation. Fig. 2 is the architecture of tasks allocation in the DMGrid.

6 Conclusions

Data mining is a nontrivial process of computation. An efficient data mining application should overcome many factors affecting its performance. With the emergence of globalization grid computing platform, it causes large-scale resource sharing and cooperation becomes possibly. This brings the new vitality for the development of distributed data mining. The grid can provide the high expensive computation resources which the data mining needs. At the same time, the grid environment is consistent with the inherent property of distribution and heterogeneity of data. It is a new trend to merge grid and data mining to meet demands of applications. In this process, optimizing the strategy of tasks allocation is extremely important feasible way to improve the performance of distributed data mining application.

References

1. I. Foster and C. Kesselman, editors. The Grid: Blueprint for a Future Computing Infrastructure. 1999.
2. Asad Samar, Heinz Stockinger. Grid Data Management Pilot (GDMP): A Tool for Wide Area Replication,IASTED International Conference on Applied Informatics (AI2001), Innsbruck, Austria, February 2001.

3. C. Baru, R. Moore, A. Rajasekar, and M. Wan, The SDSC Storage Resource Broker, in CASCON'98 Conference, Toronto, Canada, 1998.
4. A. Chervenak, E. Deelman, I. Foster, L. Guy, W. Hoschek, A. Iamnitchi, C. Kesselman, P. Kunst, M. Ripeanu,B, Schwartzkopf, H, Stockinger, K. Stockinger, B. Tierney. Giggle: A Framework for Constructing Scalable Replica Location Services, Proceedings of Supercomputing 2002 (SC2002), November 2002.
5. G. Allen, W. Benger, T. Goodale, H. Hege, G. Lanfermann, A. Merzky, T. Radke, E. Seidel, J. Shalf, The Cactus Code: A Problem Solving Environment for the Grid, Proceedings of the Ninth International Symposium on High Performance Distributed Computing (HPDC), Pittsburgh, USA, IEEE Press.
6. K.Marzullo, M. Ogg, A. Ricciardi, A. Amoroso, F. Calkins, E. Rothfus, NILE: Wide-Area Computing for High Energy Physics, Proceedings of 7th ACM SIGOPS European Workshop, Connemara, Ireland, 2-4 Sept. 1996, ACM Press.
7. W. Hoschek, J. Jaen-Martinez, A. Samar, H. Stockinger, K. Stockinger, Data Management in an International Data Grid Project, Proceedings of the 1st International Workshop on Grid Computing (Grid 2000, Bangalore, India), Springer-Verlag, Berlin, Germany, 2000.
8. D. Abramson, J. Giddy, and L. Kotler, High Performance Parametric Modeling with Nimrod/G: Killer Application for the Global Grid?, Proceedings of the International Parallel and Distributed Processing Symposium (IPDPS 2000), May 1-5, 2000, Cancun, Mexico, IEEE CS Press, USA, 2000.
9. R. Buyya, D. Abramson, and J. Giddy, An Economy Driven Resource Management Architecture for Global Computational Power Grids, Proceedings of the 2000 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2000), June 26-29, 2000, Las Vegas, USA, CSREA Press, USA, 2000.
10. H. Casanova, G. Obertelli, F. Berman, and R. Wolski, The AppLeS Parameter Sweep Template: User-Level Middleware for the Grid, Proceedings of the IEEE SC 2000: International Conference Networking and Computing, Nov. 2000, Dallas, Texas, IEEE CS Press, USA.
11. Kun Gao, Semantics Based Grid Services Publishing and Discovery, Proceedings of The 1st International Symposium on GRID COMPUTING, Corfu, Greece, August 18, 2005, ISBN 960-8457-32-7, pages 89-93.
12. Kun Gao, Towards Semantic-Driven Grid Resource Discovery, WSEAS TRANSACTIONS on SYSTEMS, Issue 10, Volume 4, October 2005, ISSN 1109-2777, Pages 1668-1675.
13. Kun Gao, Youquan Ji, Meiqun Liu, Jiaxun Chen, Rough Set Based Computation Times Estimation on Knowledge Grid, Lecture Notes in Computer Science, Volume 3470, Jun 2005, Pages 557 - 566
14. Kun Gao, Kexiong Chen, Meiqun Liu, Jiaxun Chen, Rough Set Based Data Mining Tasks Scheduling on Knowledge Grid, Lecture Notes in Computer Science, Volume 3528, May 2005, Pages 150 – 155

Adaptive Processing Scheme of Overflowed Buckets for Bucket Sort Algorithm*

Peng Liu[1], Yu-ping Ma[2], and Jian-she Dong[3]

[1] Grid Research Center, PLA University of Science and Technology, Nanjing 21007, China
[2] Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China
[3] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
liupeng00@mail.tsinghua.edu.org

Abstract. Bucket sort algorithm is an effective approach to sort very large files, whereas the probability of bucket overflow hinders its efficiency. The paper puts forward a more effective bucket sort algorithm, THShort2, which subtly handles the overflowed buckets. For a different degree of bucket overflow, we propose a corresponding processing scheme. The correctness and efficiency of THShort2 is proofed theoretically. The experiment results show that the performance of THSort2 is about triple times of NTSort, and 50% faster than THSort.

Keywords: External Sort, Bucket Sort, THSort, THSort2, NTSort.

1 Introduction

Sorting is one of the most fundamental data processing operations, especially in data mining and database system. Researchers have being put forward many sort algorithms. If the size of sort target is below GB level, we can adopt the internal sorting algorithm, such as Quick Sort [1], Heapsort [2], Flash Sort [3], Proportion Split/Extend Sort [4], Fast sorting method of separating segment [5] and sorting method by nase distribution and linking [6], etc. Internal sorting is limited by the memory space, so it is not suitable for very large data file. External sort algorithms [7] are to sort large-scale data over several GBs. Multi-line Merging Sort and Bucket Sort, both with two passes, are popular in this field.

THSort [8] algorithm adopts two arrays of RAID (Redundant Array of Inexpensive Disks) and multi-channels parallel I/O, and makes use of multithreading to realize alternant parallel I/O. When an RAID is reading input data, CPU is sorting the data in the memory, and another RAID is outputting the data has been sorted. As a result, THSort gets a maximum parallelization and won the 2002 PennySort competition [9]. Another sort algorithm, SheenkSort [10], based on bucket sort algorithm and statistic method, gets a even higher performance and won the 2003 PennySort competition.

* This work was supported by the National Science Foundation of China under the grant No. 60403043.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 495–503, 2007.
springerlink.com

The disadvantage of SheenkSort, is that the algorithm can not insure that the bucket will not overflow, even if the raw data is completely random. Found an effective method to resolve the problem of bucket overflow, this paper bring forward a new external bucket sort algorithm, THSort2, which is more effective, based on parallel processing and rapidly distributing method.

2 Overflow Probability of Direct Distributing Method

THSort2 algorithm includes two steps. First, it distributes each record of the data file to M subfiles or buckets according to its hash value. Hash value of all the records of a subfile will be bigger than those of its former subfiles. Second, it sorts each bucket in turn and link them together one by one to consequently get an ordered output file.

Suppose N is total number of records in the origin file, M is the size of bucket, then $\lceil \log_2 M \rceil$ bit can decide each record belong to any subfile. Because the data is completely random, the probability of a record distributed into a bucket is $1/M$. We use $R_{\max}$ to represent the maximal capability of bucket, and q_1 for the overflow probability. Thus we have

$$q_1 = \sum_{i=R_{\max}+1}^{N} \frac{N!}{i!(N-i)!}(\frac{1}{M})^i(1-\frac{1}{M})^{N-i} \approx \sum_{i=R_{\max}+1}^{N} \frac{\sqrt{2\pi N}(\frac{N}{e})^N}{i!\sqrt{2\pi(N-i)}(\frac{N-i}{e})^{N-i}}(\frac{1}{M})^i(1-\frac{1}{M})^{N-i}$$
$$< \frac{\lambda^\delta}{(\lambda+1)(\lambda+2)...(\lambda+\delta)} e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^i}{i!} = \frac{\lambda^\delta}{(\lambda+1)(\lambda+2)...(\lambda+\delta)} \quad (1)$$

of which $\delta = R_{\max} - \lambda$, only consider the latter $\delta/2$ items less than or equal to $\left(\frac{\lambda}{\lambda+\delta/2}\right)^{\delta/2} = (\frac{2\beta}{1+\beta})^{\frac{1-\beta}{2}R_{\max}}$, this $\beta = \lambda/R_{\max} \in (0,1)$ is the file utilization, $(2\beta/1+\beta) < 1$, and q_1 decrease as $R_{\max}$ increases. Enlarged the memory can decrease overflow

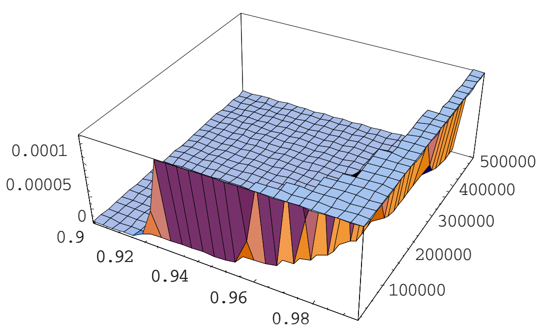


Fig. 1. The bucket overflow probability of file utilization and bucket capability as to set different value

probability of the subfile, but when $R_{\max}$、N is changelessness, it must increase amount of the subfile to realize. And then we have overflow probability less than or equal to $q_1 = 1-(1-(\frac{2\beta}{1+\beta})^{\frac{1-\beta}{2}R_{\max}})^M$. Fig. 1 shown image of the function $q_1 = 1-(1-(\frac{2\beta}{1+\beta})^{\frac{1-\beta}{2}R_{\max}})^M$, in which $\beta \in (0.9, 0.99)$, $R_{\max} \in (10000, 500000)$.

We can learn from Fig. 1 that q_1 increase as β increase. And when β is under proper condition, the larger is $R_{\max} = \frac{\lambda}{\beta}$, the less is the overflow probability, even approach 0 in the end. In a similar way, gotten different probability of $R_{bucket} > 2R_{\max}$ and $R_{bucket} > 3R_{\max}$ less than or equal to

$$q_2 = 1-(1-(\frac{2\beta_2}{1+\beta_2})^{(1-\beta_2)R_{\max}})^M \quad (\beta_2 = \beta/2) \tag{2}$$

and

$$q_3 = 1-(1-(\frac{2\beta_3}{1+\beta_3})^{\frac{3(1-\beta_3)R_{\max}}{2}})^M \quad (\beta_3 = \beta/3) \tag{3}$$

3 Scheme to Handle Overflow Buckets

According to the experience of THSort algorithm, the computer on sorting has three components of disposal: two arrays of RAID and a CPU, therefore, divided the memory into three blocks as the work space of respective component of disposal. Now given some marks will be used in the context:

r : *read* w : *write* I : *Input* O : *Output* S : *Sort* M : *Merge*

Ir: as to *I*nput data, *r*ead data from the current component

Iw: *I*nput No.i block data *w*rite into current component

Or: as to *O*utput data, *r*ead data from the current component

Ow: input data to write into the current component

$\frac{1}{2}Mr$: Merging two blocks of ordered data, of which a data block come from itself equipment

$\frac{1}{3}Mr$: Merging three blocks of ordered data, of which a data block come from itself equipment

Mw: write merged data block to itself equipment

In order to analyze conveniently and do not lose rationality, we approximatively think the whole operations as stated above expend a slice of time T to the elementary data block:

$$t_{Ir}(B) = t_{Iw}(B) = t_S(B) = t_{Or}(B) = t_{Ow}(B) = t_{\frac{1}{2}Mr}(B) = t_{\frac{1}{2}Mr}(B) = t_{Mw}(B) = T \quad (4)$$

Set up R_{bucket} as the actual quantity of data which is the bucket occurred overflow, and of R_{bucket} different to $R_{\max}$, we will offer different method to resolve:

(1) As $R_{bucket} \le R_{\max}$,the bucket do not occurred overflow. And at the moment all of components can operate on pipeline method, the efficiency of the algorithm is the maximum, a data block B each managed only expend a slice of time T.

(2) As $R_{\max} < R_{bucket} \le 2R_{\max}$, sorting again the overflow bucket on the bucket sort algorithm and separate to two blocks of memory, it is putted differently out RAID2 after sort, however, this operation can not avoid to occur overflow of the memory. Therefore, before dealing farther with the overflow bucket, utilize the algorithm to sort one by one then to merge rather than use the bucket sort algorithm again, thus it sure that the farther overflow would not occur. Through analysis, we can know dealt with double overflow bucket need expend five slices of time. As shown as the Fig. 2.

Time	Array1	Memory1	Memory2	Memory3	Array2
T1	Ir	Iw			
T2	Ir	S	Iw		
T3	Ir	Or	S	Iw	Ow
T4			Or	S	Ow
T5	Ir		Iw	Or	Ow
T7	Ir		S	Iw	
T8				S	
T9			1/2 Mr	1/2 Mr	Mw
T10	Ir	Iw	1/2 Mr	1/2 Mr	Mw
T11	Ir	S	Iw		
T12	Ir	Or	S	Iw	Ow
T13			Or	S	Ow
T14				Or	Ow

Fig. 2. The processing method as double overflow

(3) If $2R_{\max} < R_{bucket} \le 3R_{\max}$,for there are three memory exist in the system, we can also divide the spillover barrels to three memory, this can still avoid affecting the efficiency while dealing with spillover barrels in turn, through analyzing, the conclusion is that the processing time is seven timeslice, show as Figure 3.

(4) If $R_{bucket} > 3R_{\max}$, the spillover barrels data can't be accommodated in Memory, it's necessary for external storage to be used as buffer storage, here is the general algorithm.

As usual, if $(H-1)R_{\max} < R_{bucket} \le H \times R_{\max}$, the first step is not only divide the data in spillover barrels to H segments but also array and write them in another, with

Time	Array1	Memory1	Memory2	Memory3	Array2
T1	Ir	Iw			
T2	Ir	S	Iw		
T3	Ir	Or	S	Iw	Ow
T4	Ir	Iw	Or	S	Ow
T5	Ir	S	Iw	Or	Ow
T6	Ir		S	Iw	
T7				S	
T8		1/3 Mr	1/3 Mr	1/3 Mr	Mw
T9		1/3 Mr	1/3 Mr	1/3 Mr	Mw
T10		1/3 Mr	1/3 Mr	1/3 Mr	Mw
T11	Ir	Iw			
T12	Ir	S	Iw		
T13	Ir	Or	S	Iw	Ow
T14			Or	S	Ow
T15				Or	Ow

Fig. 3. Processing method as twice overflow

H forming an orderly son paper, the method in this process is similar with $R_{bucket} \le R_{\max}$. The second step requires the use piled classification algorithms to order H pieces of orderly sub-files to form a ordered file in length of R_{bucket}.

The algorithm of the second step is showed as followings, assuming that:

$A_i(p_i, m_i) \quad (i = 1,2,3...H)$ is the No. i orderly part defined by p_i and m_i, where p_i is the pointer that point to current position and m_i is the length of the initial data;

$B_i(ps_i, n_i, V) \quad (i = 1,2,3...H)$ is the i piece of memory defied by ps_i, n_i, and V, where ps_i point to the first data unordered, n_i is the data quantity of current memory, and V is the maximum of it;

$C(p, W)$ is the memory defied by p and W, where p is the current pointer, W is the capacity of C. Its chief function is to storage the combination result, and output the result to the external storage when spilled over.

HeapMerge ($A_1(p_1,m_1), A_2(p_2,m_2), A_3(p_3,m_3),...A_H(p_H,m_H)$)

Input：Part document orderly

Output： And the formation of a joint paper documents one orderly

(1) Substantially Reactor

(2) WHILE ($\exists B_i(B_i.ps_i == B_i.n_i)$ and (B_i is unmarked))

(3) IF $(A_i.p_i < A_i.m_i)$

(4) $size = MIN(V, A_i.m_i - A_i.p_i)$

(5) $COPY(B_i.ps_i, A_i.p_i, size)$

(6) $B_i.n_i = size$

(7) $A_i.p_i += size$

(8) ELSE

(9) mark B_i

(10) IF (there are some data in the stack)

(11) Output data pointed by $C.p$, and supply from data pointed by $B_j.ps_j$

(12) $B_j.ps_j{++}$

(13) $C.p{++}$

(14) ELSE

(15) GOTO (20)

(16) IF $(C.p == W)$

(17) OUTPUT(C,W)

(18) $C.p = 0$

(19) GOTO (2)

(20) IF $(C.p \neq 0)$

(21) OUTPUT(C, $C.p$)

4 Performance Analysis

According to the conclusions of previous, clearly, Record the rate of number $< R_{\max}$, $R_{\max} \sim 2R_{\max}$, $2R_{\max} \sim 3R_{\max}$ respectively not more than. Optimization algorithms using existing merger deal with $< R_{\max}$, $R_{\max} \sim 2R_{\max}$, $2R_{\max} \sim 3R_{\max}$, Each piece requires one time, 5 time films and seven hours films. However the rate of exceed $3R_{\max}$ comparing to $R_{bucket} < 3R_{\max}$ is a senior endless small, negligible. Then we can calculate the everage time E to deal with one barrel, in order to get E, we should calculate E_1 $\left(1 \le E \le E_1\right)$:

$$
\begin{aligned}
E_1 &\approx (1-q_1)*1+(q_1-q_2)*5+(q_2-q_3)*7 \\
&= -4(1-(\frac{2\beta}{1+\beta})^{\frac{1-\beta}{2}R_{\max}})^M - 2(1-(\frac{2\beta_2}{1+\beta_2})^{(1-\beta_2)R_{\max}})^M + 7(1-(\frac{2\beta_3}{1+\beta_3})^{\frac{3(1-\beta_3)}{2}R_{\max}})^M
\end{aligned} \tag{5}
$$

Put $\beta=\frac{\lambda}{R_{\max}},\beta_2=\frac{\lambda}{2R_{\max}}=\frac{\beta}{2},\beta_3=\frac{\lambda}{3R_{\max}}=\frac{\beta}{3},R_{\max}=\frac{1}{\frac{\lambda}{R_{\max}}}\bullet\lambda=\frac{1}{\beta}\bullet\lambda,M=256$ in (5), then

$$E_1=-4(1-(\frac{2\beta}{1+\beta})^{\frac{1-\beta}{2}R_{\max}})^{256}-2(1-(\frac{2\beta}{2+\beta})^{(1-\frac{\beta}{2})R_{\max}})^{256}+7(1-(\frac{2\beta}{3+\beta})^{\frac{3-\beta}{2}R_{\max}})^{256}.$$

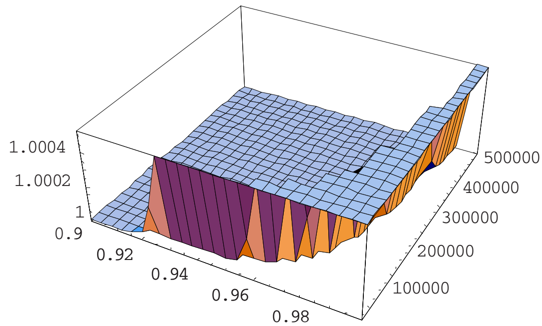


Fig. 4. The State of how E_1 changes with $R_{\max}$

Fig. 4 indicates the state of how E_1 changes with matter utilization β and the capability $R_{\max}$. It shows that, E_1 increases with β, if β is invariable, E_1 will decrease with the increase of total record, and ultimately become 1, that is to say $E_1\to 1$, so $E\to 1$, this can prove rationality of this algorithm. We can draw the conclusion that: time performance of this algorithm is almost linear, namely, time performance for treating N piece of record is $O(N)$.

The spillover probability is very small, but even a spillover, can lower the costs of exceptions to that in a performance analysis can not consider the spillover. Suppose data L is the length of each record, $L_1=\lceil\log_2 M\rceil$ is the length of key words,then $L-L_1$ is the additional data,we needn't compare the additional data when treat every record, the total time to treat with N pieces of recor is $O(\frac{L_1}{L}N)$. If the record quantity of the NO. i barrel is $k_i(i=1,2,...M)$,the total sort time is also $O(\sum_{i=1}^{M}k_i\log k_i)<O(MR_{\max}\log R_{\max})$, then the total time to treat with N pieces of records is:

$$O(\frac{L_1}{L}N)+O(\sum_{i=1}^{M}k_i\log k_i)<O(\frac{L_1}{L}N+MR_{\max}\log R_{\max})\approx O(MR_{\max}\log R_{\max})$$
$$(\log R_{\max}>>\frac{L_1}{L},MR_{\max}>N) \qquad (6)$$

Thus we can draw the conclusion that ,without spillover, the time performance of THSort2 is $O(N*\log_2(M))$ faster than the one of SheenkSort. However, the rate of spillover is too small to calculate when compare the time performance. So it's deemed that the performance of THSort2 is higher than SheenkSort by $O(N * \log_2(M))$.

5 Experiments and Results

The performance testing is on a RAID-M memorizer made by ourselves. Its configuration is 2 GB DDR 400 memories, 3.0GHz Xeon CPU, PCI Express I/O connection channels, two hard disk array cards, each constitutes RAID5 by seven hard disks, so they constitute two I/O channels that not be restricted by the PCI. Because the result of the test is related to the environment very much, this paper emphasizes particularly on comparing the performance of THSort2 to the basic algorithms NTSort (Since SheenkSort is a open source software, so this paper doesn't compare the performance to it).

Fig. 5 shows the results. We can see the performance of THSort2 is about three times of NTSort, and this means it is 50% faster than THSort.

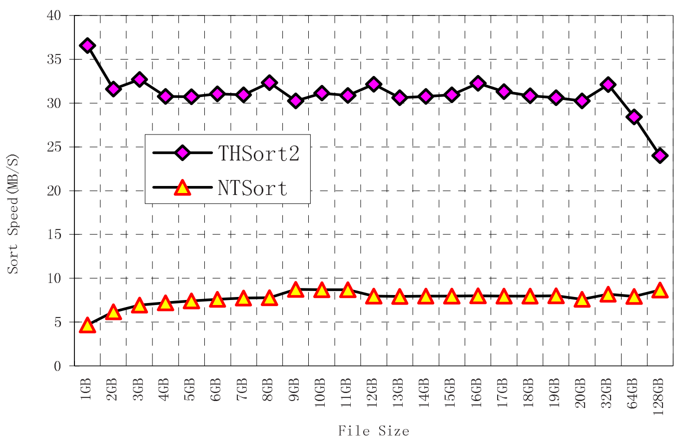


Fig. 5. Performance comparison of THSort2 and NTSort

6 Conclusion

This paper makes the further studying on the basis of THSort2 and NTSort. The analysis of the theory shows that no matter how big the SheenkSort sets the swatch, the bucket flowing can not be completely avoid. So we need high efficiency algorithms to solve the problem of bucket flowing. This paper use three Buffer Storages, masterly gives the arrangement way that the flowing bucket is smaller than three times of the length of the bucket, and the general arrangement way that the flowing bucket is bigger than three times of the length of the bucket. In fact, if the bucket flows, the rate of the size of it smaller than three times of it is mostly happen, so this algorithm is high efficiency on the statistics.

On the basis of high efficiency solving the bucket flowing, this paper analyzes the distribute point sample statistics of SheenkSort, and considers that to the basic random data, can completely use hash function do the rapid distribution of data without wasting extra time for the sample statistics. So THSort2 is faster than SheenkSort by $O(N * \log_2(M))$ on theory. Performance tests showed that the performance of THSort2 is about three times of the benchmark algorithms NTSort, and is around 50% faster than THSort.

References

1. Knuth, D. E.: The Art of Computer Programming (Volume 3/Sorting and Searcing). 2nd ed., New York: Addison-Wesley Publishing Company, 1973. 114~116.
2. Baase, S.: Computer Algorithms: Introduction to Design and Analysys. 3rd ed, English Reprint Copyright by PEARSON EDUCATION NORTH ASIA LIMITED and HIGHER EDUCATION PRESS, 2001. 182~197.
3. Neubert, K. D.: The flash sort algorithm. Dr. Dobb's Journal, 1998,23(2):123~129.
4. Chen, J. C.: Proportion extend sort. SIAM Journal on Computing, 2001,31(1):323~330. (in Chinese)
5. Tang, X. Y.: Fast sorting method of separating segment. Journal of Software, 1993,4(2):53~57 (in Chinese).
6. Wang, X. Y.: A new sorting method by nase distribution and linking, Chinese Journal of Computers, 2000,23(7):774~777 (in Chinese).
7. Elmasri, R., and Navathe, S. B.: Fundamental of Database Systems. 2nd ed., Redwood City, CA: Benjamin/Cummings Publishing Company, 1994. 83~84.
8. Shi, Y., Zhang, L., and Liu, P.: THSort: A Single-Processor Parallel Sorting Algorithm, Journal of Software, 2003, 14(2): 159~165. (in Chinese)
9. Sort Benchmark, http://research.microsoft.com/barc/SortBenchmark/
10. Yang, L., Huang, H., and Song, T.: The Sample-Seperators Based Distributing Scheme of the External Bucket Sort Algorithm. Journal of Software, 2005,16(5):643~651. (in Chinese)

A Novel Fractal Image Coding Based on Quadtree Partition of the Adaptive Threshold Value

Liangbin Zhang and Lifeng Xi

Computer Science and Information Technology College, Zhejiang Wanli University,
Ningbo, Zhejiang, 315100, R.P. China
zlb@zwu.edu.cn

Abstract. Fractal image coding is a novel technique for still image compression. Compared with the distance between the range block and the matching domain block, setting of the initial threshold value is one of the most difficult problems in Fisher Quadtree-based fractal image coding. In this paper, a novel fractal image coding based on Quadtree partition of the adaptive threshold value is proposed. Considering the input image feature fully, we put forward the computation derivation process of the adaptive threshold value progressively and declare that the adaptive threshold value has the direct proportion with the variance of the current range block. Experimental results show that compared with Fisher Quadtree-based fractal coding for the same image, the proposed coding scheme obtains better performance including the improved quality of the decoded image, shorter compression time and higher compression ratio.

Keywords: fractal image coding, Quadtree partition, adaptive threshold.

1 Introduction

Fractal image compression, which is based on the IFS (Iterated Function System) proposed by Barnsley[1], is a novel approach to image coding. Its performance relies on the presence of self-similarity between the regions of an image. Since most images process a high degree of self-similarity, fractal compression contributes an excellent tool for compressing then[2-3]. Recently, there are several methods[4] subsequently proposed to improve the performance of fractal image compression. In the range and domain block mapping, several other functions have been proposed in the literatures. Besides, various approaches are also proposed to reduce the searching within the domain pool. Among all fractal block coding schemes, the technique of variable-size blocking is included to compromise the compression ratio and the level of quality. Range block segmentation is important to code image for saving bit rate. Quadtree segmentation is a common method to partition image, since its flexibility and less overhead. In Fisher Quadtree-based fractal image coding[5-6], the threshold value compared with the distance between the range block and the searching domain block is setup by manual experimental experience fixedly. Hence experimental results will be greatly influenced via the threshold minor variety. How to get the adaptive threshold value corresponding to the current range block is one of the most difficult problems in the fractal image coding.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 504–512, 2007.
springerlink.com

In this paper, a novel fractal image compression based on Quadtree partition of the adaptive threshold value is proposed. Considering the input image feature fully, we put forward the computation derivation process of the adaptive threshold value progressively and declare that the adaptive threshold value has the direct proportion with the variance of the current range block. Experimental results show that we improve the performance of Quadtree segmentation by adapting the threshold value among each level of the Quadtree.

The balance of the paper is organized as follows: theoretical foundations including basic fractal image coding and Fisher Quadtree-based coding are stated in Section 2. Proposed methodology is described in Section 3. Experimental results and discussion is reported in Section 4. Conclusion is included in Section 5.

2 Theoretical Foundations

2.1 Basic Fractal Image Coding

Fractal image coding makes good uses of image self-similarity in space by ablating image geometric redundant. Fractal coding process is quite complicated but decoding process is very simple, which makes use of potentials in high compression ratio. The main theory of fractal image coding is based on Local Iterated Function System, attractor theorem, and collage theorem. Regard original compressible image as attractor, how to get LIFS parameters is main problem of fractal coding.

We explain the basic procedure for the fractal image coding [7].

1. A given image I is divided into non-overlapping M range blocks of size $B \times B$ and into arbitrarily located N domain blocks of size $2B \times 2B$. The range blocks are numbered from 1 to M, and represented by $R_i(1 \le i \le M)$. Similarly, the domain blocks are from 1 to N, and represented by $D_j(1 \le j \le N)$.
2. For each range block R_i, the best matched domain $D_k(1 \le K \le N)$ and an appropriate contractive affine transformation τ_{ik} which satisfy the following equation are found through

$$d(R_i, \tau_{ik}(D_k)) = \min d(R_i, \tau_{ij}(D_j)) \tag{1}$$

Where τ_{ij} is an contractive affine transformation from the domain block D_j to the range block R_i; the distortion measure $d(R_i, \tau_{ij}(D_j))$ is the Mean Square Error (MSE) between the range block R_i and the contractive domain block $\tau_{ij}(D_j)$. The contractive affine transformation τ_{ij} is composed of two mappings ϕ_j and θ_{ij} as follows:

$$\tau_{ij} = \theta_{ij} \circ \phi_j \tag{2}$$

The first mapping ϕ_j is the transformation of domain-block size to the same size as range blocks. This transformation can be described as follows: The domain block

D_j is divided into non-overlapping unit blocks of size 2×2; and each pixel value of the transformed block $\phi_j(D_j)$ is an average value of four pixels in each unit block in D_j. The second mapping θ_{ij} consists of two steps: The first step transforms the block $\phi_j(D_j)$ by one of the following eight transformations: rotation around the center of the block $\phi_j(D_j)$, through 0^0,$+90^0$,$+180^0$, and $+270^0$, and each rotation after orthogonal reflection about mid-vertical axis of the block $\phi_j(D_j)$.Those eight transformations are called isometries. The second step is the transformation p_{ij} of pixel values of a block obtained by the first step. This transformation p_{ij} is defined as

$$p_{ij}(v) = s_{ij}v + h_{ij} \tag{3}$$

where v is a pixel value of the block obtained by the first step, and the parameters s_{ij} and h_{ij} are computed by the least square analysis of pixel values of the range block R_i and the block obtained by the first step. We call the parameters s_{ij} and h_{ij} a scaling coefficient and an offset, respectively.

The LIFS parameters listed below are encoded:

(1) Parameters to indicate a location of the best matched domain block;
(2) A parameter to indicate an isometric on the best matched domain block;
(3) A scaling coefficient and an offset.

The proposed method quantizes these LIFS parameters [7].

2.2 Fisher Quadtree-Based Fractal Image Coding

In the basis automatic fractal image coding, they cannot represent the image features well because an image is divided into the blocks of a fixed size. In fact, if a larger block is used, it is difficult to find a good matching for areas, which have fine details and complex regions such as Lena's eye. It will lead to loss in quality in the decoding image. On the other hand, if a smaller block is used, although a good matching can be found, as many smooth areas such as Lena's background are divided to many blocks while they are well fitted with a larger block, the number of blocks will increase. Hence, it causes less compression ratio as well as larger computation time.

Jacquin and Fisher gained a good deal of enlightenment from the Quadtree method of a grey image and solved the above problem by means of the Quadtree method to separate such an image[5-7]. Quadtree segmentation expresses an image with a tree structure[8]. Fig.1 shows this spatial structure. On the top it is a father node which has 4 son-nodes corresponding to 4 blocks of image. And each son-node has 4 son-nodes of its own which maps to the 4 quadrants of next subblock image. The root of Quadtree is original image. Fig.2 shows quadrants of a block and their labels. Fig.3 shows the divided image and its corresponding Quadtree.

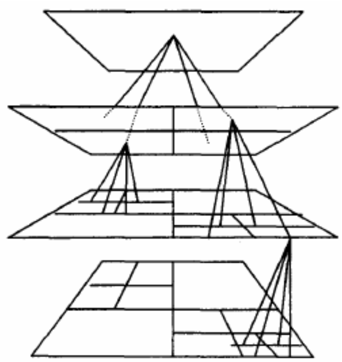


Fig. 1. Image quadtree segmentation

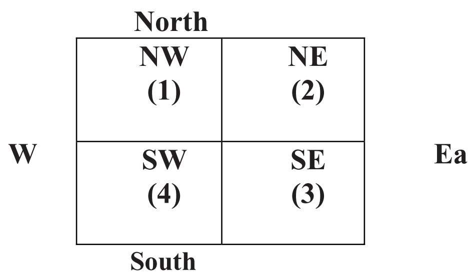


Fig. 2. Quadrants of a block and their labels

Fisher Quadtree-based fractal image coding is described as followed:

Before division, we first set maximal and minimal depth of the Quadtree and a maximal allowable fixed threshold to decline the number of range blocks. Then we continuously partition a range into four square ranges of the same size by the Quadtree

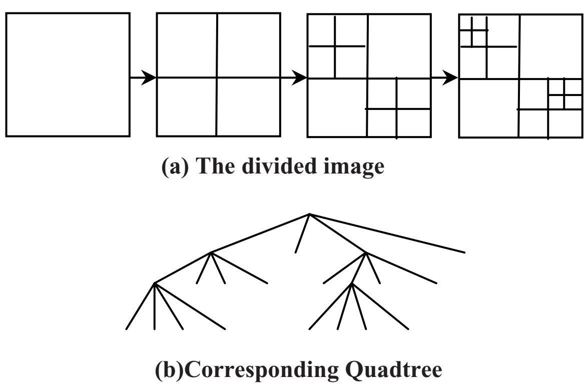


Fig. 3. The divided image and its Quadtree

method until minimal depth is met. An optimum matching block will be can be marked as D_j and the range corresponding with it can be marked as R_i and the partition is not done again. Otherwise they are further partitioned into four ranges. This process continues until the minimal depth is met.

3 Proposed Methodology

In Fisher Quadtree-based fractal image coding, the fixed threshold value decides the number of the range block partition and affects coding efficiency directly. The larger value the fixed threshold is, the more numbers range blocks with large size has, and compression ratio improves while PSNR of the decoded image decreases. The smaller value the fixed threshold is, the less numbers range blocks with large size has, and PSNR of the decoded image improves while compression ratio decreases. The threshold value is the key to Fisher Quadtree-based fractal image compression[9].

Considering the input image feature fully, we put forward the computation derivation process of the adaptive threshold value progressively and declare that the adaptive threshold value has the proper proportion with the variance of the current range block. Some basic definitions is described as following:

Average grey value of the range block R_i is defined as $\overline{a}$:

$$\overline{a} = \frac{1}{N}\sum_{i=0}^{N-1} a_i \tag{4}$$

Average grey value of the domain block D_i is defined as $\overline{b}$:

$$\overline{b} = \frac{1}{N}\sum_{i=0}^{N} b_i \tag{5}$$

Variance of the range block R_i is defined as σ_r^2 :

$$\sigma_r^2 = \frac{1}{N}\sum_{i=0}^{N-1}(a_i - \overline{a})^2 \tag{6}$$

Variance of the range block D_i is defined as σ_d^2 :

$$\sigma_d^2 = \frac{1}{N}\sum_{i=0}^{N-1}(b_i - \overline{b})^2 \tag{7}$$

Covariance between the range block and the domain block is defined as $\mathrm{cov}(a,b)$:

$$\mathrm{cov}(a,b) = \frac{1}{N}\sum_{i=0}^{N-1}(a_i - \overline{a})(b_i - \overline{b}) \tag{8}$$

The distance of Mean Square Error (MSE) between the range block and the searching domain block is defined as $dis(a,b)$:

$$dis(a,b) = \frac{1}{N}\sum_{i=0}^{N-1}[a_i - (sb_i + h)]^2 \tag{9}$$

Relative covariance between the range block and the domain block is defined as $\mathrm{cov}_R^2(a,b)$:

$$\mathrm{cov}_R^2(a,b) = \frac{\mathrm{cov}^2(a,b)}{\sigma_d^2} \tag{10}$$

From the given definition, conclusion is obtained that according to a given range block, function $dis(a,b)$ the distance between the range block and the searching domain block is relative with $\mathrm{cov}_R^2(a,b)$ and also satisfied with form,

$$dis(a,b) = \sigma_r^2 - \mathrm{cov}_R^2(a,b) \tag{11}$$

Here, for each of the range block, σ_r^2 is a constant and function $dis(a,b)$ has only be relative with $\mathrm{cov}_R^2(a,b)$. Moreover, function $dis(a,b)$ has relation with the function $f(x) = m \cdot e^{-k^2x^2}$ $(a > 0, k > 0)$ in close coordination. When condition is included as follows: $m = \sigma_r^2, mk^2 = 1$, $dis(a,b)$ is main component for the Maclaurin expansion of function $f(x)$.

$$f(x) = m - mk^2x^2 + R_n(x) \tag{12}$$

Here, $R_n(x)$ is the rest component for the Maclaurin expansion of function $f(x)$. Given $m = \sigma_r^2, mk^2 = 1, x = \mathrm{cov}_R(r,d)$, we can get form

$$f(cov_R(a,b)) = \sigma_r^2 - cov_R^2(a,b) + R_n(cov_R(a,b)) = dis(a,b) + R_n(cov_R(a,b)) \tag{13}$$

Form (13) shows the threshold of $dis(a,b)$. Fig.4 shows function $f(x)$ has relation with the function $dis(a,b)$.

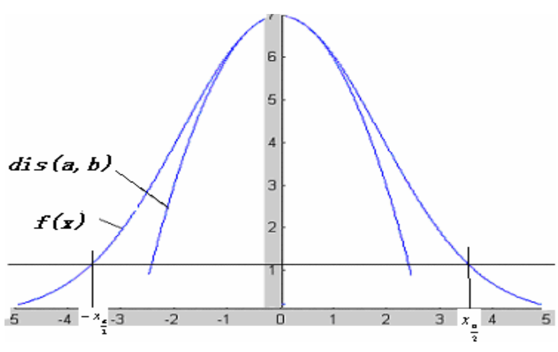


Fig. 4. Function f(x) relation to dis(a,b)

During the interval range $(-\infty,+\infty)$, definite integral of *f(x)* is $\int_{-\infty}^{=\infty} f(x)dx = \frac{m}{k}\sqrt{\pi}$. In probability theory , believe metric of $(1-\alpha)$ of the function *f(x)* is $\int_{-x_{\frac{a}{2}}}^{x_{\frac{a}{2}}} f(x)dx = (1-a)\int_{-\infty}^{+\infty} f(x)dx$. So $f(x_{\frac{a}{2}}) = \sigma_r^2 e^{-\frac{x_a^2}{2}}$. Finally we regard $f(x_{\frac{a}{2}}) = \sigma_r^2 e^{-\frac{x_a^2}{2}}$ as the adaptive threshold of $dis(a,b)$.

4 Experimental Results

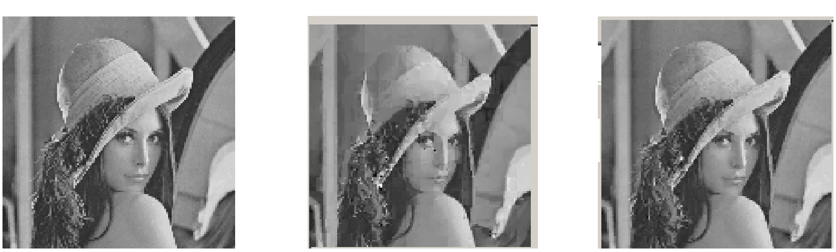
All the experiments are carried out on a computer with Intel 2.5Ghz and 512M RAM in the Win2000 professional operating system and VC6.0 language is used. Original image is classical 128×128 grey-level Lena face image coded with 8 bits per pixel.

An optimal bit allocation strategy is as follows: 14 bits for the location of the matched domain block (horizontal and vertical coordinate), 3 bits for isomorphic types, 5 bits for contrast scaling and 7 bits for the offset, 3 bit for the depth of the Quadtree. For each of the range block, fractal code includes 32 bits allocation via writing into a text file as a fractal coding file. During the iteration process of the image decoding, those grey value either exceeding integer 255 or less than integer 0 is replaced by the average of its four neighbors to avoid block diverging.

In Fisher Quantree-based fractal image coding, our setting is that the maximal range size is 64×64, and the minimal range size is 2×2.The fix threshold value is setup up by real number 0.5 and 0.2 respectively. Table.1 shows the detail experimental data and Fig.5 shows the decoded image of 10 iteration.

We can see from the experimental data. The larger value fixed threshold is, the more numbers range blocks with large size has, then compression ratio improves while PSNR of the decoded image decreases. The smaller value fixed threshold is, the less numbers range blocks with large size has, then PSNR of the decoded image improves while compression ratio decreases. For obtaining the matching domain of Lena face image, range blocks of eye and fair regions are partitioned into small size and range blocks of shoulder and background are partitioned into large size.

In our coding scheme (fractal image compression based on Quadtree partition of the adaptive threshold value), the maximal and the minimal range block size is as above. The believe metric is setup up by real number 0.92 and 0.96 respectively.



(1)original image, (2)decode image with 0.5 theshold, (3)decoded image with 0.2 threshold

Fig. 5. Fisher Quadtree-based fractal image coding

Table 1. Fisher Quadtree-based fractal image coding

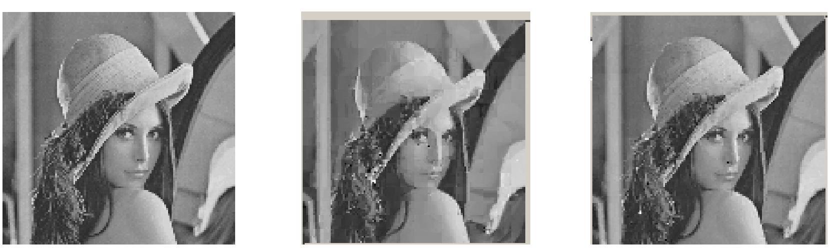
Fixed threshold	0.5	0.2
Different size Range numbers (maximal range size 64*64, minimal range size 2*2)	Total Range number: 460 4*4 range number: 292 8*8 range number: 163 16*16 range number: 5	Total range number: 1021 2*2 range number: 340 4*4 range number: 595 8*8 range number: 86
Coding time	55 S	135 S
Compression ratio	9.82:1	4.42:1
PSNR	25.8	32.2

Table.2 shows the detail experimental data of our coding scheme and Fig.6 shows the decoded image of 10 iteration of our coding scheme.

We can see from the experimental data. For each range block, the adaptive threshold has direct proportion with its variance. The larger variance of the range block is, the larger adaptive threshold is, and the range block may be a midrange or an edge block. The smaller variance of the range block is, the smaller adaptive threshold is, and the range block may be a smooth or shade block. Hence, our coding scheme can be better adapted with human vision characteristic and obtains less distortion of the decoded image. Compared with Fisher coding for the same image, the proposed scheme obtains better performance including the improved quality of the decoded image, shorter compression time and higher compression ratio.

Table 2. Our coding scheme

Believe metric	0.92	0.96
Different size Range numbers (maximal range size 64*64, minimal range size 2*2)	Total range number:420 4*4 range number: 272 8*8 rang number: 144 16*16 range number: 4	Total range number: 986 2*2 range number: 321 4*4 range number: 603 8*8 range number: 62
Coding time	49 S	128 S
Compression ratio	9.91:1	4.62:1
PSNR	26.9	33.6



(1) original image, (2)decoded image with 0.92 believe metric, (3)decoded image with 0.96 believe metric

Fig. 6. Our coding scheme

5 Conclusion

In this paper, a novel fractal image compression based on Quadtree partition of the adaptive threshold value is proposed. We put forward the computation derivation process of the adaptive threshold value progressively and declare that the adaptive threshold value has the direct proportion with the variance of the current range block, so that threshold rely on manual experimental experience is solved. Experimental results show that we improve the performance of Quadtree segmentation by adapting the threshold value among each level of the Quadtree. How to reduce computing variance time of each range block and make effective classification searching among the region, how to make use of the fractal characteristic to encode and decode image, such as fractal dimension[10] and other related topics[11-14], are our future work.

Acknowledgments. This research is supported by National Natural Science Foundation of China (Grant No.10671180), and Research Project (No.20060183) of Education Department of Zhejiang Province of China.

References

1. Barnsley MF. : Iterated Function Systems and the Global Construction of Fractals. Proc. Roy. Soc. London (1985) 243-275
2. A.E. Jacquin.: A Novel Fractal Block-coding Technique for Digital Image. Proceedings of ICASSP IEEE International Conference on ASSP, Vol.4 (1990) 34-42
3. A.E. Jacquin.: Fractal Image Coding a Review. Proceeding of the IEEE, Vol.81 (1993) 212-223
4. Mingshui Li, Shanhu Ou and Heng Zhang.: The New Progress in Research Approach of Fractal Image Compression. Journal of Engineering Graphics, Vol. 4 (2004) 143-152
5. Y.Fisher.: Fractal Image Compression. Fractals, Vol.2 (1994) 25-36
6. Y.Fisher.: Fractal Image Compression: Theory and Application. Springer-Verlag, Berlin Heidelberg New York (1995)
7. Shouji Chen, Liming Zhang: Fractal And Image Compression. Shanghai Science And Technology Education Publishing House(1998)
8. Chunling Chun, Shaodi Wang and Bingzhen Shan.: A Fractal Image Coding Based on the Quadtree. Proceeding of ICSP'98. (1998) 783-786
9. Jinglei Guo.: Grey Image Compression Based on Fractal Theory and Genetic Algorithm. Wu Han Science and Engineering University Master Paper(2002)
10. L. F. Xi.: Hausdorff Dimension of Lever Set of Rademacher Series. C. R. Acad. Sci. Paris Serie I 331 (2000) 29-33
11. Mei Sun, Lixin Tian, Jian Yin.: Hopf Bifurcation Analysis of the Energy Resource Chaotic System. International Journa of Nonlinear Science, Vol.2 (2006) 49-53
12. Guoliang Cai, Juanjuan Huang, Lixin Tian, Qingchao Wang.: Adaptive Control and Slow Manifold Analysis of a New Chaotic System. International Journa of Nonlinear Science. Vol.2(2006) 50-60
13. Xuedi Wang, Lixin Tian, Liqin Yu.: Linear Feedback Controlling and Synchronization of the Chen's Chaotic System. International Journa of Nonlinear Science, Vol.2 (2006) 43-49
14. Meifeng Dai, Xi Liu.: Lipschitz Equivalence Between Two Sierpinski Gasketse. International Journal of Nonlinear Science, Vol.2 (2006) 77-82

Towards the Application of Distributed Database in University MIS

Xiaoyong Wang and Yaofeng Fang

Computer Science Department, Junior College of Zhejiang Wanli University,
Ningbo, 315101, R.P. China

Abstract. The paper explores some crucial technologies of Distributed Database in University MIS, such as the distribution and replication of data, and studies its feasibility on the basis of campus networks.

Keywords: Distributed database, University MIS, Data distribution, Coherence and parallel of data, Data copy.

1 Introduction

Distributed Database is a set of data that belong to a complete unit in logical sense but are distributed to various nodes in physical sense, distinct from the traditional one which we call Centralized Database. Compared with Centralized Database, Distributed Database is characterized by distribution and logic coordination: the former refers to the arrangement that the data are distributed to some data subsets stored at various nodes with overall consideration, instead of being stored in a particular computer; the latter refers to the arrangement that the data subsets at various nodes are regulated separately by strict rules but belong to an integral unit by logic [1].

Therefore, the distribution of Distributed Database makes it different from Centralized Database, its logic coordination makes it distinct from Decentralized Database through network connections and it is superior to the Decentralized Database concerning the data's independence. Nevertheless, Centralized Database is the basis of Distributed Database and the network provides the indispensable environment for Distributed Database [2].

The University Management involves management of various aspects, the students management, the teaching management, the scientific research management, the management of human resources, the assets management, to name a few. The current situation in many universities where computers are utilized in their management is that each administrative division explores its own management system such as the students grading management system explored by teaching administration, the personnel management system by the personnel division, the library management system by library and so on, which operates mainly on the basis of the personal computer (PC) or the local area network (LAN), and the resource share and information exchanges visits can not be achieved among different divisions. As the modernization of university management has been advancing, the traditional mode of management, while abusing many resources, is inadequate to meet the requirement of open information management, resource share and information exchanges visits [3].

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 513–519, 2007.
springerlink.com

In the past few years, the campus network has been utilized as the basic facilities of university administration, whose application has greatly promoted the modernization of education and provided the solid foundation for constructing the information college as well as a new network platform for the exploration of university MIS with Distributed Database as its core technology.

2 Some Crucial Technologies Concerning the Application of Distributed Database in the University MIS

Next, Some crucial technologies concerning the application of Distributed Database in the University MIS and the realization of these technologies are to be discussed as follows, supposing the operating system is WindowsNT and the network relation type database is SQL Server of Microsoft.

2.1 Drawing Up the Scheme of Data Distribution

Like Centralized Database, Distributed Database consists of two parts: the integral of all necessary applicable data, known as physical database, which is the principle part of the Distributed Database, and the definition of data structure, the distribution of the overall data and its description, known as descriptive database. The data of Distributed Database can be divided into the partial data provided for the local application of a particular node, and the overall data that are involved in the overall application and can be visited by many nodes as they are stored in various nodes physically. The data distribution of Distributed Database are not stored in a particular node, but partitioned into some logical segments according to different requirement, which are distributed to the various nodes through some strategies. As the data distribution affects the capability, the reliability and efficiency of the whole system, it is of extreme importance to decide on the proper strategy of data distribution [4] [5].

2.1.1 Models of Data Distribution

The data distribution of Distributed Database can be roughly divided into four types, the centralized model, the intersected model, the copied model and the composite model. The characteristics and demerits of each model are illustrated in table 1.

2.1.2 Elaboration on the Project of University Data Distribution

The flow of campus data information nodes are time dependent as the information exchange visits mainly focus on some periods of each semester and a large amount of information consultation are usually limited within each division. In accordance with this practical situation, we can employ the distribution strategy with slight variation on the basis of the composite model; namely, all data are divided into a number of subsets according to the operation requirement of each division, each of which is stored in a particular database server of the LAN where different divisions are located. The copes of all subsets for resource share are stored in the central database server of the campus network, thus decreasing the complication of data visits of inter-divisions. The concrete solution is that the database server of each division works as a network node of LAN, with SQL Server7.0 installed, in which the local data is stored;

the interactive visits among different divisions are made through the visit to the data copy stored in the central databank server, and the coherence is maintained through distributed transaction or copy. Besides, the central database server needs to register to local servers of various divisions so as to make the visit of each division to the central server available. If this strategy is taken, on one hand, the load of local servers can be relieved and the interactive visits among different divisions become less complicated; on the other hand, the whole system can still function well in case of the malfunction in a certain division, thus the reliability of the system is improved, since there is copy in the central database server for each local database, the prompt recovery of data through the copy of SQL Server is possible if the malfunction at a particular node leads to the destruction of data.

Table 1. Comparison among various data distribution models

type	characteristics	demerits
The centralized model	As all data segments are stored in the same node, the easy operation and management of data as well as the coherence and completion of data can be guaranteed.	The reliability is not high since the node where the data are stored is susceptible to overburden and occurrence of bottleneck, leading to the collapse of the whole system in case of malfunction
The intersected model	All data are integrated and intersected to be some logic segments, each of which is allocated to a particular node. Therefore, the memory equipment of each node can be fully exploited and the memory space becomes larger	There is difficulty in maintaining the coherence and completion of data since data are distributed to various nodes
The copied model	As the overall data have many copies and there is a complete data copy at each node, the reliability of system is higher, the response is faster and the recovery of databank is easier.	It costs more to maintain the simultaneous data revision of all nodes; there is more redundancy of data, and the memory space is affected.
The composite model	All data are divided into a number of subsets. Each one is stored in a particular node and none of the nodes store all data, thus combining the merits of both intersected model and copied model, and resulting in the more efficient operation.	As it is the combination of the intersected model and copied model, it inherits the demerits of both models.

The blue print of data distribution that is feasible in accordance with the characteristics of university Distributed Database is illustrated in the following Fig. 1.

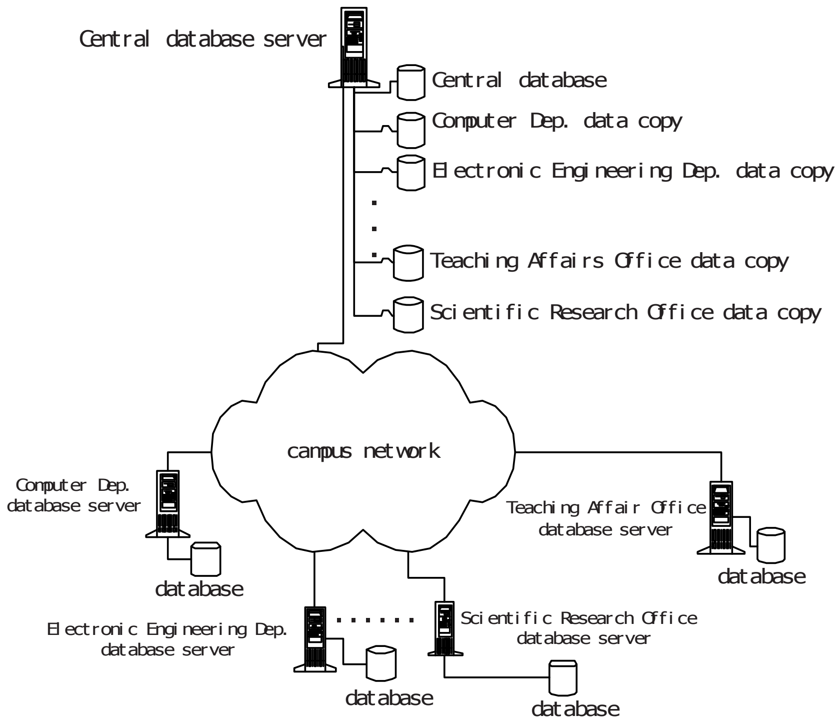


Fig. 1. The blue print of the Distributed Database applicable in university MIS

2.2 Coherence And Parallel of Data

In the MIS that employs Distributed Database, transaction is an array of distributed processing and the data to be processed are likely to be distributed to various nodes. As the distribution of data is transparent to users who often make overall demands on data likely to be stored in any of the nodes comprising the Distributed Database, problems tend to arise when quite a few transactions are operated in parallel, despite the improving efficiency of the system. There are three problems brought about by transaction parallel operating: loss of updating, incoherence of data processing, and dependence on the unsubmitted updating. In case the problems can not be properly solved, the completion and coherence of the data are ruined. In order to maintain the coherence and parallel of data, lock scheme contained in the MIS of SQL SERVER can be applied to close off the data being revised by a transaction in case other users should visit the incoherent data.

2.3 The Scheme of Data Copy

Another problem that should not be overlooked in the date distribution to various nodes and the processing of distributed data, is how to maintain the data synchronization of

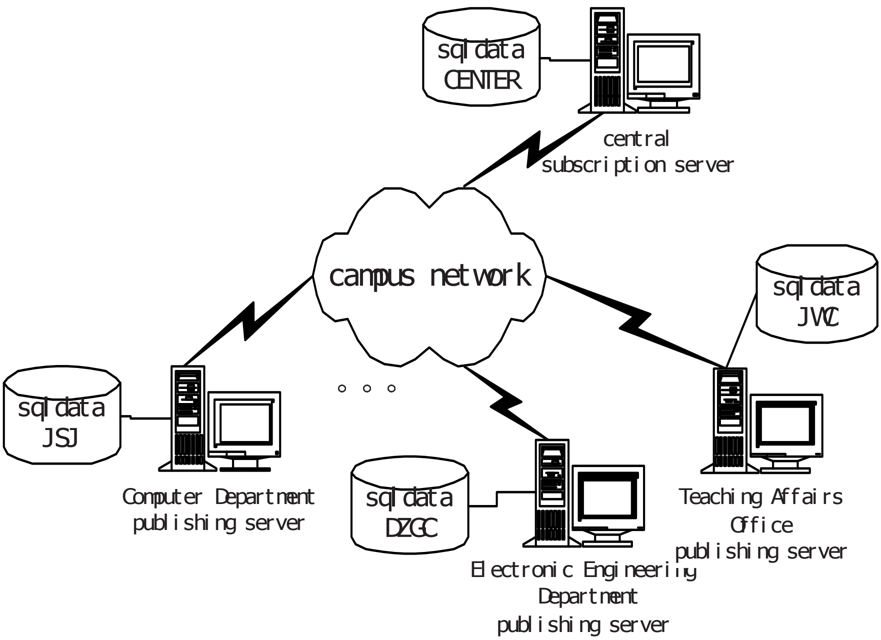


Fig. 2. The blue print of the publishing/ subscription model in university distributed MIS

various databanks. For instance, suppose the courses information of a specialty in the computer department has been revised, but the revision has not been reported to the master database server, consequently, when another division pays visit to the master database server, it only get the outdated courses information, which is inconsistent with the revised version owing to the failure of the data synchronization between local databases and the master one. The reprography service provided by SQL Server can ensure the replication of information among different databank servers, capable of acknowledging revision and sending it to the other systems including long-distance systems.

The reprography model of SQL is made up of three physical layers – subscription layer, publishing layer and distribution layer, which are closely linked with each other to constitute a functional replication model on the principle of "you publish and I subscribe" that are feasibly observed and applied. A brief introduction to these three layers is as follows.

1. Publishing layer provides information. When a system is designed to provide information to other systems through the reprography, it becomes a publisher and the database server where it is located is the publishing server, providing information for other systems and that information can be regarded as a publishing with all entries, ranging from the whole content of the databank to a record or a search result.
2. Subscription layer receives information. The users should turn their attention to subscribers and have them aware where to find received information after the information entries are published. Subscribers refer to the information receivers who need to install the connection with the distribution servers.

3. Distribution server with its databank working as a connector between the publishing server and the subscription server, is the resource of information. The users, upon installation, can choose the local servers or the long-distance severs as their distribution servers, where all information related to the publishers and subscribers are stationed.

Above all, the publishing server and distribution server should be installed in order to realize the replication of Distributed Database of university MIS. The two servers are to be installed in different computers when there are a lot of subscriptions; otherwise, it's more proper to put them together (see Fig. 2).

2.4 Technology of Database Accessing of MIS

Next, the technology of Database Accessing of MIS is to be discussed with an example of the MIS subsystem applied in the Computer Department, which employs PowerBuilder6.5 as the client's developing tools on the basis of the distributed MIS of the campus network. Making use of transaction objects to make visit to the database, the clients program of PowerBuilder is directly connected with the sub-databases through the specific interface. A detailed description of how users of the MIS subsystems connect with and make visit to various databases upon registration is given in the following script.

```
//to read from the init files the parameters connected
to the databank and to connect with the database
through the specific interface.

disconnect;

SQLCA.DBMS=ProfileString(“.\jsjprofile.ini”,"Database",
"DBMS", " ")

SQLCA.Database=ProfileString(“.\jsjprofile.ini”,"Databa
se","DataBase", " ")

SQLCA.LogID=ProfileString(“.\jsjprofile.ini”,"Database"
,"LogID", " ")

SQLCA.LogPass=ProfileString(“.\jsjprofile.ini”,"Databas
e","LogPassword"," ")

SQLCA.ServerName=ProfileString(“.\jsjprofile.ini”,"Data
base","ServerName", " ")

SQLCA.UserID=ProfileString(“.\jsjprofile.ini”,"Database
","UserID"," ")

SQLCA.DBPass=ProfileString(“.\jsjprofile.ini”,"Database
","DatabasePassword", " ")

SQLCA.DBParm=ProfileString(“.\jsjprofile.ini”,"Database
","DbParm"," ")

SQLCA.AutoCommit=true

CONNECT;//connected to the above-mentioned parameters

......
```

JSJprofile.ini, an init file in the script, is a structured text similar to INI of windows and made up of many sections each of which consists of some variable Assignment Statements. Through Profile String parameters provided by PowerBuilder, each attribute value of SQLCA can be retrieved from the structured text.

3 Conclusions

It has been a big game to explore the large scale MIS on the basis of Distributed Database operating in the campus network. This mode of database application system that employs the distributed computing and Distributed Database technology ensures more direct, prompt and convenient data management in various divisions that are widely separated from each other. In addition, the MIS on the basis of Distributed Database is easy for system expansion and it is a new task to take the campus network as its operation environment. However, owing to the complication of the distributed environment, a great many difficulties will occur in the process of exploration concerning the maintenance of the coherence, completion, parallel and security of data in Distributed Database. The paper has probed into some crucial technologies and the feasibility of distributed database in University MIS. With the development of Distributed Database technology and the campus network, the explorations in this respect will surely flourish in near future.

References

1. F. Bennour, Un Gestionnaire de Structures Distribuées et Scalables pour les multiordinateurs Windows: Fragmentation par Hachage, PhD thesis in French, Paris Dauphine University, 2000.
2. M. Blaum, J. Brady, J. Bruck & J. Menon, EVENODD: An Optimal Scheme for Tolerating Double Disk Failures in RAID Architectures, IEEE 1994.
3. J. Blomer, M. Kalfane, R. Karp, M. Karpinski, M. Luby & D.Zuckerman, An XOR-Based Erasure-Resilient Coding Scheme, ICSI Tech. Rep. TR-95-048, 1995.
4. L. Hellerstein, G.A. Gibson, R.M. Karp, R.H. Katz & D.A. Patterson, Coding Techniques for handling Failures in Large Disk Arrays, Algorithmica, 1994, 12, pp.182-208.
5. M. Ljungström, Implementing LH*RS: a Scalable Distributed Highly-Available Data Structure, Master Thesis, Feb. 2000, CS Dept., U. Linkoping, Suede.

Fractal Interpolation Fitness Based on BOX Dimension's Pretreatment

Qin Wang[1], Min Jin[1], Lifeng Xi[1], and Zhaoling Meng[2]

[1] School of Computer Science and Information Technology,
Zhejiang Wanli University, Ningbo 315100, P.R. China
[2] Mathematics Unit, Beijing Institute of graphic Communication,
Beijing, 102600 P.R. China
{Qin WANG, Min JIN, Lifeng XI, Zhaoling MENG, qinwang@126.com}

Abstract. For graphs of various local complex degrees, this paper will investigate their fitting approach and conduct experiments by using the mixture processing method which is a combination of the Box dimension's pretreatment with self-affine fractal interpolation function (AFIF).

Keywords: Fractal, fitness, interpolation, AFIF.

1 Introduction

As a new tool in data fitness and interpolation, fractals are always self-similar or self-affine, which means the fractals' local complexity is same as the whole. However, this property sometimes is a restriction of further application in the data fitness.

This paper concerns a special class of fractals, AFIF. AFIF can simulate not only the graph of smooth function, but also can effectively and accurately fit rough curves and vibrating data, such as mountain range outlines, electrocardiograms ...etc [3]. It is a new interpolation tool after polynomials and splines. As for the general theory of fractal interpolation function and affine fractal interpolation function, the reader is referred to [1-6].

Based on above discussion, AFIF have same fractal dimension or same complex degree at each location. However, in the practical application, the graphic complex degrees and the sensitivities may be absolutely different when data respond to time in the disparate time periods. Thus, it is obvious that the fitting may not be effective if we directly use AFIF, which has the same complexity everywhere.

Due to the defect of using only one fractal interpolation function in the data fitting, in this paper we will use the mixture processing method, which is a combination of Box dimension's pretreatment with AFIF, to conduct experimentation and analysis. Clustering analysis is adopted according to the each sub-graph's box dimension. Then the sub-graphs are reconstructed together according to adjacent Box dimension, and fractal interpolation is adopted separately. Finally, it is readjusted and resumed.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 520–526, 2007.
springerlink.com

2 Fractal and Interpolation

2.1 Fractal Dimension

Definition 1. Let E be a compact subset of $\mathbf{R}^2$. Given $\delta > 0$, let $N_\delta(E)$ be the smallest cardinality of family of solid squares with side length δ such that the union of these squares covers E.

If the limit $\lim_{\delta\to 0} \frac{\log N_\delta(E)}{-\log \delta}$ exists, then

$$\dim_B E = \lim_{\delta\to 0} \frac{\log N_\delta(E)}{-\log \delta} \tag{1}$$

is called the Box dimension of E.

Example 1. Smooth curve
The graph of any smooth function has Box dimension 1. This result means that polynomials and splines are too smooth to approximate the graph with high complexity.

Example 2. C×C

Let $C = \{\sum_{i=1}^{\infty} a_i 3^{-i} : a_i = 0 \text{ or } 2 \text{ for each } i\}$, which is called standard Cantor set. Then $C \times C$ can be covered by 4^n square of side 3^{-n}, and we can check that $N_{3^{-n}}(C\times C) = 4^n$.
Therefore,

$$\dim_B (C\times C) = \lim_{n\to\infty} \frac{\log N_{3^{-n}}(E)}{-\log 3^{-n}} = \frac{\log 4}{\log 3}.$$

Example 3. Graph of continuous function
For the graph E of a continuous function f, we let $\delta = 2^{-n}$, and

$$M_\delta(E) = \sum_{k=1}^{2^n} [2^n \omega(f, [\frac{k-1}{2^n}, \frac{k}{2^n}])],$$

where $\omega(f,[a,b]) = \max_{x_1\in[a,b]} f(x_1) - \max_{x_1\in[a,b]} f(x_2)$ is the oscillation of f restricted on the interval [a,b], and [x] denotes the smallest integer greater than or equal to x, for example, [3]= [2.5]=3. Then there is a constant $C > 1$ such that $C^{-1}M_\delta \le N_\delta \le CM_\delta$, and thus

$$\dim_B E = \lim_{\delta\to 0} \frac{\log M_\delta(E)}{-\log \delta} \tag{2}$$

It is easy to show that for any *smooth* function, we have $D^{-1}\delta \le M_\delta \le D\delta$ for any δ, where $D>1$ is a constant, therefore the graph has Box dimension 1.

2.2 Affine Fractal Interpolation Function

Given points $\{(x_i, y_i)\}_{i=0}^{N}$ in the plane, we suppose $\{\omega_1, \omega_2, \ldots, \omega_N\}$ is an *iterated function system* satisfying

$$\omega_i\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_i & 0 \\ c_i & d_i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} g_i \\ h_i \end{pmatrix} \tag{3}$$

$$\omega_i\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix},\ \omega_i\begin{pmatrix} x_N \\ y_N \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \tag{4}$$

where $|d_i|<1$ and $a_i \in (0,1)$ for any i with $1 \le i \le N$. By (3) and (4), we notice that $\{\omega_i\}_{i=1}^{N}$ is determined by $\{(x_i, y_i)\}_{i=0}^{N}$ and. We always call $\{d_i\}_{i=1}^{N}$ vertical factors and $\{(x_i, y_i)\}_{i=0}^{N}$ interpolation points respectively.

Definition 2. Suppose $f(x)$ is a continuous function on the interval $[x_0, x_N]$. Let

$$\Gamma = \{(x, f(x)) : x \in [x_0, x_N])\}$$

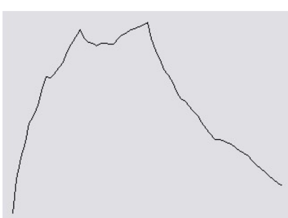
be the graph of $f(x)$. We say that $f(x)$ is an **affine fractal interpolation function**, if

$$\Gamma = \bigcup_{i=1}^{N} \omega_i(\Gamma) \tag{5}$$

Example 4

For AFIF defined by (3)-(5), the dimension $\dim_B \Gamma$ of the graph Γ satisfies the following dimension formula ([1]):

$$\sum_{i=1}^{N} |d_i| \cdot |a_i|^{\dim_B \Gamma - 1} = 1 \tag{6}$$



Interpolation points	(0,0.1),(0.5,0.8),(1,0.2)
Vertical factors	d_1=0.5, d_2= - 0.2

Fig. 1. An Example of AFIF

Remark 1: Formula (6) holds when the interpolation points do not lie in a line simultaneously and $\sum_{i=1}^{N} |d_i| > 1$. Any connected part of the graph Γ of AFIF has the same dimension $\dim_B \Gamma$.

3 Algorithm

- **Step 1:** We divide the interval into several subintervals and thus obtain some sub-graph.
- **Step 2:** By using formula (2) to estimate the dimension of each sub-graph.
- **Step 3:** Clustering the sub-graphs according to their dimensions, we reconstruct some new graph $\Gamma_1, \Gamma_2, \cdots \Gamma_k$, each of which is composed of some sub-graphs with adjacent values of Box dimension. In the process of reconstruction, we should make translation for each sub-graph along y-axis to ensure the *connectedness* of graph.
- **Step 4:** For each new graph Γ_i in Step 3, we use AFIF to approximate it.
- **Step 5:** Reconstruct these AFIFs to obtain an approximation of the original graph.

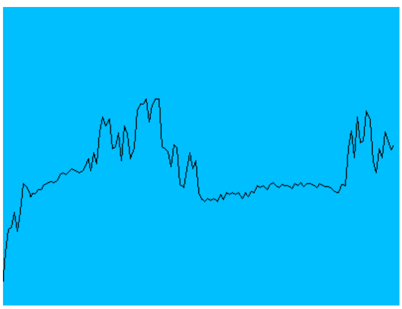


Fig. 2. Original graph

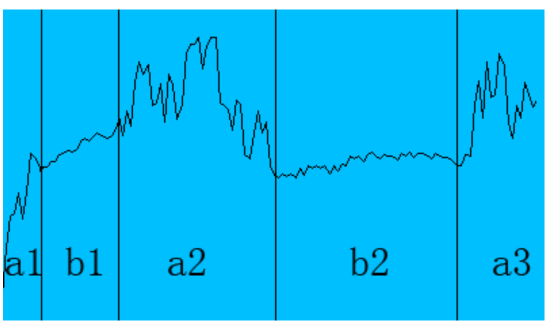


Fig. 3. Partition

4 Experimental Results

The graph is a piece of the tendency picture of captial stock certificate. We use it as an example for fractal interpolation.

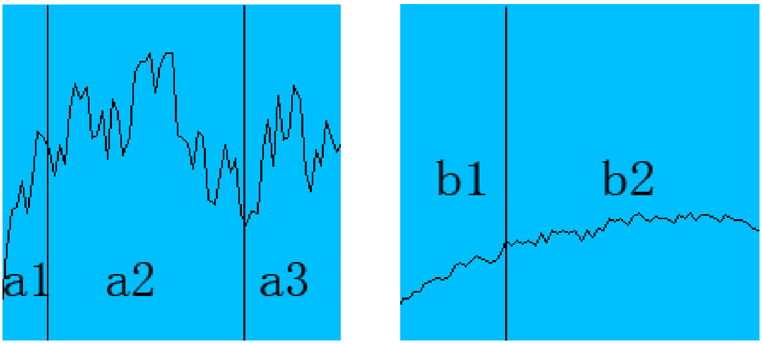
Here we give a partition of the original graph to obtain 5 parts: a1, b1, a2, b2 and a3. By formula (2), we get

$\dim_B$(Part a1) ≈0.37, $\dim_B$ (Part a2)≈0.38, $\dim_B$ (Part a3)≈0.38

and

$\dim_B$ (Part b1)≈0.14 , $\dim_B$ (Part b2)≈0.13.

And thus we reconstruct the blow graphs.



graph (a) graph (b)

Fig. 4. Reconstruction of graphs according to Box dimension

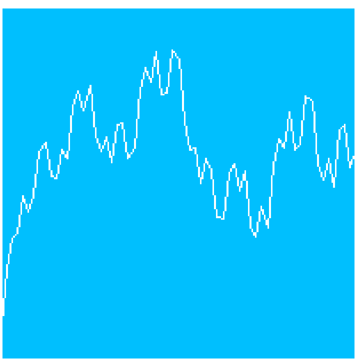


Fig. 5. AFIF (a') with respect to graph (a)

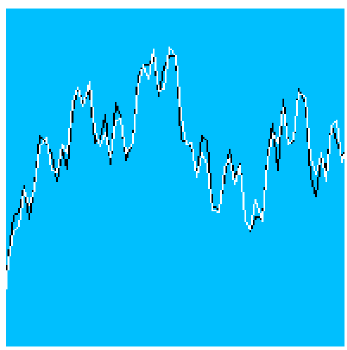


Fig. 6. AFIF (a') and graph (a)

Interpola-tion points	（0,0.1），（0.5,0.85），（1,0.6）
Vertical factors	d_1=0.6, d_2=-0.7

Fig. 7. AFIF (b') with respect to graph (b)

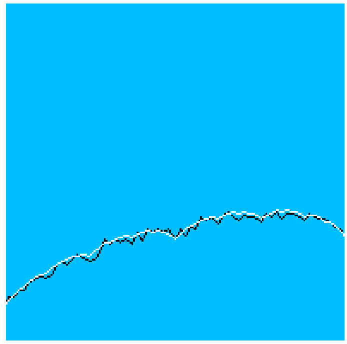


Fig. 8. AFIF (b') and graph (b)

Interpolation points	(0,0.1) (0.5,0.3) (1,0.2)
Vertical factors	d_1=0.5, d_2=0.6

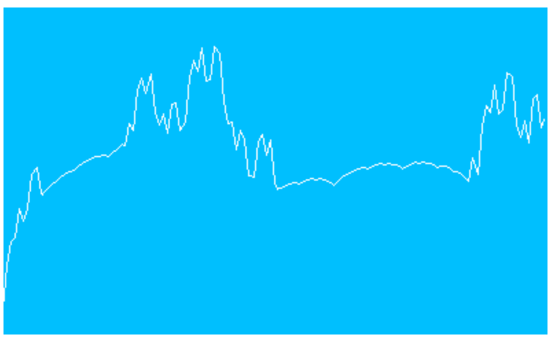


Fig. 9. Reconstruction of AFIF (a') and (b')

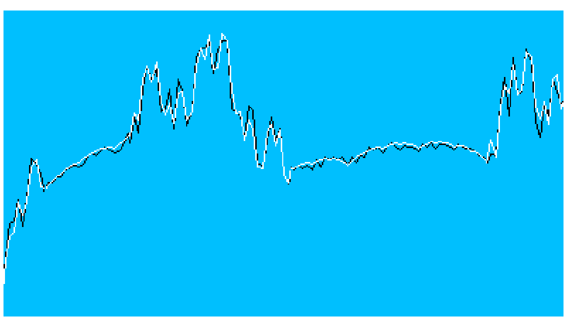


Fig. 10. Original graph (black) and graph of AFIFs with reconstruction (white)

References

1. Barnsley, M. F., Fractal functions and interpolation. Constr. Approx., 1986, 2 , 303–329.
2. Barnsley, M. F., Elton, J. H. and Hardin, D. P., Recurrent iterated function systems. Constr. Approx., 1989, 5, 3–31.
3. Barnsley, M. F., Elton, J., Hardin, D. and Massopust, P., Hidden variable fractal interpolation functions. SIAM J. Math. Anal., 1989, 20, 1218–1242.
4. Barnsley, M. F. and Harrington, A. N., The calculus of fractal interpolation functions. J. Approx. Theory, 1989, 57, 14–34.
5. Dalla L. and Drakopoulos V., On the parameter identification problem in the plane and the polar fractal interpolation functions. J. Approx. Theory, 1999, 101, 289–202
6. Mazel D. S. and Hayes M. H., Using iterated function system to model discrete sequences. IEEE Trans. on Signal, 1992, 40(7), 1724–1734.

A Statistical Spam Filtering Scheme Based on Grid Platform*

Peng Liu[1], Jian-she Dong[2], and Wei Zhao[3]

[1] Grid Research Center, PLA University of Science and Technology, Nanjing 21007, China
[2] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
[3] Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China
liupeng00@mail.tsinghua.edu.org

Abstract. Spam is in spate, which accounts for over 60 percent of all emails in the world recently. Researchers are trying to develop ways to fight it but few are effective. The paper put forward a new filtering scheme based on grid technology and statistical method, which regards the user computers and email servers as nodes of the grid. They contribute and consume statistics information on the grid platform. If the number of copies of an email is obviously err from normal value, to flag it as a spam then can be a reasonable operation. As more and more nodes join the platform, the filtering precision can be further improved, just as the simulation study shows.

Keywords: spam, grid, CNGrid, statistics.

1 Introduction

Since the Internet and its application acquire rapid development, a series of problems related to Internet have accompanied, some of which may lead to a lot of troubles, for example, the Spam. The flooding of Spam will result in a mass of network resources being wasted, and the normal email corresponding being affected.

The problem of spam email is apparent to any frequent email user: unwanted, unsolicited bulk messages are emailed to a large number of users indiscriminately, which is similar to bulk mails sending the traditional postal service. In September 2001, 8% of all emails in US were spam. By July 2002, this fraction had increased to 35% [1]. More recent studies report that, in North America, a business user received 10 spam emails on average per day in 2003, and that this number is expected to grow by a factor of four by 2008 [2]. Furthermore, AOL and MSN report a daily blocking of 2.4 billion spam emails from reaching their customers' inboxes. This traffic corresponds to about 80% of daily incoming emails at AOL [3]. It is reported by the Anti-spam center of ISC [**4**] that in China a user received 19.33 spam emails on average

* This work was supported by the National Science Foundation of China under the grant No. 60403043.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 527–534, 2007.
springerlink.com

per week and 63.97% of all emails were spam in Mar. 2006, this is 2.03 spam emails more than Oct. 2005.

Spammers conduct marketing, commercial, and even unethical activities by sending out a huge amount of spam. This high volume is required as it is the only way to receive enough economical benefit. There is therefore a heavy maldistribution on e-mail traffic, making document space density a good index to identify spam. Although ordinary users seldom send more than 1000 similar e-mails, spammers have to send the same spam far more than that. Note that some of the unethical spam mail are said to be difficult to judge even for a human. However, the existence of over thousand identical e-mails makes the fact clear. Actually, experimental results reported in Section 4 showed that simple threshold is enough to distinguish spam from other e-mails.

The evident difference between spam mail and normal mail is that the same spam mail will be delivered to a large number of users, but most of normal only have one single receiver. Based on this observation, this paper presents a counting method based on CNGrid for spam mail filtering, In order to avoid normal mail being classified as spam mail, we also use a white list (WL) to improve the precise of spam mail filter.

The rest of the paper is organized as follows: In Section 2, we present the related works of spam mail filtering. In Section 3, we present the proposed filtering system in detail. Then, in Section 4, we show the results of the experiment which reveals the effectiveness of the proposed anti-spam filter. Finally, section 5 presents conclusions.

2 Related Works

Over the past few years, different approaches have been presented to provide resistance against spammers. Some of them use a Bayesian-like approach, or a rule-based approach, and some use a cryptographic solution to protect against spamming problem.

The simplest and most intuitive of all technique used to curb spam was to keep a blacklist of addresses to be blocked, or a white list of addresses to be allowed are also used. However this technique is not proved to be successful – since the spammers started sending spam mails either without the senders address or by spoofing the sender address.

Somebody suggests the method which to increase spammer's cost, such as filters and fight back (FFB) [5], slow senders method and penny per mail method [6]. The working of FFB resembles DoS (Denial of Services) a little bit. It sends junk messages to the spammers to increase their working load. The slow senders method and penny per mail method require all email sender to carry out a calculation which will consume their work time or to pay a little fee for each email. These methods must be support by new protocol, so they are not easy to be popularized.

Another one kind method is to distinguish the email sender's identity. Such as the ePrivacy E-mail open the standard [7] and Questions-Answering filtering. The ePrivacy request all senders to declare taking part in "the no sending spam" alliance. Every declarer will gain the figure signature and insert the signature into each email header to insure the sender's identity. An email without the signature will be chucked. The Questions-Answering method requires that the mail sender to fill in a table at a

Web page, otherwise the email with not be granted to send out. These methods are effective, but increased burden for senders and are easy to lose the legal mail.

Compared with the above method, the filtering methods are recipient by more people. Cohen [8] suggesting that incoming mail can be categorized according to its contents based on automatic learning rules. Some sophisticated rule based methods have good performance in spam filtering. SpamAssassin [9] is a successful case to filter spam emails based on rules. But the higher false positive ratio is its greatest shortcoming for

The concept of Bayesian Junk Mail filters suggested by Sahami et al. [10] got popularity. The filter was based on naive Bayes classifier. This method achieved a relative high degree of precision, but the recall was slightly low. It means that study have been found that more spam mails were classified as normal incorrectly. It was also found that outright deletions of spam brought about relatively high costs.

The cryptographic solution to protect against spamming problem was presented by Ioannidis [11]. In that solution the email address was encoded with certain policies. There policies were encrypted using symmetric keys and generated the message authentication code. The drawback to this solution was quite lengthy mail address, which proved to be difficult to adopt in commercial solutions. Not all methods presented for spam classification are suitable for both desk top based and server-based mail classification. The spam classification at the desktop is often more customizable and accurate, but such solutions often need too much computing and analysis and they are not suitable for massive spam mail process. The server-based mail classification should consider more about performance and avoiding normal mail being classified as spam mail.

3 Architecture of the Anti-spam Grid Base on CNGrid

In this paper we chose Grid [12] technique as the basis of the anti-spam system. It based on the following considerations: (1) Spam is delivered globally, so we need a global infrastructure to gather information on spam. (2) As central control system may result in bottlenecks, a collaboration of distributed services will efficiently serve local users. (3) It is a most dynamic environment that all the servers, clients, and e-mails keep changing all the time, so we need to form a virtual organization which is adaptive to changes.

The architecture of Anti-Spam Grid with distributed statistic of figure signature and distributed Bayesian filter. We call this architecture as Anti-Spam Grid. The whole system is built of Grid server, Anti-Spam Grid mail server (ASG Server) and Anti-Spam Grid client (ASG Client).

The Grid Server is with responsibility for the ASG Servers' scheme, registration and detecting the ASG Servers' running state. It is also with responsibility for the client user's registration, granting safe certification to client users and assigning ASG Server for client users which will serve it.

The ASG Server storages the fingerprint corpus with the designated range and provides index server schemed by the Grid Server. At the same time, it provides the searching service of approximate email and sharing of Bayesian knowledge repository.

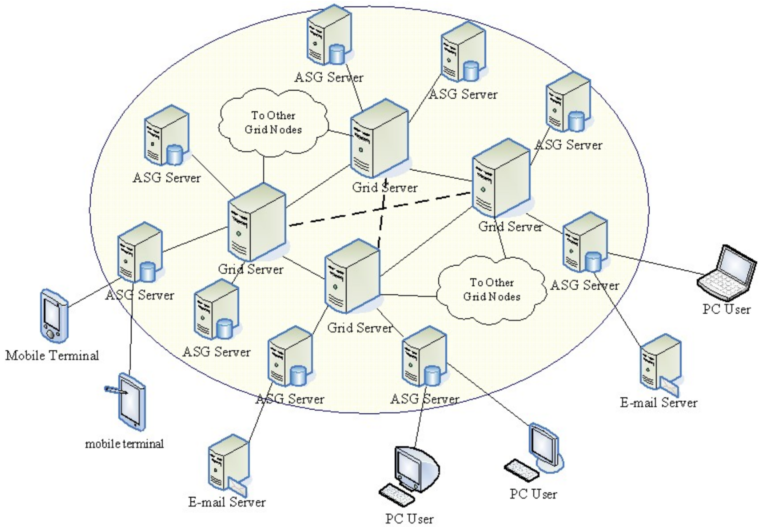


Fig. 1. The structure of Anti-Spam Grid

ASG Client includes the ultimate user of email service and the email servers that need service of spam filtering. When a new client takes part in the ASG, it suggests the request to a Grid Server. The Grid Server auditing the request and provides an authority certificate to the client and assigns an ASG Server who will serve the new client. ASG Client takes charge of the fingerprint abstracting. When a client send a fingerprint to ASG Server to query the amount of approximate email, the assigned ASG Server will responses the query and the new fingerprint will be stored in the global fingerprint repository. ASG Client also reports the local Bayesian knowledge to the assigned ASG Server.

The main work steps of Anti-Spam Grid are as followings:

- ASG Server publishes its service to someone Grid Server. Grid servers share their information each other.
- Once a ASG Client joined the ASG, it suggest request to a Grid Server.
- The Grid server assigns a ASG Server to serve it according to the rule of workload balance or serving nearby
- When the ASG Client suggesting connection request firstly to an ASG Server, the ASG Server should send the client's authority certificate to Grid Server to check it. If the certificate is qualified, it will be stored locally and the later checking will be implemented locally.
- ASG Client report email fingerprints and Bayesian knowledge to ASG Server.
- ASG Server returns the amount of approximate emails and other Bayesian knowledge to the client.

Since each ASG Server only stores the fingerprints of limited range, the query of approximate email may be processed in other ASG Server. The range of which fingerprint be stored in an ASG Server is assigned by a Grid Server. In order to increase the efficiency of approximate email query, every ASG Server maintains a fingerprint table. It only cost one hop to route the object ASG Server in approximate fingerprint searching.

The scalability of Anti-Spam Grid is perfect. Not only Grid Server but also ASG Server can join the system dynamically. When a new ASG Server join the system, it can burden apart of client and workload schemed by Grid Server.

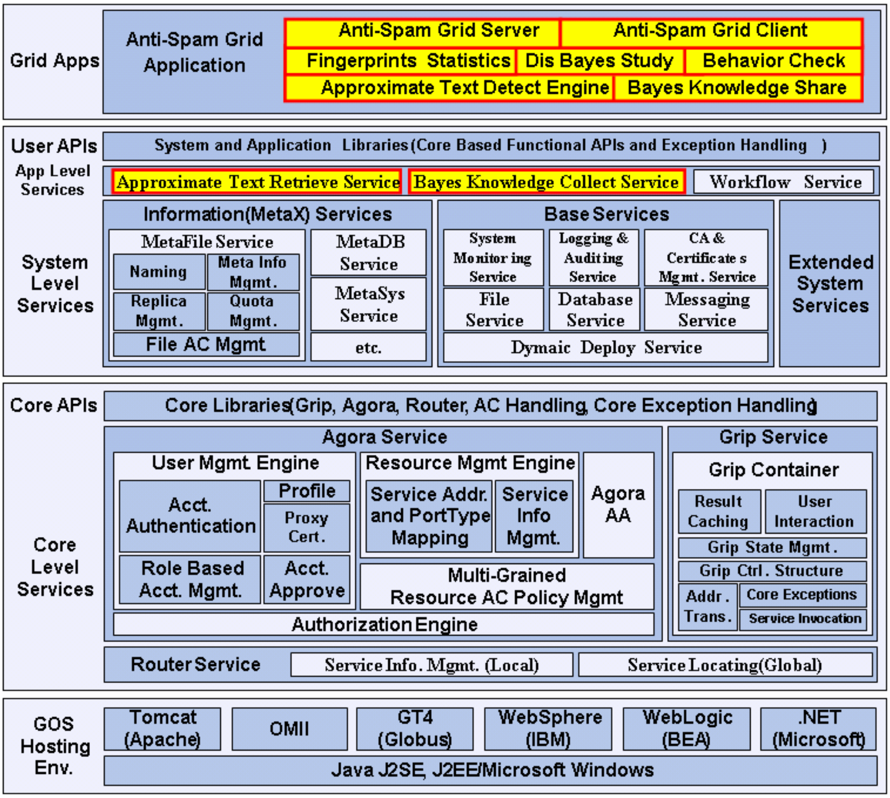


Fig. 2. The architecture of CNGrid-ASG

The proposed system is constructed on the CNGrid software version 2.0. The structure is shown in Fig. 2. We append the service of approximate text detection and the service of Bayesian knowledge collection and integrating, combined with the Information Service, System Monitoring Service, Logging Service, CA & Certificates Managements Service in the System Level and the User Management Engine in the Core Level, we realized the source management, user management and safe management.

Based on the CNGrid basis service and the extended service above, the system carried out Grid Server, ASG Server in Grid Application Lever. The system can filter spam in effect based on the integration of approximate email statistic technology, Bayesian knowledge sharing and real time communication protocol behavior analysis technology.

4 Simulation and the Result

There are many mature spam corpus for experiment of English spam filter such as Ling-Spam and PU corpus. In this paper we use a larger spam corpus, Genspam [13]. It includes 32332 spam and 9072 legit emails. In the spam corpus there are some spam are approximate. We used 90% of the corpus to train our Bayesian filter and use the other emails to simulate the process of email sending. In the simulate experiment, the amount of ASG users is from 10 to 100. We send a spam to *n* users, *n* is random from 1 to the user amount. We set the threshold as 10, when an email is receipted by users more than the threshold it will be judged as spam.

In classification tasks, two commonly used evaluation measures are accuracy (Acc) and error rate ($Err = 1 - Acc$):

$$Acc = \frac{n_{L\to L} + n_{S\to S}}{N_L + N_S}, \quad Err = \frac{n_{L\to S} + n_{S\to L}}{N_L + N_S} \tag{1}$$

N_L and N_S are the numbers of legit and spam emails to be classified. $n_{L\to L}$ is the number of spam emails that be classified as spam, and so on $n_{S\to S}$, $n_{L\to S}$, $n_{S\to L}$ can be deduced by analogy.

Accuracy and error rate assign equal weights to the two error types ($L \to S$ and $S \to L$). When selecting the threshold of the filter, but for users it is common believed that $L \to S$ is more costly than $S \to L$. To make accuracy and error rate sensitive to this cost, we adopt the cost-sensitive evaluation measures proposed by Androutsopoulos [14]: when a legitimate message is misclassified, this counts as λ errors; and when it is classified correctly, this counts as λ successes. This leads to weighted accuracy (WAcc) and weighted error rate (WErr =1−WAcc):

$$WAcc = \frac{\lambda \cdot n_{L\to L} + n_{S\to S}}{\lambda \cdot N_L + N_S}, \quad WErr = \frac{\lambda \cdot n_{L\to S} + n_{S\to L}}{\lambda \cdot N_L + N_S} \tag{2}$$

And the total cost ration (TCR):

$$TCR = \frac{WErr^b}{WErr} = \frac{N_S}{\lambda \cdot n_{L\to S} + N_{S\to L}} \tag{3}$$

In our experiment we set the λ as 9. The results are shown in Fig.3 and Fig. 4. For the Bayesian filter itself, the WAcc is about 91% and the TCR is about 2.5 and they are not change follow the user amount changing. For the ASG system, the WAcc and TCR are increasing with the in increase of users. In the experiment, the value of threshold is fixed. How to dynamic amend the threshold value is a problem should to research and experiment in our future work.

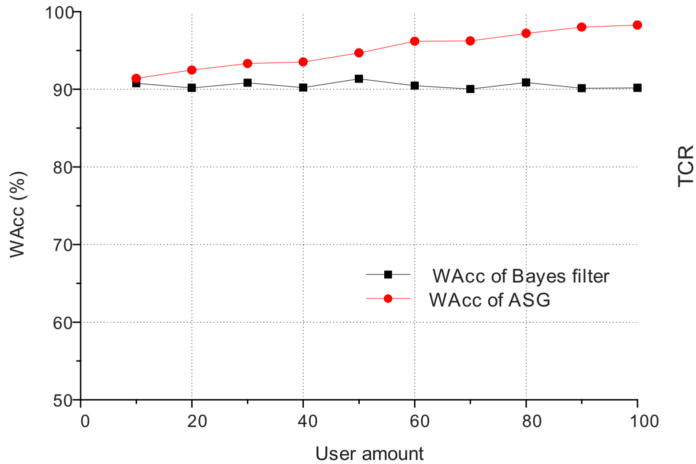


Fig. 3. WAcc of Bayesian filter and ASG

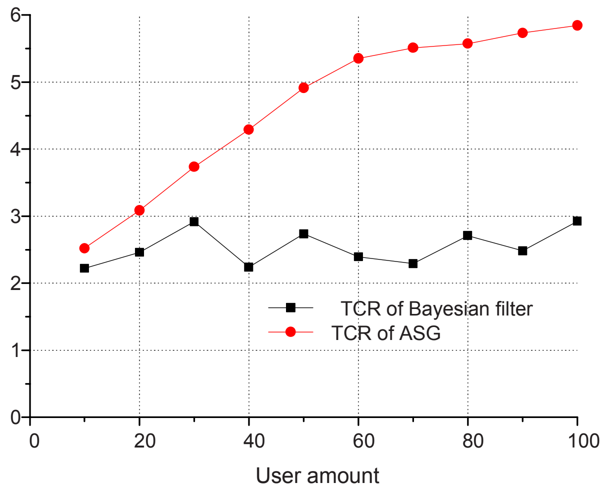


Fig. 4. TCR of Bayesian filter and ASG

5 Conclusions

We present design and evaluation issues of Anti-Spam Grid, an infrastructure dedicated to filter unsolicited bulk e-mails. Based on the CNGrid, the ASG users and servers can be dynamically added to the system. At the same time, the system can run properly whenever any ASG Server or Grid Server fails. So we can say the system reflects the core idea of the grid: virtual organizations. The result of experiment shows it is much more effective than contemporary anti-spam approaches.

References

1. Paul, N. C.and Monitor, C. S.: New strategies aimed at blocking spam e-mail. http://newsobserver.com/24hour-/ technology/story/655215p-4921708c.html.
2. Nelson, M.: Anti-spam for business and isps: Market size 2003-2008. Ferris research – analyzer information service report (2003)
3. Fallows, D.: Spam: How it is hurting e-mail and degrading life on the internet. Tech. Rep. 1100, PEW Internet & American Life Project (2003)
4. http://www.anti-spam.cn/ShowArticle.php?id=2713
5. Filters that Fight Back, http://www.paulgraham.com/ffb.html
6. Hird, S.: Technical Solutions for Controlling Spam. In proceedings of AUUG2002, Melbourne, September 2002.
7. ePrivacy Group, http://www.eprivacygroup.com/
8. Cohen, W. W.: Learning rules that classify e-mail.In: Proceeding of the AAAI Spring Symposium on Machine Learning in Information Access, 18-25, 1996
9. Spamassassin, http://au2.spamassassin.org/doc.html
10. Sahami, M., Dumais, S., Heckerman, D., and Horvitz, E.: A Bayesian Approach to Filtering Junk E-Mail. AAAI Tech Rep. WS-98-05 (1998)
11. Ioannidis, J.: Fighting Spam by Encapsulating Policy in Email Addressed. 10th Network and Distributed System Security Symposium (2003)
12. Foster, I., and Kesselman, K. (ed.): The Grid 2: Blueprint for a New Computing Infrastructure, 2nd edition, Morgan Kaufmann, Nov. 2003, ISBN: 1558609334
13. http://www.cl.cam.ac.uk/~bwm23/
14. Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., Paliouras, G., and Spyropoulos, C. D.: An Evaluation of Naïve Bayesian Anti-Spam Filtering. Proc. of Workshop on Machine Learning in the New Information Age, 11th European Conference on Machine Learning, Barcelona (2000)

An Approach to Web Prefetching Agent Based on Web Ontology with Hidden Markov Model

Xin Jin

School of Information, Central University of Finance & Economics, Beijing, 100081, P.R. China
d0229jin@mail.dhu.edu.cn

Abstract. With the rapid growth of web services on the Internet, users are experiencing access delays more often than ever. Recent studies showed that web prefetching could alleviate the WWW latency to a larger extent than the traditional caching. Web prefetching is one of the most popular strategies in web mining research domain, which are proposed for reducing the perceived access delay, improving the service quality of web site and mining the user requirement information in advance. In this paper, we introduce the features of the web site model named web ontology, and build a web prefetching agent-WebAGENT based on the web ontology and the hidden Markov model. With the agent, we analyze the user access path and how to mine the latent information requirement concepts, then we could make semantic-based prefetching decisions. Experimental results show that the web prefetching scheme of the WebAGENT has better predictive mining effect and prefetching precision.

1 Introduction

Due to the rapid development of Internet technique and the exponential growth of online information, Internet has become one of the most important information sources. However, owing to the limitation of the bandwidth, users always suffer from long delay time when they access web pages. In order to solving the problem, experts proposed a lot of solutions. Web prefetching is the most prominent one [1,2,3].

Web prefetching is an active caching scheme [4]. Compared with the normal passive caching scheme, web prefetching predicts the next request for web documents based on the current request of users through analyzing the server log data, fetches them in advance and loads into the server cache. It reduces the perceived access delay in some exent and improves the service quality of web server [5,6,7].

In this paper, we propose a novel web prefetching approach based on web ontology with the Hidden Markov Model(HMM). By utilizing the HMM, we analyze the user's browsing track, capture the user's actual information requirement intention, and make semantic-based web prefetching decisions.

Remainder of the paper is organized as follows. Section 2 describes the details of our prefetching tasks and algorithms of the web fetching agent named WebAGENT. Section 3 describes the experiments designed and evaluates the performance of the scheme based on our web prefetching agent. Section 4 provides a summary of this work.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 535–543, 2007.
springerlink.com

2 Web Prefetching

In general, a user always accesses the web site with certain intention. Driven by the intention, the user follows a link on a page that he is currently visiting and browses continually until that he or she satisfied. In other words, access path contains the user's certain requirement intention. If we extract the latent information requirement concepts from the user access path, we can make accurate decisions for web prefetching based on it. For example, let us consider that a given user has accessed the following pages Pa, Pb and Pc, according to the following sequence Pa→Pb→Pc . Page Pd, Pe and Pf are cited by page Pc. If we find that the concept c is latent intention that the access path implies and page Pe contains the concept c, we can pre-fetch page Pe for the user's next request.

Obviously, mining information requirement concepts that access path implies is crucial to web prefetching. Our approach is based on the following observations:

Observation 1: The author of web page always uses hyperlink to implement the organization of server host Hyperlink establishes the relation between web pages in certain concepts.

Observation 2: Anchor text of hyperlink can provide users with sufficient information. It generalizes the major content of linked page.

The conceptualization architecture of the web site, organized by hyperlink, can always be considered as a web ontology, which can help us to describe and analyze the web site structure. Based on the two observations above, we utilized the web ontology and the HMM to implement the concept mining. The web pages that access path includes can be denoted as the states of the HMM, the concepts that web pages contain correspond to the observer symbols, respectfully. Therefore, the observation symbol sequence of the HMM is the concept sequence actually. In fact, the probability of concept sequence about a certain concept is the possibility of the latent requirement concept implied by user access path. According to the probability calculated, we can choose some concepts as the latent requirement concepts, and evaluate the web pages that the current page cited. The web pages that satisfied these concepts as more are picked out as prefetching objects.

2.1 Model Description

2.1.1 Web Ontology

Proposition: Consider web ontology as conceptualization architecture of the web site, which is a tuple structure<D, R> where D is a domain, and R is a set of relevant relations on D.

For a web site ontology, the domain D can be considered as the set of all existing web pages, denoted as $D=\{p| p\in D\}$, and R is the set of hyperlinks in all web pages to link the different pages, denoted as $R=\{l| l\in R\}$.

- *Definition 1*: The web page p can be denoted as p =(P_ID, P_Url), where P_ID is the unique ID of the page, P_Url is the URL of the page.
- *Definition 2*: The hyperlink can be defined as a quadruple l=(L_ID,Anchor_Text,SP_ID,EP_ID), where L_ID is the unique ID of the hyperlink,

Anchor_Text is the anchor text around the hyperlink, SP_ID is ID of the web page that contains the hyperlink , EP_ID is ID of the page which the hyperlink points to.

- *Definition 3*: User access path E is a request sequence on domain D, defined as $E=<p_1,p_2,\ldots p_i,\ldots,p_n>$, where pi is the web page that the user requests for in the i^{th} step , p_n is the current page, n is the length of access path. The candidate pages for web prefetching are all web pages that the current page cited.
- *Definition 4*: Session S is a request sequence of the user in a certain time interval, defined as $S=<s_1,s_2,\ldots,s_n>$, where each $s_i \in S$ is the web page the user accesses in the ith step of the session, n is the length of the session. The page b follows page a, denoted as a→b, iff there exists s_i =a and s_i+1 =b($1\le i<n$).
- *Definition 5*: C is the Concept set of user's information requirement, denoted as $C=\{c_1,c_2,\ldots,c_m\}$, which is fetched from the web ontology.

In order to fetch the concepts of user's requirement in the web ontology, we use the HMM to analyze the user access path E from the domain D. The HMM is well known and widely used statistical method of characterizing the spectral properties of the frames of a pattern, which was proposed firstly by Baum and his colleagues in the late 1960s .

2.1.2 Describe the HMM

- Set the HMM to be $\lambda = (A, B,\pi)$, [7]
- N, the number of states in the model.
- Described the individual states as {1,2,…,N}
- Denoted q_t as the state at time t.
- M, the number of distinct observation symbols per state.
- Denoted the individual symbols as
  $V=\{v_1,v_2,\ldots,v_M\}$.
- $A=\{a_{ij}\}$, The state transition probability distribution, where

$$a_{ij} = P(q_{t+1} = j \mid q_t = i), \qquad i,j\le N \tag{1}$$

- $B=\{b_j(k)\}$, The observation symbol probability distribution, in which

$$b_j(k) = P(v_k(t) \mid q_t = j), \quad 1\le k\le M \tag{2}$$

defines the symbol probility distribution in state j, j =1,2,…,N.

- The initial state distribution $\pi=\{\pi_i\}$, in which

$$\pi_i = P(q_1(t) = i), \; 1\le i\le N,\ t=0 \tag{3}$$

In order to make the HMM model suitable to web prefetching from the web ontology, we let the states of the HMM corresponding to the web pages of user access path *E* respectively, and the observation symbols correspond to the concepts of the Set C respectively. That is, the state q_i corresponds to the web page p_i and the symbol v_i corresponds to the concept c_i.

2.2 Web Prefetching Agent - WebAGENT

Based on the web ontology with HMM, we build a web prefetching agent named WebAGENT, whose tasks include four-phase processing as the Fig.1: preprocess the server log data, extract the requirement concepts, analysis user access path and fetch the web pages. Before fetching web page, we will analyze the user access path, mine the latent information requirement concepts, then, based on it, we make predictive prefetching decisions. We will give a detailed introduction as follows.

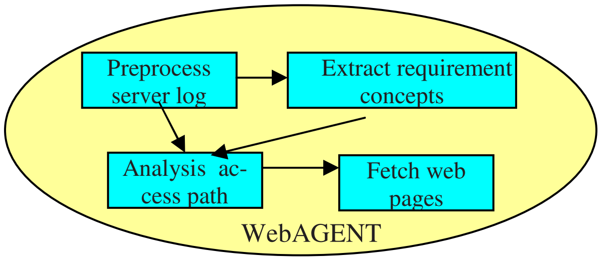


Fig. 1. The architecture of the WebAGENT. This also shows the tasks of the WebAGENT

2.2.1 Preprocess the Server Log Data for Prefetching

The web prefetching is constructed on the basis of analyzing the server log. The WebAGENT preprocess the server log file, extract user sessions and form a collection of user sessions. Algorithm 1 shows the main procedure.

Algorithm 1. Preprocessing the Server Log

1) Removing all false requests and requests for web pages such as graphic files, .cgi files.

2) Segregating the log file into the independent collections of requests according to IP address.

3) Processing each independent request collection, respectively.

- Sorting the requests by access time.
- Extracting sessions from the user request collection. For any two adjacent requests, if the time interval between its access time is smaller than the time threshold *tw*, these requests belong to the same user session.
- Gathering all sessions of the user into the user's session collection.

4) All user session collections compose the server session collection.

2.2.2 Extract the Requirement Concepts

According to Observation 1, access path can be regard as the deep search for certain concepts. Observation 2 shows that anchor text of hyperlink generalizes content of page linked. The user can judge directly whether the page linked is worth browsing or not, without having actually read the page. Therefore, Let *p* be the current page of access path, anchor texts of all hyperlinks that page *p* contains are the basis on which the user chooses the next page. So All concepts that those anchor texts contain compose the user's information requirement concept collection. We define a pseudo document, denoted as *HyperDoc* , which contains all anchor texts of hyperlinks in the current page *p*. We have:

$$HyperDoc = \bigcup_{l' \in L} l'.Anchor\_Text \tag{4}$$

where, *L* is the set of hyperlinks,

$$L = \{ l' \mid l'.SP\_ID = p.P\_ID \}$$

HyperDoc is processed for generating the requirement concept set. The main steps can be described in the following algorithm 2.

Algorithm 2. The Main Procedure of Composing the Requirement Concept Set *C*
1) Removing all html tags from *HyperDoc*.
2) If *HyperDoc* is in Chinese, word segmentation is for *HyperDoc* .
3) Removing all words that belong to a stop-word list.
4) If *HyperDoc* in English, all words are stemmed.
5) All results from above steps form the user's requirement concept set *C*.

2.2.3 Analysis User Access Path

Following the above steps, the WebAGENT calculates the probability of concept sequence of each concept in the requirement concept set. For example, Let E=<p1,p2,…,pn-1,pn> be the user path access, where pn is the current page, n is the length of access path. $o = (\underbrace{c_i c_i \ldots c_i}_{n})$ is the observation sequence, where ci ∈ C. The probability of *O* represents possibility that the concept c_i is latent requirement concept that access path *E* implies.

While responding access requests of users, the web server imports the request into corresponding user buffer. The user buffer is flushed once a time threshold *tw* is exceeded. The request sequence in the time intervals *tw* will be regarded as the current user's access path. Because the access path *E* is fixed, we only consider the access path as the fixed–state sequence. The probability of the observation sequence O is

$$P(o \mid q, \lambda) = \prod_{t=1}^{n} P(o_t \mid q_t, \lambda) = b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdot \ldots \cdot b_{q_n}(o_n) \tag{5}$$

where, $b_{q_j}(o_i)$ is the probability of observation symbol o_i in state q_j. Transforming it into the probability of the concept observation sequence, we have

$$P(o \mid q, \lambda) = b_{p_1}(c_i) \cdot b_{p_2}(c_i) \cdot \ldots \cdot b_{p_n}(c_i) \tag{6}$$

where, $b_{p_j}(c_i)$ is the observation concept probability of the concept c_i in the page p_j.

Considering the characteristics of access path, main action of access path details certain concepts, and leads users to web pages that they satisfy. So, we calculate the page's navigation capability for these concepts as observation concept probability. For each concept in the requirement concept collection, the weight in a page formula is defined as follow,

$$W_p(c_i) = \frac{\left(\frac{tf_i}{tf_{\max}}\right)}{\sqrt{\sum_{j=1}^{m}\left(\frac{tf_j}{tf_{\max}}\right)^2}} \tag{7}$$

where, tf_i is the frequency of the concept c_i in the document p, $tf_{\max}$ is the maximum concept frequency of all concepts in the requirement concept set C.

Suppose the user's session set is T. Let $S(p)$ be the page set that contains all pages that follow the page p in the user sessions,

$$S(p) = \{ p' \mid \forall s' \in T, p' \in s', p \rightarrow p' \} \tag{8}$$

the navigation capacity of page p to the concept c_i is estimated as follows

$$Nav_p(c_i) = \sum_{p' \in S(p)} W_{p'}(c_i) \tag{9}$$

It is normalized by dividing each weight of a concept by the sequence root of the sum of the squared weights.

$$\overline{Nav}_p(c_i) = \frac{Nav_p(c_i)}{\sum_{c' \in C} Nav_p(c')} \tag{10}$$

The $b_p(c_i)$ is computed by using the following formula,

$$b_p(c_i) = \overline{Nav}_p(c_i) \tag{11}$$

The modified function $P(o \mid q, \lambda)$ is stated as follows:

$$P(o \mid q, \lambda) = \overline{Nav}_{p_1}(c_i) \cdot \overline{Nav}_{p_2}(c_i) \cdot \ldots \cdot \overline{Nav}_{p_n}(c_i) \tag{12}$$

2.2.4 Fetch the Web Pages

Based on the modified function described above, the agent sort the concept set by the probability of the observation sequence calculated and choose the first τ concepts to form the concept set η. In other word, we have $\eta = \{c_1, c_2, \ldots, c_\tau\}$, where c_i is the ith concept. In our implementation, it is reasonable to set τ to 7. The prefetching prior score of page p is estimated with the following formula.

$$Score(p) = \sum_{c' \in \eta} W_p(c') \tag{13}$$

Using this formula, the agent calculates the prior score of all web pages that the current page of the user access path cited. According to the prefetching threshold θ, the first θ pages are the prefetching pages and are load into the server cache in advance.

3 Experimental Evaluation

3.1 Experiment Design

3.1.1 Dataset

To evaluate the performance of the web prefetching scheme of the WebAGENT, we conducted experiments on the Test web server (http://test.dhu.edu.cn). The Test web server includes 865 html pages, main topics of which are the related techniques about Chinese information processing. We obtained the server log from Jun. 1, 2004 to Sept. 30, 2004, which contains 22169 user requests. Set the time threshold *tw* to be two hours. Given the preprocessing step outlined above, the characteristics of the server log are shown in the Table 1. For evaluation purpose, we divide the complete dataset into the training dataset and the testing dataset, which the training dataset contains the first 1500 sessions and the testing dataset contains the remaining sessions.

Table 1. The characteristics of the server log preprocessed

Total of user requests	22169
Total of user session collections	212
Total of Sessions.	3128
Avg. Session Length	7
Min. Sesesion Length	1
Max.Session Length	10

3.1.2 Evaluation Metrics

We define the following measures to evaluate the performance of the WebAGENT:

Definition 6: *Request Hit Ratio* is the ratio of number of pages requested that are accurately prefetched and the total pages requested.

Definition 7: *Session Hit Ration* is the ratio of number of pages requested that are accurately prefetched and the total pages requested in a session.

3.2 Experimental Results and Analysis

3.2.1 Experimental Results

Firstly, we measured *Request Hit Ratio* in the different step of the user access path. So as to do it, we set the prefetching threshold θ to be 4, regard each session in the testing set as the current user access path and calculate the *Request Hit Ratio* of the model in each step. Figure 2 shows the relation between *Request Hit Ratio* and number of the steps of the current access path(θ =4).

Secondly, we measured the avg. *session Hit Ratio* in the different prefetching threshold θ. Figure 3 shows the relationship between avg. *Session Hit Ratio* and the prefetching threshold θ.

3.2.2 Experimental Analysis

As can be seen, In figure 2, we observed that *Request Hit Ratio* of the scheme increases rapidly from 50.4% to 58.6% when number of steps increases from 2 to 7, but number of steps exceeds 7, it begins to drop and decreases from 58.6% to 55.5%. The main reason is that ongoing bias of the user's interests may occur as step exceeds the certain value, introduce noise into predictive web prefetching, and reduce the predictive precision, but the schema is always higher than the average value 50%. As Figure 3 shows, avg. *Session Hit Ratio* of the scheme increases rapidly while the prefetching threshold θ increases from 1 to 4, the rate of increase declined after θ exceeds 4. we observe that our scheme has better avg. *Session Hit Ratio* than the non-prefetching web access.

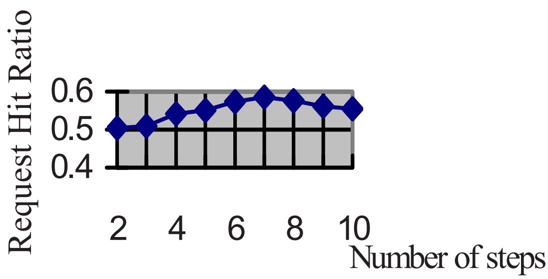


Fig. 2. Request Hit Ratio Vs Number of Steps (θ =4). This shows the relation between *Request Hit Ratio* and number of the steps of the current access path(θ =4).

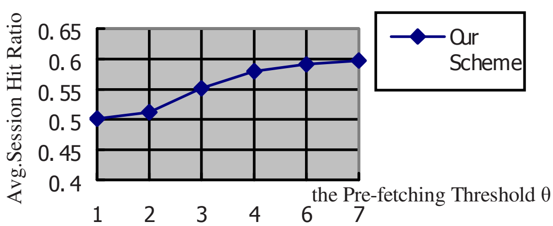


Fig. 3. Avg. Session Hit Ratio Vs the Prefetched Threshold θ. This shows the relationship between avg. *Session Hit Ratio* and the prefetching threshold θ.

4 Conclusions

Web prefetching reduces significantly the perceived latency and improves the service quality of web site, which is implemented successfully in many correlated applications. In this paper, we built the web prefetching agent-WebAGENT based on web ontology with HMM. The web prefetching scheme of the WebAGENT is based on the idea that it could make semantic-based prefetching decision in virtue of mining the latent information requirement concepts that the user access path implies. Web prefetching has common feature with other web applications that involve prediction of user access pattern. We hope this approach can be useful for reference by some relative research domains.

References

1. Evangelos P. Markatos and Catherine E.Chironaki, A Top 10 Approach for prefetching the web, *Proceedings of INET'98: Internet Global Summint* , July 1998
2. Styart Schechter, Murali Krishnan, and Michael D. Smith. Using Path profiles to predict http requests. *Proceedings of WWW7*, 1998.
3. SuYoung Yoon,Eunsook Jin, Jungmin Seo, Multimedia Technology ResearchLab, Korea, Telcom,http://www.isoc.org/inet99/proceedings/posters/106.
4. M.Deshpande, G.Karypis, Selective Markov Models for Predicting Web_Page Accesses, *Proceedings SIAM Int.Conference on Data Mining(SDM'2001)*,Apr.2001.
5. Sarukkai,Ramesh R.. Link Prediction and Path analysis Using Markov Chains, *9th World Wide Web Conference* ,May 2001.
6. XU Bao-wen,ZHANG Wei-feng, Applying Data Mining to Web Prefetching. *Chinese J.Computers,*2001,24(4):1-7.
7. Lawrence Rabiner, Riing-Hwang Juang, *Fundamentals of speech recognition*, Prentice Hall PTR. 1993:312-389.

Part XIII

Fuzzy Logic Theory

A Causal Model with UncertainTime-Series Effect Based on Evidence Theory

Vilany Kimala[1] and Koichi Yamada[2]

[1] 1603-1,Kami-tomioka, Nagaoka, Niigata, Japan, 940-2188 Information Science and Control Engineering Nagaoka University of Technology
kimala@mis.nagaokaut.ac.jp
[2] Management and Information Systems ScienceNagaoka University of Technology
yamada@kjs.nagaokaut.ac.jp

Abstract. Probability and possibility are both convenient scales of uncertainty, because they are defined by a distribution function. They also have complementary properties in the sense that probability is a quantitative and objective ratio scale, while possibility is a qualitative and subjective ordinal scale. The paper discusses probabilistic and possibilistic causal models with a time-series effect from the viewpoint of Evidence theory, and shows that they can be defined by a single equation with different conditions of focal elements using the basic probability assignments. The equation could be recognized as a causal model with a general representation of uncertainty in the form of Evidence theory. The paper finalizes the discussion with the properties of the generalized uncertain causal model.

Keywords: Uncertainty, Time-series effect, Possibility, Probability, Evidence theory, Plausibility.

1 Introduction

The paper studies a generalized uncertain causal model with time-series events as its effect. It is a model that a cause generates time-series effects with uncertainty.

The aim of this paper is to propose a new method based on evidence theory for dealing the disparity between knowledge types of time-series effects under uncertainty causal model. First, defines two types of probabilistic and possibilistic causal models, where prior knowledge is modeled by probability distribution and possibility distribution. Then, represent them by a common equation using the basic probability assignment, though they have different constraints on the same equation. The paper then discusses the characteristics of the causal model defined by the equation with no specific constraints. It is a generalized causal model in the sense of uncertainty. Lastly the paper shows that the generalized causal model is a plausibilistic causal model with similar characteristics to the probabilistic and possibilistic causal models defined at first.

2 Causal Model

The section describes the causal model with a time-series effect, which will be discussed in the paper. First, a causal model, which does not have any uncertainty information are defined, and then uncertainty measures are introduced in the model.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 547–555, 2007.
springerlink.com

2.1 Causal Model with Time-Series Effect

The causal model which has a time-series event as its effect is defined by the next definition.

Definition 1: *An event c arises at* $t=0$, *and a time-series event* $E_T =< e_1, e_2, ..., e_T >$ *follows as its effect. The event c is called cause, and must be an element of the set of possible causes* $U = \{u_1, u_2, ..., u_N\}$. E_t *is called time-series effect.* e_t *denotes an elemental event at time* t *in* E_t *. The actual entity of* e_t *is an element of the set of possible events* $V = \{v_1, v_2, ..., v_M, v_\varnothing\}$ *. It is assumed that* $U \cap V = \phi$, $e_t \neq e_{t'}$ ($t \neq t'$) *and* $e_T = v_\varnothing$ *,where* $v_\varnothing$ *is a terminal. Then, the causal model defined above is called causal model with a time-series effect.*

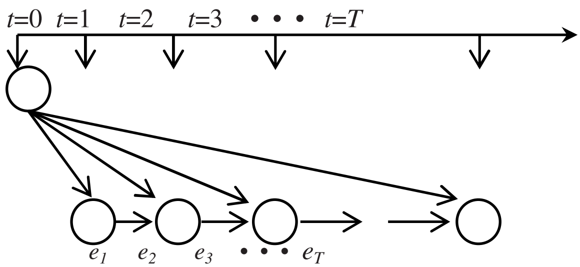


Fig. 1. A Causal Model with Time-series Effect

What should be noticed in the above definition is $e_t \neq e_{t'}$ for $t \neq t'$, which means that the same element in V never arises twice in a time-series effect. Thus, the length T of the time-series effect must be less than or equal to $|V| = M+1$. The terminal $v_\varnothing$ is introduced to recognize whether a given time-series effect is finished or not.

The causal model given by the definition can be utilized to model many systems: a mechanism that a disease causes a series of symptoms a series of the user's operations caused by her/his intention, etc. In the latter case, the user's intention could be a cause and a series of operations given by him/her could be a time-series effect.

In the following subsections are defined two types of uncertain causal models: probabilistic and possibilistic models. In addition, each model has two types of sub models with different characteristics: a submodel where the initial ratios among conditional probabilities/possibilities of e_1 given the causes are preserved until $t = T$, and a submodel where the initial ratios among transition probabilities/possibilities given the cause are preserved.

2.2 Probabilistic Causal Models

At first, we define a simple probabilistic model which preserves the ratios among conditional probabilities of elemental effects given the cause until $t = T$.

Definition 2: *Let the probability that cause* u_i *arises be given by a probability distribution* $p(u_i) \equiv p(c = u_i)$ *on* U. *Conditional probabilities* $p_1(v_j \mid u_i) \equiv p(e_1 = v_j \mid c = u_i)$ *are also given.* $p(u_i) > 0.0$ *for any* u_i. *Then, if* $p_t(v_j \mid u_i, E_{t-1}) \equiv p(e_t = v_j \mid u_i, E_{t-1})$ *is given by the next equation when* $t \geq 2$, *the causal model is called IPrC (Initial Probability ratio Conservation) model.*

$$p_t(v_j \mid u_i, E_{t-1}) = \begin{cases} \dfrac{p_1(v_j \mid u_i)}{\displaystyle\sum_{v_k \in V - \{E_{t-1}^1, \ldots, E_{t-1}^{t-1}\}} p_1(v_k \mid u_i)}, & \text{if } v_j \notin E_{t-1}, \\ 0, & \textit{Otherwise}, \end{cases} \tag{1}$$

where $p_1(v_\varnothing \mid u_i) > 0$ *and* $v_\varnothing \notin E_{t-1}$ *meaning that the terminal* $v_\varnothing$ *is not an elemental event of* $E_T =< e_1, e_2, \ldots, e_{t-1} >$. E_{t-1}^s *designates the entity of* e_s *in* E_{t-1} *(* $s \leq t-1$ *).*

The above equation means that the probability of $e_t = v_j$ conditioned by the cause u_i and the time-series effect E_{t-1} is zero if v_j is an elemental event of E_{t-1} and is equal to the normalized probability of $p_1(v_j \mid u_i)$ otherwise. The normalized probability of $p_1(v_j \mid u_i)$ is the one divided by the sum of $p_1(v_k \mid u_i)$ where v_k is not an elemental event of E_{t-1}. Since $v_\varnothing$ is the terminal, $p_t(v_j \mid u_i, E_{t-1})$ is not defined when $v_\varnothing \in E_{t-1}$. $p_1(v_\varnothing \mid u_i) > 0$ is required in order that $v_\varnothing$ is possible.

The model where the ratios among transition probabilities given the cause are preserved is defined by the following.

Definition 3: *In addition to the probability distribution* $P(u_i) \equiv P(c = u_i)$ *on* U *and Conditional probabilities* $p_1(v_j \mid u_i) \equiv p(e_1 = v_j \mid c = u_i)$, *the initial transition probabilities* $p_{12}(v_j \mid u_i, v_k) \equiv p(e_2 = v_j \mid c = u_i, e_1 = v_k)$ *are also given. Then, if* $p_t(v_j \mid u_i, E_{t-1}) \equiv p(e_t = v_j \mid u_i, E_{t-1})$ *is give by the next equation when* $t \geq 2$, *the causal model is called TPrC (Transition Probability ratio Conservation) model.*

$$p_t(v_j \mid u_i, E_{t-1}) = \begin{cases} \dfrac{p_{12}(v_j \mid u_i, E_{t-1}^{t-1})}{\displaystyle\sum_{v_k \in V - \{E_{t-1}^1, \ldots, E_{t-1}^{t-1}\}} p_{12}(v_k \mid u_i, E_{t-1}^{t-1})}, & \text{if } v_j \notin E_{t-1}, \\ 0, & \textit{Otherwise}, \end{cases} \tag{2}$$

where $v_\varnothing \notin E_{t-1}$, $p_{12}(v_\varnothing \mid u_i, v_k) \neq 0$ *and* $p_{12}(v_k \mid u_i, v_k) = 0$ *for any* $v_k (\neq v_\varnothing)$ *are required.*

The above equation means that the probability $p_t(v_j \mid u_i, E_{t-1})$ is equal to the normalized probability of $p_{12}(v_j \mid u_i, v_k)$, when v_j is not an elemental effect of E_{t-1}, and zero otherwise. The term normalization is used in the same sense as described before.

2.3 Possibilistic Causal Models

The subsection defines possibilistic versions of IPrC and TPrC models, which are called Initial Possibility ratio Conservation (IPoC) model and Transition Probability ratio Conservation (TPoC) model, respectively.

Definition 4: *Let the possibility that cause* u_i *arises be given by a possibility distribution* $\pi(u_i) \equiv \pi(c = u_i)$ *on* U. *Conditional possibilities* $\pi_1(v_j \mid u_i) \equiv \pi(e_1 = v_j \mid c = u_i)$ *are also given.* $\pi(u_i) > 0.0$ *for any* u_i. *Then, if* $\pi_t(v_j \mid u_i, E_{t-1}) \equiv \pi(e_t = v_j \mid u_i, E_{t-1})$ *is given by the next equation when* $t \geq 2$ *the causal model is called IPoC (Initial Possibility ratio Conservation) model.*

$$\pi_t(v_j \mid u_i, E_{t-1}) = \begin{cases} \dfrac{\pi_1(v_j \mid u_i)}{\underset{v_k \in V - \{E_{t-1}^1, \ldots, E_{t-1}^{t-1}\}}{Max} \pi_1(v_k \mid u_i)}, \text{if } v_j \notin E_{t-1}, \\ 0, \textit{Otherwise}, \end{cases} \tag{3}$$

where $v_\varnothing \notin E_{t-1}$ *and* $\pi(v_\varnothing \mid u_i) > 0$.

Definition 5: *In addition to the possibility distribution* $\pi(u_i) \equiv \pi(c = u_i)$ *on* U *and Conditional possibilities* $\pi_1(v_j \mid u_i) \equiv \pi(e_1 = v_j \mid c = u_i)$, *the initial transition probabilities* $\pi_{12}(v_j \mid u_i, v_k) \equiv \pi(e_2 = v_j \mid c = u_i, e_1 = v_k)$ *are also given. Then, if* $\pi_t(v_j \mid u_i, E_{t-1}) \equiv \pi(e_t = v_j \mid u_i, E_{t-1})$ *is given by the next equation when* $t \geq 2$, *the causal model is called TPoC (Transition Possibility ratio Conservation) model.*

$$\pi_t(v_j \mid u_i, E_{t-1}) = \begin{cases} \dfrac{\pi_{12}(v_j \mid u_i, E_{t-1}^{t-1})}{\underset{v_k \in V - \{E_{t-1}^1, \ldots, E_{t-1}^{t-1}\}}{Max} \pi_{12}(v_k \mid u_i, E_{t-1}^{t-1})}, \text{if } v_k \notin E_{t-1}, \\ 0, \quad \textit{Otherwise}, \end{cases} \tag{4}$$

where $v_\varnothing \notin E_{t-1}$, $\pi_{12}(v_\varnothing \mid u_i, v_k) \neq 0$ *and* $\pi_{12}(v_k \mid u_i, v_k) = 0$ *for any* $v_k (\neq v_\varnothing)$ *are required.*

IPoC and TPoC models defined above are basically the same as the probabilistic ones, respectively, except that the uncertainty is represented by possibility instead of probability.

3 Representation by Basic Probability Assignment

We discuss how to represent the causal models defined in the previous section in the form of basic probability assignment. In the following, $m(B)$ is the *bpa* of $B \subseteq U$ for cause c and $m_1(A \mid u_i)$ is the *bpa* of $A \subseteq V$ for the first elemental event e_1 given the cause u_i. $m_t(A \mid u_i, E_{t-1})$ denotes the *bpa* of A for e_t in the situation where cause $c = u_i$

and $E_{t-1} =< e_1,...,e_{t-1} >$ are observed ($t \geq 2$). In addition, $m_t(A \mid u_i, e_{t-1} = v_j)$ denotes the *bpa* of *A* for e_t when $c = u_i$ and $e_{t-1} = v_j$ are observed. Then we define $m_{12}(A \mid u_i, v_j) \equiv m_2(A \mid u_i, e_1 = v_j)$, which means $m_{12}(A \mid u_i, v_j)$ is the *bpa* for e_2 given the cause u_i and the first elemental effect $e_1 = v_j$.

First, consider a causal model with a time-series effect whose basic probability assignment $m_t(A \mid u_i, E_{t-1})$ is given by the following equation.

$$m_t(A \mid u_i, E_{t-1}) = \begin{cases} 0, \text{ if } A = \varnothing \text{ or } E_{t-1}^{t-1} \in A, \\ \dfrac{m_{t-1}(A \mid u_i, E_{t-2}) + m_{t-1}(A \cup \{E_{t-1}^{t-1}\} \mid u_i, E_{t-2})}{1 - m_{t-1}(\{E_{t-1}^{t-1}\} \mid u_i, E_{t-2})}, \\ \textit{otherwise}, \end{cases} \tag{5}$$

where $m_1(A \mid u_i, E_0) = m_1(A \mid u_i)$, $v_\varnothing \notin E_{t-1}$. $m(B), B \subseteq U$ *and* $m_1(A \mid u_i), A \subseteq V$ *are given a priori.*

The above equation shows that *bpa* of *A* for e_t given the cause u_i and the time-series effect E_{t-1} is zero if the last elemental event E_{t-1}^{t-1} of $E_{t-1} =< E_{t-1}^1, E_{t-1}^2 ..., E_{t-1}^{t-1} >$ is an element of *A*. Otherwise, it is equal to the sum of $m_{t-1}(A \mid u_i, E_{t-2})$ and $m_{t-1}(A \cup \{E_{t-1}^{t-1}\} \mid u_i, E_{t-2})$ divided by $1 - m_{t-1}(\{E_{t-1}^{t-1}\} \mid u_i, E_{t-2})$, which means 1) the *bpa* of a subset containing E_{t-1}^{t-1} becomes zero, 2) the value of $m_{t-1}(A \cup \{E_{t-1}^{t-1}\} \mid u_i, E_{t-2})$ is added to $m_{t-1}(A \mid u_i, E_{t-2})$ if $A \neq \varnothing$, and 3) the value of $m_{t-1}(\{E_{t-1}^{t-1}\} \mid u_i, E_{t-2})$ is discarded and the other values are normalized by the rest of the basic probability assignment.

The following propositions hold for the causal model defined above.

Proposition 1: *The causal model given by eq. (5) is an IPrC model, if all focal elements of* $m(B)$ *and* $m_1(A \mid u_i)$ *are singletons.*

Proof : Omitted due to the space limitation.

Proposition 2: *The causal model given by eq. (5) is an IPoC model, if all focal element of* $m(B)$ *are consonant (nested), and so as* $m_1(A \mid u_i)$.

Proof : Omitted due to the space limitation.

Then, consider another causal model defined by the next equation.

$$m_t(A \mid u_i, E_{t-1}) = \begin{cases} 0, \text{if } A = \varnothing \text{ or } A \neq F - S_{t-1} \text{ for any } F, \\ \dfrac{\sum_{F; F - S_{t-1} = A} m_{12}(F \mid u_i, E_{t-1}^{t-1})}{1 - \sum_{F: F - S_{t-1} = \varnothing} m_{12}(F \mid u_i, E_{t-1}^{t-1})}, \textit{otherwise}, \end{cases} \tag{6}$$

where $m_1(A|u_i,E_0)=m_1(A|u_i)$, $v_\varnothing \notin E_{t-1}$ *and* $S_{t-1}=\{v_j|v_j\in E_{t-1}\}$. F *is a focal element of* $m_{12}(\bullet|E_{t-1}^{t-1})$. *When there is no* F *such that* $F-S_{t-1}=\varnothing$ *or* $F-S_{t-1}=A$, *the sum of* $m_{12}(F|u_i,E_{t-1}^{t-1})$ *for* F *is zero.*

Then, the following proposition hold,

Proposition-3 : *The causal model defined by eq. (6) is a TPrC model, if all focal elements of* $m(B)$, $m_1(A|u_i)$ *and* $m_{12}(F_k|u_i,v_j)$ *are singletons.*

Proof : Omitted due to the space limitation.

Proposition-4 : *The causal model defined by eq. (6) is a TpoC model, if all focal elements* $m(B)$ *are consonant and so as* $m_1(A|u_i)$ *and* $m_{12}(F_k|u_i,v_j)$.

Proof : *Omitted due to the space limitation.*

The section showed that the causal models given by eq. (5) and (6) become IPrC and TPrC models respectively, when all focal elements of the given bpas are singletons. They also become IpoC and TpoC models, when focal elements of all the given bpas are consonant. In the next section, we investigate the characteristics of the causal models given by eq. (5) and (6) in general, namely, in the case where there are no conditions about the focal elements.

4 Plausibilistic Causal Models

This section discuses what kinds of models the eq.(5) and (6) will be, in the case where there are no conditions about the focal elements. In the following, $Pl(B)$ is the plausibility of $B\subseteq U$ for cause c and $Pl_1(A|u_i)$ is the plausibility of $A\subseteq V$ for the first elemental event e_1 given the cause u_i. $Pl_t(A|u_i,E_{t-1})$ denotes the *plausibility* of A for e_t in the situation where cause $c=u_i$ and $E_{t-1}=<e_1,...,e_{t-1}>$ are observed ($t\geq 2$). In addition, $Pl_t(A|u_i,e_{t-1}=v_j)$ denotes the *plausibility* of A for e_t when $c=u_i$ and $e_{t-1}=v_j$ are observed. Then we define $Pl_{12}(A|u_i,v_j)\equiv Pl_2(A|u_i,e_1=v_j)$.

First, consider a causal model with a time-series effect whose plausibility $Pl_t(A|u_i,E_{t-1})$ is given by the following equation.

$$Pl_t(A|u_i,E_{t-1})=\begin{cases}\dfrac{Pl_1(A|u_i)}{Pl_1(F|u_i)}, if \quad E_{t-1}^{t-1}\notin A,\\ 0, if \qquad Otherwise \quad ,\end{cases} \tag{7}$$

where $v_\varnothing \notin E_{t-1}$, $F=V-\{E_{t-1}^1,E_{t-1}^2,...,E_{t-1}^{t-1}\}$ *and* $Pl_1(\{\bullet\}|u_i,E_0)=Pl_1(\{\bullet\}|u_i)$ $Pl(A)>0.0$ *for any* A.

The above equation means that the plausibility of A for e_t given the cause u_i and time-series effect E_{t-1} is zero if E_{t-1}^{t-1} is an element of A. Otherwise, it is equal to the normalized plausibility of $Pl_1(\bullet \mid u_i)$. The normalized plausibility is the one divided by $Pl_1(F \mid u_i)$.

Then, the following propositions hold,

Proposition 5 : *The causal model presented by plausibility given by eq.(7) is the same as the one given by eq.(5).*

Proof : Omitted due to the space limitation.

Then, consider another causal model defined by next equation.

$$Pl_t(A \mid u_i, E_{t-1}) = \begin{cases} \dfrac{Pl_{12}(A \mid u_i, E_{t-1}^{t-1})}{Pl_{12}(F \mid u_i, E_{t-1}^{t-1})}, if & E_{t-1}^{t-1} \notin A, \\ 0, if & Otherwise \ , \end{cases} \tag{8}$$

where $v_\varnothing \notin E_{t-1}$, $F = V - \{E_{t-1}^1, E_{t-1}^2, ..., E_{t-1}^{t-1}\}$, $Pl_1(\{\bullet\} \mid u_i, E_0) = Pl_1(\{\bullet\} \mid u_i)$ $Pl_{12}(\{v_k\} \mid u_i, v_k) = 0$ *for any* $v_k \neq v_\phi$ *are required.*

Proposition 6: *The causal model presented by plausibility given by eq.(8) is the same as the one given by the models given by eq.(6).*

Proof : Omitted due to the space limitation.

The section showed the causal models given by eq.(7) and (8) are the same as eq.(5) and (6) in general, respectively. Therefore, in the case where there are no conditions about the focal elements, the eq.(5)and(6) could be called initial/transition plausibility ratio conservation models, respectively. Thus, eq.(5)and(6) are generalized uncertain causal model. The following example can help us understand of this point.

Example: Assume $V = \{v_1, v_2, v_3, v_4\}$ and time $T = 4$. *bpa* $m_1(\{\bullet\} \mid u_i)$, and $m_{12}(\{\bullet\} \mid u_i, v_j) = \tau_{ij}$ are given by the following vector λ and matrix τ ,where the subset $Z_i \mid i = 1,2,...,15$, $Z_i \subset V$ are given in table1.

$$\lambda = (m_1(Z_j \mid u_i)),\ j = 1,2,...,15$$

$$\lambda = (0.065, 0.065, 0.055, 0.055, 0.07, 0.065, 0.1, 0.12, 0.05, 0.05, 0.045, 0.065, 0.045, 0.08, 0.07)$$

$$\tau = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{ccccccccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Z_5 & Z_6 & Z_7 & Z_8 & Z_9 & Z_{10} & Z_{11} & Z_{12} & Z_{13} & Z_{14} & Z_{15} \end{array} \\ \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.095 & 0.12 & 0.15 & 0.25 & 0.09 & 0.195 & 0.1 \\ 0.09 & 0.0 & 0.15 & 0.15 & 0.0 & 0.0 & 0.25 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.18 & 0.08 \\ 0.12 & 0.145 & 0.0 & 0.135 & 0.0 & 0.195 & 0.0 & 0.0 & 0.125 & 0.0 & 0.18 & 0.0 & 0.0 & 0.0 & 0.1 \\ 0.15 & 0.155 & 0.165 & 0.0 & 0.195 & 0.0 & 0.0 & 0.0 & 0.09 & 0.145 & 0.0 & 0.0 & 0.1 & 0.0 & 0.0 \end{bmatrix} \end{array}$$

Table 1. Subset of V

$Z_1 = \{v_1\}$	$Z_4 = \{v_1, v_4\}$	$Z_7 = \{v_1, v_3, v_4\}$	$Z_{10} = \{v_2, v_3\}$	$Z_{13} = \{v_3\}$
$Z_2 = \{v_1, v_2\}$	$Z_5 = \{v_1, v_2, v_3\}$	$Z_8 = \{v_1, v_2, v_3, v_4\}$	$Z_{11} = \{v_2, v_4\}$	$Z_{14} = \{v_3, v_4\}$
$Z_3 = \{v_1, v_3\}$	$Z_6 = \{v_1, v_2, v_4\}$	$Z_9 = \{v_2\}$	$Z_{12} = \{v_2, v_3, v_4\}$	$Z_{15} = \{v_4\}$

Because there are $2^4 - 1 = 15$ subsets in 2^V, there are 15 columns in matrix.

Suppose that the initial event $e_1 = v_1$ is given. thus $E_1 =< v_1 >$.Then even at $t = 2$ is $e_2 = v_2$, thus $E_2 =< v_1, v_2 >$, and then event at $t = 3$ is $e_3 = v_3$, $E_3 =< v_1, v_2, v_3 >$. In this case, from knowledge of λ and τ, $m_t(\{\bullet\} \mid u_i, E_t)$ at $t \geq 2$ can be obtained according to eq.(6). Then $Pl_t(\{\bullet\} \mid u_i, E_t)$ at $t \geq 2$ can calculate by using $Pl_t(A \mid u_i) = \sum_{B \cap A \neq \phi} m_t(B \mid u_i)$. They are summarized in Table2.

Moreover, from λ and τ, $Pl_1(A \mid u_i)$ and $Pl_{12}(A \mid u_i, E_{t-1}^{t-1})$ can be obtained according to $Pl_t(A \mid u_i) = \sum_{B \cap A \neq \phi} m_t(B \mid u_i)$ [6]. They are shown in vector Pl_λ and matrix Pl_τ.

$$Pl_\lambda = (0.595, 0.805, 0.835, 0.855, 0.9, 0.955, 0.95, 1.0, 0.53, 0.81, 0.835, 0.935, 0.585, 0.82, 0.6),$$

$$Pl_\tau = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{ccccccccccccccc} Z_1 & Z_2 & Z_3 & Z_4 & Z_5 & Z_6 & Z_7 & Z_8 & Z_9 & Z_{10} & Z_{11} & Z_{12} & Z_{13} & Z_{14} & Z_{15} \end{array} \\ \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.615 & 0.9 & 0.91 & 1.0 & 0.655 & 0.905 & 0.695 \\ 0.64 & 0.0 & 0.92 & 0.9 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.68 & 0.91 & 0.66 \\ 0.595 & 0.9 & 0.0 & 0.975 & 0.0 & 1.0 & 0.0 & 0.0 & 0.645 & 0.0 & 0.88 & 0.0 & 0.0 & 0.0 & 0.61 \\ 0.665 & 0.9 & 0.91 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.585 & 0.85 & 0.0 & 0.0 & 0.605 & 0.0 & 0.0 \end{bmatrix} \end{array},$$

Then, From obtained knowledge of Pl_λ and Pl_τ, $Pl_t(\bullet \mid u_i, E_t)$ at $t \geq 2$ can calculate by using eq.(8). The calculate result are the same as Table2.

In table2 we note $Pl_t = Pl_t(\bullet \mid u_i, E_{t-1}),\ t = 1,2,3,4$

Table 2. Calculated result of plausibility at time t

Subsets	Pl_1	Pl_2	Pl_3	Pl_4
Z_1	0.595	0.0	0.0	0.0
Z_2	0.805	0.0	0.0	0.0
Z_3	0.835	0.0	0.0	0.0
Z_4	0.855	0.0	0.0	0.0
Z_5	0.93	0.0	0.0	0.0
Z_6	0.955	0.0	0.0	0.0
Z_7	0.95	0.0	0.0	0.0
Z_8	1.0	0.0	0.0	0.0
Z_9	0.53	0.615	0.0	0.0
Z_{10}	0.81	0.9	0.0	0.0
Z_{11}	0.835	0.91	0.0	0.0
Z_{12}	0.935	1.0	0.0	0.0
Z_{13}	0.585	0.655	0.7472	0.0
Z_{14}	0.82	0.905	1.0	0.0
Z_{15}	0.6	0.695	0.7253	1.0

How to calculate the values of $Pl_t(A|u_i,E_{t-1})$ by using eq.(8).In the case $t=1, E_1 =< v_1 >$, we choose the line of v_1 in matrix Pl_τ to calculate by using eq.(8), for example.

$$Pl_2(\{v_2\}|u_i,v_1) = \frac{Pl_{12}(\{v_2\}|u_i,v_1)}{Pl_{12}(\{v_2,v_3,v_4\}|u_i,v_1)} = \frac{0.615}{1.0} = 0.615$$

In the case $t=2, E_2 =< v_1,v_2 >$

$$Pl_3(\{v_3\}|u_i,v_1,v_2) = \frac{Pl_{12}(\{v_3\}|u_i,v_2)}{Pl_{12}(\{v_3,v_4\}|u_i,v_2)} = \frac{0.68}{0.91} = 0.7472$$

From calculated result of $Pl_t(\bullet\,|\,u_i,E_t)$ at $t \geq 2$ in above, it clearly that eq.(6) and eq.(8) is the same equation. Then, the obtain result in Table2 shows, when event, $e_1 = v_1$ arise the initial ratios among transition plausibility $Pl_{12}(\{\bullet\}|u_1,v_1)$ given the cause are preserved until time $t=2$, and then $e_2 = v_2$ arise the initial ratios among transition plausibility $Pl_{12}(\{\bullet\}|u_1,v_2)$ given the cause are preserved until time $t=3$. Therefore, it is clear that the properties of causal models given by eq.(5)and(6) are initial plausibility and initial transition plausibility preserved respectively.

5 Conclusion

The paper defined two types of uncertain causal models: probabilistic and possibilistic causal models. Then authors showed that the probabilistic and possibilistic model can be defined by a single equation with different specific constraints of focal element using basic probability assignment. In addition, the paper discussed the characteristics of a generalized uncertain causal model given by the equation with no specific constraints. The model is a plausibilistic one with characteristics similar to those of the probabilistic and possibilistic causal models defined at first.

References

1. G.Shafer: *A mathematical theory of evidence*, Princeton Univ.Press(1979)
2. G.J.Klir and T.A.Folger, Fuzzy sets, uncertainty and information, Prentice Hall, 1988.
3. D.Dubois and H. Prade: Epistemic entrenchment and possibilistic logic, *Artificial Intelligence*, Vol.50, 1991, pp.223-239.
4. D.Dubois and H. Prade, A survey of belief revision and updating rules in various uncertainty models, Int. J. *Intelligent systems*, Vol.9, 1994, pp.61-100.
5. K.Yamada, Probability-Possibility transformation based on Evedence theory, *IFSA World Congress 2001*, 2001, pp.70-75
6. G.J.Klir: *Fuzzy Sets and Fuzzy Logic Theory And Applications*, Prentice Hall PTR (1995)
7. V.Kimala, K.Yamada “Causal Model with Uncertain Time-series Events for Reasoning Causes,” *Proc. the 21th Fuzzy Systems Symposium,* pp.439-442, 2005(in Japanese)

Parametric Fuzzy Linear Systems

Iwona Skalna

Department of Applied Computer Science
AGH University of Science and Technology

Abstract. Systems of linear equations with elements being affine linear functions of fuzzy parameters are relevant to many practical problems. A method for solving such systems is proposed. It consists of two steps. First a finite number of parametric interval linear systems is solved using the direct method. Then membership functions of fuzzy solution elements are interpolated. Parameters are modeled by arbitrary fuzzy numbers with convex membership function and compact support. Conditions for existence of the fuzzy solution are given. The performance of the proposed method is presented using an illustrative example of truss structure.

1 Introduction

Many economical, financial, physical, engineering and electrical problems boil down to solution of linear systems of equations. When the problem characteristics are imprecise, then the linear system of equations is no longer crisp. Imprecise or unknown values of the system parameters can be modeled using probability distributions, intervals or fuzzy values.

Several methods for solving linear interval system [13], [17] have been developed since '60s [11] when interval arithmetic became more and more popular. Most of these methods assume implicitly that system coefficients vary independently within the lower and upper bound of the corresponding intervals (such system will be called classical). Usually this is not the case. Moreover, the assumption of coefficients independence makes many problem to be unsolvable [10]. This beget the need to develop methods for solving parametric linear systems [5], [6-9], [14], [18], [19] – parametrisation is used to eliminate drawbacks of interval arithmetic [13], especially the problem of dependency [3], [10]. Using the direct method proposed in [20] large systems of parametric linear equations with a lot of interval parameters can be solved efficiently. Since the solution set of interval system is usually non-convex and has a very complicated shape, instead of the solution set itself, an outer and inner interval solutions are computed. The best outer solution coincides with an interval hull solution.

Some investigations have also been reported in literature on the solution of classical fuzzy linear systems [1], [3], [12]. However, based on the extension principle, the solution of fuzzy linear equations can be viewed as a result of series of solutions of a nested family of interval systems. Hence the problem of dependency is also relevant to fuzzy linear systems. An optimization-based scheme for numerical solution of parametric fuzzy linear systems has been proposed in [16]. However, the approach appears to be difficult and inefficient for solving large systems of equations. The

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 556–564, 2007.
springerlink.com

search-based algorithm for solving parametric fuzzy linear systems was presented in [15]. In both paper considerations were restricted to triangular fuzzy numbers.

In this paper an efficient method for solving parametric linear systems with elements linearly dependent on a vector of fuzzy parameters is proposed. In this context to solve means to compute an outer interval solution. Two major steps constitute the framework of the proposed methodology: 1) solve a finite number of linear systems with interval parameters corresponding to the α-cuts non-uniformly distributed on [0, 1] interval; 2) interpolate membership functions of fuzzy solution elements with cubic splines. Non-uniform distribution of the α-cuts reflects the shape of the membership function of fuzzy parameters. Interpolation is used to compute approximations of the intermediate values of the membership functions of the fuzzy solution elements.

Problem parameters are modeled by arbitrary fuzzy numbers with continous convex membership function and compact support. Conditions for existence of the fuzzy solution are given. The performance of the proposed method is presented using an illustrative example of parametric fuzzy linear system that arise in analysis of truss structures.

2 Preliminaries

A real compact interval $\boldsymbol{a} = [\underline{a}, \overline{a}] = \{x \in R \mid \underline{a} \le x \le \overline{a}\}$. A family of real compact intervals will be denoted by IR. For an interval $\boldsymbol{a}$ define a midpoint $\breve{a} = \mathrm{mid}(\boldsymbol{a}) = (\overline{a} + \underline{a})/2$ and a radius $\breve{a} = \mathrm{rad}(\boldsymbol{a}) = (\overline{a} - \underline{a})/2$.

A square interval matrix $\boldsymbol{A}$ is an array of intervals:

$$\boldsymbol{A} = \{\boldsymbol{a}_{ij}\}_{i,j=1}^{n}, \ \boldsymbol{a}_{ij} \in IR, \ i, j = 1, \ldots, n \ .$$

One column ($n\times 1$) interval matrix is just an interval vector. IR^n, $IR^{n\times n}$ will denote a family of interval vectors, respectively, square interval matrices.

An interval matrix $\boldsymbol{A} \in IR^{n\times n}$ is regular (or non-singular) if all real matrices $A \in \boldsymbol{A}$ are non-singular.

An interval matrix $\boldsymbol{A} \in IR^{n\times n}$ is called an H-matrix [13] (should not be confused with a Hadamard matrix) iff there exist $u \in R^n$, $u > 0$ such that $\langle \boldsymbol{A}\rangle u > 0$. Here $\langle \boldsymbol{A}\rangle$ is a real $n\times n$ matrix with entries:

$$\langle \boldsymbol{a}\rangle_{ii} = \begin{cases} \min\{|\underline{a}_{ii}|, |\overline{a}_{ii}|\}, & 0 \notin \boldsymbol{a}_{ii} \\ 0 & 0 \in \boldsymbol{a}_{ii} \end{cases}, \ \langle \boldsymbol{a}\rangle_{ij} = -\max\{|\underline{a}_{ii}|, |\overline{a}_{ii}|\}, \ i \ne j \ .$$

and is called an Ostrovsky matrix [13].

Theorem 1 (Neumaier [13]). If $\boldsymbol{A} \in IR^{n\times n}$ is an H-matrix and $\boldsymbol{B} \subseteq \boldsymbol{A}$, $\boldsymbol{B} \in IR^{n\times n}$, then $\boldsymbol{B}$ is an H-matrix.

For an arbitrary set $X \subset R^n$ an interval hull X is defined as

$$X = \bigcap \{Y \in IR^n \mid X \subseteq Y\} \ .$$

A convex fuzzy number $\boldsymbol{a}$ with compact support is defined by its membership function $\mu_{\boldsymbol{a}}:R \to [0, 1]$, such that:

- there exist a unique $m \in R$ with $\mu_{\boldsymbol{a}}(m)=1$,
- a support $\text{supp}(\boldsymbol{a}) = Cl\{x \in R \mid \mu_{\boldsymbol{a}}(x) > 0\}$ is bounded in R,
- $\mu_{\boldsymbol{a}}$ is fuzzy convex on supp($\boldsymbol{a}$),
- $\mu_{\boldsymbol{a}}$ is upper semi-continous on R.

The function $\mu_{\boldsymbol{a}}$ is called fuzzy convex on supp($\boldsymbol{a}$) if

$$\mu_{\boldsymbol{a}}(\lambda x_1+(1-\lambda)x_2) \geq \min\{\mu_{\boldsymbol{a}}(x_1),\mu_{\boldsymbol{a}}(x_2)\} \ ,$$

for $0 \leq \lambda \leq 1$ and $x_1 \neq x_2$. A family of convex fuzzy numbers with continous membership function and compact support will be denoted by F.

For $\boldsymbol{a} \in F$, an interval corresponding to α-value of a membership function

$$\boldsymbol{a}(\alpha) = \{ x \mid \mu_{\boldsymbol{a}}(x) \geq \alpha \} \in IR, \ \alpha \in (0, 1] \ ,$$

is called a (weak) α-cut; a 0-cut is defined separately as $\boldsymbol{a}(0) = \text{supp}(\boldsymbol{a})$.

For each $\alpha \in [0, 1]$, $\boldsymbol{a}(\alpha)$ is compact and connected. A fuzzy number $\boldsymbol{a} \in F$ can be viewed as a nested family of α-cuts in a sense that

$$\boldsymbol{a}(\alpha) \subseteq \boldsymbol{a}(\beta), \quad \text{for } \alpha > \beta \ .$$

A square fuzzy matrix $\boldsymbol{A}$ is an array of fuzzy numbers:

$$\boldsymbol{A} = \{\boldsymbol{a}_{ij}\}_{i,j=1}^{n}, \ \boldsymbol{a}_{ij} \in F, \ i, j = 1, \ldots, n \ .$$

Fuzzy vector is a one column ($n\times1$) fuzzy matrix. F^n, $F^{n\times n}$ will denote a family of fuzzy vectors, respectively square fuzzy matrices.

Let $\boldsymbol{A} = \{\boldsymbol{a}_{ij}\}_{i,j=1}^{n}$ be a fuzzy matrix, then an α-cut

$$\boldsymbol{A}(\alpha) = \{\boldsymbol{a}_{ij}(\alpha)\}_{i,j=1}^{n}$$

is computed componentwise. $\boldsymbol{A}(\alpha)$, $0 \leq \alpha \leq 1$, is an interval matrix.

Corollary 1. Let $\boldsymbol{A} \in F^{n\times n}$. If $\boldsymbol{A}(0)$ is regular (an H-matrix), then $\boldsymbol{A}(\alpha)$ is regular (an H-matrix) for each $\alpha \in [0, 1]$.
Proof: Obvious.

3 Fuzzy Linear Systems

The following matrix equation

$$\boldsymbol{Ax} = \boldsymbol{b} \ , \tag{1}$$

with $\boldsymbol{a}_{ij}$, $\boldsymbol{b}_i \in F$ $(i, j = 1, \ldots, n)$, is called a fuzzy linear system. For a fuzzy linear systems different solution sets have been considered [3], [12]. Finally, three types of the solutions can be distinguished: classical solution $\boldsymbol{S}_C$, marginal solutions $\boldsymbol{S}_E$ and $\boldsymbol{S}_I$, and joint (or vector) solution $\boldsymbol{S}_J$. Classical solution $\boldsymbol{S}_C$ employs α-cuts and interval arithmetic in order to work out the solution. Taking α-cuts of Eq. (1) the system

$$\sum_{j=1}^{n}[\underline{a}_{ij}(\alpha), \overline{a}_{ij}(\alpha)][\underline{x}_j(\alpha), \overline{x}_j(\alpha)] = [\underline{b}_i(\alpha), \overline{b}_i(\alpha)] \ , \tag{2}$$

for $0 \le \alpha \le 1$, $1 \le i \le n$, is obtained. Interval multiplication and addition is used to evaluate the left-hand side of Eq. (2). Hence, for each $\alpha \in [0, 1]$, the original $n\times n$ system is transformed into a $2n\times 2n$ system. Eq. (2) is solved for $\underline{x}_j(\alpha)$, $\overline{x}_j(\alpha)$ hoping they produce the α-cuts of fuzzy numbers $\boldsymbol{x}_j$. Buckley et al. [2], [3] point out that very often the classical solution doesn't exist, which is mainly due to the dependency problem. $\boldsymbol{S}_E$ and $\boldsymbol{S}_I$ solutions are obtained by solving the corresponding crisp system using Cramer's rules, and then evaluating the solution using extension principle, respectively, interval arithmetic for each α-cut. The joint solution $\boldsymbol{S}_J$, on which the paper is focused, is a fuzzy vector that coincides with united solution set [13] of linear interval systems for each α-cut.

Set

$$\boldsymbol{S}(\alpha) = \{x \in R^n \mid Ax = b, A \in \boldsymbol{A}(\alpha), b \in \boldsymbol{b}(\alpha)\} \tag{3}$$

and define $\boldsymbol{S}_J$, a fuzzy subset of R^n, by its membership function

$$\mu_{\boldsymbol{S}_J}(x) = \begin{cases} \sup\{\alpha \mid x \in \boldsymbol{S}(\alpha)\}, & x \in \boldsymbol{S}(0) \\ 0 & x \notin \boldsymbol{S}(0) \end{cases} . \tag{4}$$

Theorem 2 (Buckley [3]). If $\boldsymbol{A}(0)$ is regular, then $\boldsymbol{S}_J$ is a fuzzy vector.

To eliminate the dependency problem, parametric fuzzy linear systems are considered.

4 Parametric Fuzzy Linear Systems

A system

$$\boldsymbol{A}(\boldsymbol{p})x = \boldsymbol{b}(\boldsymbol{p}) \ , \tag{5}$$

where $\boldsymbol{p} \in F^k$ is a vector of fuzzy numbers, is called a parametric fuzzy linear system. The problem of solving the fuzzy system (5) can be transformed [15] into an equivalent problem of solving a nested family of parametric linear interval systems

$$\boldsymbol{A}(\boldsymbol{p}(\alpha))\mathrm{x} = \boldsymbol{b}(\boldsymbol{p}(\alpha)), \quad \alpha \in [0, 1] \ . \tag{6}$$

Interval solution set of a parametric interval linear system $A(\boldsymbol{p})x = b(\boldsymbol{p})$, $\boldsymbol{p} \in IR^k$ is defined as

$$\boldsymbol{S}(\boldsymbol{p}) = \{x \in R^n \mid Ax = b, A \in \boldsymbol{A}(\boldsymbol{p}), b \in \boldsymbol{b}(\boldsymbol{p})\} \ . \tag{7}$$

Theorem 3. If $A(\boldsymbol{p})$ is regular, then interval solution of parametric interval linear system exists.

Proof: Obvious.

Accordingly to (4) define the joint vector solution $\boldsymbol{S}_J(\boldsymbol{p})$ of the parametric system (5) by its membership function

$$\mu_{S_J(\boldsymbol{p})}(x) = \mu(x \mid \boldsymbol{S}_J(\boldsymbol{p})) = \begin{cases} \sup\{\alpha \mid x \in \boldsymbol{S}(\boldsymbol{p}(\alpha))\}, & x \in \boldsymbol{S}(\boldsymbol{p}(0)) \\ 0 & x \notin \boldsymbol{S}(\boldsymbol{p}(0)) \end{cases} . \tag{8}$$

Theorem 4. If $\boldsymbol{A}(\boldsymbol{p}(0))$ is regular, then the join solution of parametric fuzzy linear exists.

Proof: See theorem (2).

In what follows parametric linear systems with elements linearly dependent on elements of a vector of fuzzy parameters $\boldsymbol{p} \in F^k$

$$\boldsymbol{a}_{ij}(\boldsymbol{p}) = \omega(i,j)^{\mathrm{T}}\boldsymbol{p}, \quad \boldsymbol{b}_j(\boldsymbol{p}) = \omega(0,j)^{\mathrm{T}}\boldsymbol{p} \tag{9}$$

are considered, where $\omega \in (R^k\})^{n\times n}$ is a matrix of real k-dimensional vectors. Elements of the induced family of parametric interval linear systems can be expressed as

$$\boldsymbol{a}_{ij}(\boldsymbol{p}(\alpha)) = \omega(i,j)^{\mathrm{T}}\,\boldsymbol{p}(\alpha), \quad \boldsymbol{b}_j(\boldsymbol{p}(\alpha)) = \omega(0,j)^{\mathrm{T}}\boldsymbol{p}(\alpha) \ , \tag{10}$$

for $\alpha \in [0, 1]$. Systems (6) will be solved using a direct method [20]. A brief description of the method is given in the next section.

5 Description of the Direct Method

An efficient direct method (DM for short) for solving parametric linear systems with elements linearly dependent on a set of interval parameters have been proposed in [20]. The method can be used to solve large systems with a lot number of interval parameters. The method is based on the following.

Theorem 5 (Skalna [20]). Let $\boldsymbol{A}(\boldsymbol{p})x = \boldsymbol{b}(\boldsymbol{p})$ with $\boldsymbol{p} \in IR^k$, $R \in R^{n\times n}$, and $\tilde{x} \in R^n$. If $\boldsymbol{C} \in IR^{n\times n}$ given by formula

$$\boldsymbol{C}_{\mathrm{ij}} = \sum_{k=1}^{n} R_{ik}\,\omega(k,\,j)^{\mathrm{T}}\,\boldsymbol{p} \,..,\tag{11}$$

is an H-matrix then

$$\boldsymbol{S}(\boldsymbol{p}) \subseteq \tilde{x} + \langle \boldsymbol{C}\rangle^{-1} |\, \boldsymbol{Z} \,|[-1, 1] \tag{12}$$

with

$$Z_i = \sum_{j=1}^{n} R_{ij} \left(\omega(0, j) - \sum_{k=1}^{n} \tilde{x}\, \omega(j, k) \right)^{\mathrm{T}} \boldsymbol{p} \tag{13}$$

It is recommended to choose $R = A^{-1}(\breve{p})$ and $\tilde{x} = A^{-1}(\breve{p})b(\breve{p})$ so that $\boldsymbol{C}$ and $\boldsymbol{Z}$ are of small norms (see theorem 4.1.10 [13]).

6 Illustrative Example

Consider a 20-floor cantilever truss structure depicted in Fig. 1. There are 42 nodes, 81 beams, full support at node 1 and partial support (sliding along Y-axis) at node 2. This results in 81 variables and 81 fuzzy parameters. Set Young's modulus $E = 2.0{\times}10^{11}$[Pa], cross section area $A = 0.005[\mathrm{m}^2]$, and the length of the vertical beams $L = 1$[m].

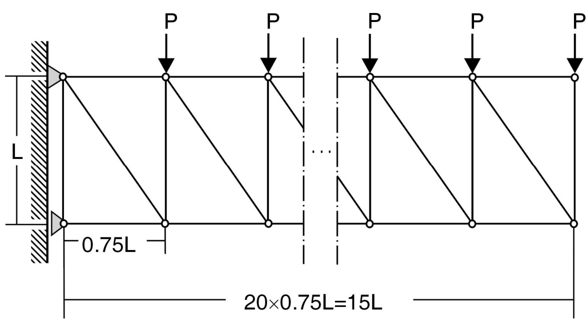


Fig. 1. 20-floor cantilever truss structure

To compute the displacements of the nodes, the following parametric fuzzy linear system

$$\boldsymbol{K}(\boldsymbol{s})d = \boldsymbol{Q}(\boldsymbol{s}) \ ,$$

has to be solved, where $\boldsymbol{K}(\boldsymbol{s})$ is a fuzzy stiffness matrix, $\boldsymbol{Q}(\boldsymbol{s})$ is a fuzzy vector of forces, $\boldsymbol{d}$ is unknown vector of fuzzy displacements, and $\boldsymbol{s}$ is a vector of fuzzy parameters.

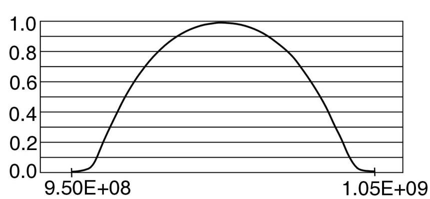


Fig. 2. Membership function of the fuzzy parameters

Assume the fuzzy parameters have the membership function (Fig. 2) given by formula

$$\mu_{s_{ij}}(x) = \begin{cases} e^{4(m_1-x)(x-m_2)-(m_2-m_1)^2/4(m_2-x)(x-m_1)}, & m_1 < x < m_2 \\ 0, & \text{elsewhere} \end{cases}$$

where $m_1 = E\cdot A\cdot(1-\varepsilon)$, $m_2 = E\cdot A\cdot(1+\varepsilon)$, $\varepsilon = 5\%$.

The results produced by DM method (10 α-cuts), for two chosen coordinates x_{41}, y_{41}, are presented in Table 1.

Table 1. Results of the DM method: 10 α-cuts

α	x_{41}	y_{41}
0.0	[0.00568536, 0.00927714]	[-0.331072, -0.0422026]
0.1	[0.00633381, 0.00862869]	[-0.260963, -0.112312]
0.2	[0.00648488, 0.00847762]	[-0.247216, -0.126059]
0.3	[0.00661057, 0.00835193]	[-0.236577, -0.136698]
0.4	[0.00672633, 0.00823617]	[-0.227446, -0.145829]
0.5	[0.00683787, 0.00812463]	[-0.219279, -0.153996]
0.6	[0.0069482, 0.0080143]	[-0.211839, -0.161436]
0.7	[0.00705947, 0.00790303]	[-0.205015, -0.16826]
0.8	[0.007174, 0.0077885]	[-0.198758, -0.174517]
0.9	[0.00729645, 0.00766605]	[-0.193018, -0.180257]
1.0	[0.00748125, 0.00748125]	[-0.186638, -0.186637]

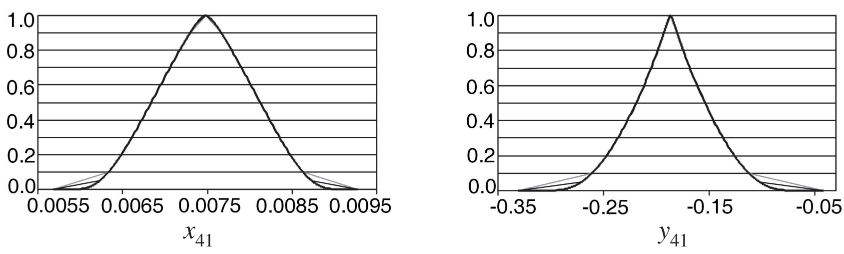


Fig. 3. Comparison of the results of the DM method obtained for 10 α-cuts (gray thin line), 20 α-cuts (black thin line) and 100 α-cuts (black thick line)

The differences between the shapes of the membership functions (based on 10, 20 and 100 α-cuts), depicted in the Fig. 3, are significant, especially for small values of α. The differences between the results of the DM method (451 α-cuts) and interpolation, based on 25 points (α-cuts) of non-uniform distribution (Fig. 4), is quite neligible. The benefit of the interpolation is that once the coefficients of cubic functions are computed, the approximation of the fuzzy solution can be easily computed for any α-value.

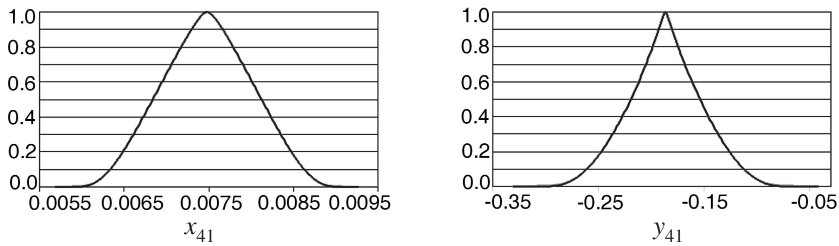


Fig. 4. Comparison of the results of the DM method for 451 α-cuts with the results of interpolation based on 25 points non-uniformly distributed on [0, 1] interval

6 Conclusions

Parametric linear systems with coefficients being affine linear functions of convex fuzzy parameters with compact support have been studied. A method for approximating the vector solution of such systems has been introduced. Conditions for existence of the vector solution have been given. It has been shown, using an illustrative example of 20-floor cantilever truss structure, that the method can be applied to large systems with a lot of fuzzy parameters. The method is very efficient and is easy to implement, which is very important for practical use.

References

[1] Buckley JJ (1992) Solving fuzzy equations. Fuzzy Sets and Systems 50:1-14
[2] Buckley JJ, Eslami E, Feuring T (2002) Fuzzy mathematics in economics and engineering. Studies in Fuzziness and Soft Computing 91. Physica-Verlag, Heidelberg, Germany
[3] Buckley JJ, Qu Y (1991) Solving systems of linear fuzzy equations. Fuzzy Sets and Systems 43:33-43
[4] Jankowski M (1990) Elements of Computer Graphics. Wydawnictwo Naukowo-Techniczne, Warsaw
[5] Jansson C (1991) Interval linear systems with symmetric matrices, skew-symmetric matrices and dependencies in the right hand side. Computing 46(3):265-274
[6] Kolev LV (1991) Interval methods for circuit analysis. Word Scientific Ltd., Singapore, New Jersey, London
[7] Kolev LV (2002) Outer solution of linear systems whose elements are affine functions of interval parameters. Reliable Computing 6:493-501
[8] Kolev LV (2002) Worst-case tolerance analysis of linear DC and AC electric circuits. IEEE Trans. on Circuits and Systems, I, Fund. Theory and Appl. 49:1693-1702
[9] Kolev LV (2004) A method for outer interval solution of linear parametric systems. Reliable Computing 10:227-239
[10] Kulpa Z, Pownuk A, Skalna I (1998) Analysis of linear mechanical structures with uncertainties by means of interval methods. Computer Assisted Mechanics and Engineering Sciences 5(4):443-477
[11] Moor RE (1966) Interval Analysis. Prentice Hall, EngleWood Cliffs N.J.

[12] Muzzioli S, Reynaerts H (2005) Fuzzy linear systems of the form A1x+b1 = A2x + b2. 4th International Symposium on Imprecise Probabilities and Their Application. Pitsburg, Pensylwania
[13] Neumaier A (1990) Interval methods for solving systems of equations. Cambridge University Press, Cambridge, UK
[14] Popova ED (2001) On the solution of parametrised linear systems. Scientific Computing, Validated Numerics, Interval Methods 127-138
[15] Rao SS, Chen L (1998) Numerical solution of fuzzy linear equations in engineering analysis. International Journal for Numerical Methods in Engineering 43: 391-408
[16] Rao SS, Sawyer JP (1995) Fuzzy finite element method for analysis of imprecisely defined systems. AIAA Journal 33(12):2364-2370
[17] Rohn J (1989) Systems of linear interval equations. Linear Algebra and its Applications 126:39-78
[18] Rump SM (1994) Verification methods for dense and sparse systems of equations. Studies in Computational Mathematics 5:63-135. Amsterdam, The Netherlands, Elsevier
[19] Skalna I (2007) Evolutionary optimization method for approximating the solution set hull of parametric linear systems. Lecture Notes in Computer Science
[20] Skalna I (2006) A method for outer interval solution of parametrized systems of linear interval equations. Reliable Computing 12(2):107-120

Lattice Ordered Monoids and Left Continuous Uninorms and t-norms

Marta Takacs

Budapest Tech, John von Neumann Faculty of Informatics 1036 Budapest,
Bécsi út 96/b, Hungary
takacs.marta@nik.bmf.hu

Abstract. The proposition about generalization of the existence of the residuum for left continuous uninorm U on a commutative, residuated l-monoid, with a neutral element is proved. The question raised previously was whether there are general operation groups which satisfy the residuum-based approximate reasoning, but at the same time are easily comprehensible and acceptable to application-oriented experts. The basic backgrounds of this research are the distance-based operators.

Keywords: uninorms, residuum, lattices.

1 Introduction

In the theory of approximate reasoning introduced by Zadeh in 1979 [15], the knowledge of system behavior and system control can be stated in the form of if-then rules. In Mamadani-based sources [7] it was suggested to represent an "if *x* is *A* then *y* is *B*" rule simply as a connection (for example as a t-norm, $T(A,B)$ or as *min*) between the so called rule premise: *x* is *A* and rule consequence: *y* is *B*.

Engineering applications based on Fuzzy Logic Systems (FLS) are mostly based on minimum and maximum operators, but from a mathematical point of view it is interesting to study the behavior of other operators in inference mechanism, which will give new possibilities for further applications. This gives as the link with the theory of pseudo-analysis, which serves also as a universal base for the system theory [6], [8]. Uninorm operator groups [17], researched in the past decade, have yielded exceptionally good results.

Applying the uninorm operators with the changeable parameter *e* in fuzzy approximate reasoning systems is born in mind that the underlying notions of soft-computing systems are flexibility and the human mind. The choice of the fuzzy environment must support the efficiency of the system and it must comply to the real world. This is more important than trying to fit the real world into the inflexible models. [1], [15], [16].

In recent years there have been numerous forms of research which analyze the application and theoretical place of the uninorms. As there is a near-complete theoretical basis for the t-norm based residuums, in a similarly detailed theoretical basis being constructed for uninorms.[4], [5], [14]

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 565–572, 2007.
springerlink.com

In the paper the basic definitions for uninorm, distance based operators and its residuum are given. Having results from [3], we introduce the residuum-based inference mechanism using uninorms, especially distance-based uninorms.

2 Uninorms and Distance Based Operators

Both the neutral element 1 of a t-norm and the neutral element 0 of a t-conorm are boundary points of the unit interval. However, there are many important operations whose neutral element is an interior point of the underlying set. The fact that the first three axioms (commutativity, associativity, montonocity) coincide for t-norms and for t-conorms, i.e., the only axiomatic difference lies in the location of the neutral element, has led to the introduction of a new class of binary operations closely related to t-norms and t-conorms.

A *uninorm* is a binary operation U on the unit interval, i.e., a function $U : [0,1]^2 \to [0,1]$ which satisfies the following properties for all $x, y, z \in [0,1]$

(U1) $U(x, y) = U(y, x)$, i.e. the uninorm is commutative,

(U2) $U(U(x, y), z) = U(x, U(y, z))$, i.e. the uninorm is associative,

(U3) $x \le y \Rightarrow U(x, z) \le U(y, z)$, i.e. the uninorm monotone,

(U4) $U(e, x) = x$, i.e., a neutral element exists, which is $e \in [0,1]$.

On the one hand the practical motivations for the introduction of uninorms were the applications from multicriteria decision making, where the aggregation is one of the key issues. Some alternatives are evaluated from several points of view. Each evaluation is a number from the unit interval. Let the level of satisfaction be $e \in]0,1[$. If all criteria are satisfied to at least *e*-extent then we would like to assign a high aggregated value to this alternative. The opposite of that is if all evaluations are below *e* then we would like to assign a low aggregated value to this alternative. But if there are evaluations below and above *e*, an aggregated value ought to be assigned somewhere in between. Such situations can be modelled by uninorms, leading to the particular classes introduced by [13].

2.1 Distance Based Operators

The distance-based operators can be expressed by means of the min and max operators as follows:

the *maximum distance minimum operator with respect* to $e \in [0,1]$ is defined as

$$\max_e^{\min} = \begin{cases} \max(x, y), & \text{if } y > 2e - x \\ \min(x, y), & \text{if } y < 2e - x \\ \min(x, y), & \text{if } y = 2e\text{-}x \end{cases} \tag{1}$$

the *minimum distance minimum operator* with respect to $e \in [0,1]$ is defined as

$$min_e^{min} = \begin{cases} min(x,y), & \text{if } y > 2e - x \\ max(x,y), & \text{if } y < 2e - x \\ min(x,y), & \text{if } y = 2e\text{-}x \end{cases} \tag{2}$$

the *maximum distance maximum operator with respect to* $e \in [0,1]$ is defined as

$$max_e^{max} = \begin{cases} max(x,y), & \text{if } y > 2e - x \\ min(x,y), & \text{if } y < 2e - x \\ max(x,y), & \text{if } y = 2e\text{-}x \end{cases} \tag{3}$$

the *minimum distance maximum operator with respect to* $e \in [0,1]$ is defined as

$$min_e^{max} = \begin{cases} min(x,y), & \text{if } y > 2e - x \\ max(x,y), & \text{if } y < 2e - x \\ max(x,y), & \text{if } y = 2e\text{-}x \end{cases} \tag{4}$$

The modified distance based operators described above are changed in the boundary condition for neutral element e:

- th*e* maximum distance minimum operator and the the maximum distance maximum operator with respect to $e \in]0,1]$,
- the minimum distance minimum operator and the minimum distance maximum operator with respect to $e \in [0,1[$.

More about distance based operators we can find in [9], [10].

The distance-based operators have the following properties: max_e^{min} *and* max_e^{max} are uninorms, the dual operator of the uninorm max_e^{min} *is* max_{1-e}^{max} ,and the dual operator of the uninorm max_e^{max} *is* max_{1-e}^{min}.

Based on results from [3] we conclude:

Operator $max_{0.5}^{min}$ is a conjunctive left-continuous idempotent uninorm with neutral element $e \in]0,1]$ with the super-involutive decreasing unary operator $g(x) = 2e - x = 2 \cdot 0.5 - x \Rightarrow g(x) = 1 - x$.

Operator $min_{0.5}^{max}$ is a disjunctive right-continuous idempotent uninorm with neutral element $e \in]0,1]$ with the sub-involutive decreasing unary operator $g(x) = 2e - x = 2 \cdot 0.5 - x \Rightarrow g(x) = 1 - x$.

3 Lattice Ordered Monoids and Left Continuous Uninorms and t-norms

Let L be a non-empty set. Lattice is a partially (totally) ordered set which for any two elements $x, y \in L$ also contains their *join* $x \vee y$ (i.e., the least upper bound of the set

$\{x, y\}$), and their *meet* $x \wedge y$ (i.e., the greatest lower bound of the set $\{x, y\}$), denoted by $(L, \preceq)$. Secondly, $(L,*)$ is a semi-group with the neutral element. Following [2] let the following be introduced:

Definition 1.
Let $(L, \preceq)$ be a lattice and $(L,*)$ a semi-group with the neutral element.[6]
The triple $(L,*,\preceq)$ is called a *lattice-ordered monoid* (or an *l-monoid*) if for all $x,y,z \in L$ we have

($LM1$) $x*(y \vee z) = (x*y) \vee (x*z)$ and
($LM2$) $(x \vee y)*z = (x*z) \vee (y*z)$.

An $(L,*,\preceq)$ *l*- monoid is said to be *commutative* if the semi-group $(L,*)$ is commutative.

A commutative $(L,*,\preceq)$ *l*- monoid is said to be *commutative, residuated l-monoid* if there exists a further binary operation $\rightarrow_*$ on *L*, i.e., a function $\rightarrow_*: L^2 \rightarrow L$ (*the * residuum*), such that for all $x,y,z \in L$ we have

$$x*y \preceq z \text{ if and only if } x \preceq (y \rightarrow_* z) \tag{5}$$

An *l*-monoid $(L,*,\preceq)$ is called an *integral* if there is a greatest element in the lattice $(L,\preceq)$ (often called the universal upper bound) which coincides with the neutral element of the semi-group $(L,*)$.

Obviously, each *l*-monoid $(L,*,\preceq)$ is a partially ordered semi-group, and in the case of commutativity the axioms ($LM1$) and ($LM2$) are equivalent.

In the following investigations the focus will be on the lattice $([0,1],\leq)$, we will usually work with a complete lattice, i.e., for each subset *A* of *L* its join $\bigvee A$ and its $\bigwedge A$ exist and are contained in *L*. In this case, *L* always has a greatest element, also called the *universal upper bound.*

Example 1. If we define $*: [0,1]^2 \rightarrow [0,1]$ by

$$x*y = \begin{cases} min(x,y) & if \quad x+y \leq 1 \\ max(x,y) & otherwise \end{cases} \tag{6}$$

then $([0,1],*,\leq)$ is a commutative, residuated *l*-monoid, and the *-residuum is given by

$$x \rightarrow_* y = \begin{cases} max(1-x,y) & if \quad x \leq y \\ min(1-x,y) & otherwise \end{cases} \tag{7}$$

It is not an integral, since the neutral element is 0.5.

The operation * results in an *uninorm*, and special types of distance based operators (see [12] and [6]).

The following result is on important characterization of left-continuous uninorms.

Theorem 1.

For each function $U : [0,1]^2 \to [0,1]$ the following are equivalent:

(*i*) $([0,1], U, \leq)$ is a commutative, residuated *l*-monoid, with a neutral element

(*ii*) *U* is a left continuous uninorm.

In this case the *U*-residuum $\to_U$ is given by

$$x \to_U y = sup\{z \in [0,1] | U(x,z) \leq y\} \tag{8}$$

Proof.

It is easy to see, that $([0,1], U, \leq)$ is a commutative, residuated *l*-monoid with a neutral element if and only if *U* is a uninorm.

Therefore, in order to prove that (*i*)⇒(*ii*), assume that $([0,1], U, \leq)$ is residuated, fix and arbitrary sequence $(x_n)_{n\in N}$ in $[0,1]$ and put $x_0 = \sup_{n\in N} x_n$.

Let $y_0 \in [0,1]$, and $z_0 = \sup_{n\in N} U(x_n, y_0)$.

Obviously $z_0 \leq U(x_0, y_0)$, and (8) implies $(y_0 \to_U z_0) \geq x_n$ for all $n \in N$, subsequently, $(y_0 \to_U z_0) \geq x_0$.

Applying again (8) in the opposite direction, we obtain $U(x_0, y_0) \leq z_0$. Because of the monotonicity of uninorm U, (U3), and based on *Proposition 1.22.* from [6], we have $U(x_0, y_0) = z_0$, i.e., $\sup_{n\in N} U(x_n, y_0) = U\left(\sup_{n\in N}(x_n), y_0\right)$

Conversely, if the uninorm *U* is left-continuous, define the operation $\to_U$ by (8). Then it is clear that for all $x, y, z \in [0,1]$, $x \leq (y \to_U z)$ whenever $U(x, y) \leq z$. The left-continuity of U then implies $U(y \to_U z, y) \leq z$, which together with the monotonicity (U3), ensures that $\to_U$ is indeed the *U*-residuum. ◊

The work of De Baets, B. and Fodor, J. [3] presents general theoretical results related to residual implicators of uninorms, based on residual implicators of t-norms and t-conorms.

Residual operator R_U, considering the uninorm U, can be represented in the following form:

$$R_U(x, y) = sup\{z | z \in [0,1] \wedge U(x,z) \leq y\} \tag{9}$$

Uninorms with the neutral elements $e = 0$ and $e = 1$ are t-norms and t-conorms, respectively, and related residual operators are widely discussed, we also find suitable definitions for uninorms with neutral elements $e \in \,]0,1[$.

If we consider a uninorm *U* with the neutral element $e \in \,]0,1[$, then the binary operator R_U is an implicator if and only if $(\forall z \in \,]e,1[)(U(0,z) = 0)$. Furthermore R_U is an implicator if U is a disjunctive right-continuous idempotent uninorm with unary operator g satisfying $(\forall z \in [0,1])(g(z)=0 \Leftrightarrow z=1)$.

The residual implicator R_U of uninorm U can be denoted by Imp_U.

Consider a uninorm U, then R_U is an implicator in the following cases:

- U is a conjunctive uninorm,
- U is a disjunctive representable uninorm,

U is a disjunctive right-continuous idempotent uninorm with unary operator g satisfying $(\forall z \in [0,1])(g(z)=0 \Leftrightarrow z=1)$.

Theorem 1. implies in a special case Proposition 2.47. from [6]:

Corollary 1

For each function $T : [0,1]^2 \to [0,1]$ the following are equivalent:

(i) $([0,1], T, \leq)$ is a commutative, residuated integral l-monoid,

(ii) T is a left continuous t-norm.

In this case the T-residuum $\to_T$ is given by ($ResT$)

$$x \to_T y = sup\{z \in [0,1] | T(x,z) \leq y\}.$$

Because of its interpretation in [0,1]-valued logics, the T-residuum $\to_T$ is also called *residual implication* (or briefly *R-implication*). [12]

3.1 Residual Implicators of Distance Based Operators

According to Theorem 8. in [3] we introduce implicator of distance based operator $max_{0.5}^{min}$.

Consider the conjunctive left-continuous idempotent uninorm $max_{0.5}^{min}$ with the unary operator $g(x)=1-x$, then its residual implicator $Imp_{max_{0.5}^{min}}$ is given by

$$Imp_{max_{0.5}^{min}} = \begin{cases} max(1-x, y) & if \quad x \leq y \\ min(1-x, y) & elsewhere \end{cases} \tag{10}$$

3.2 Residuum-Based Approximate Reasoning with Distance Based Operator

In many sources it was suggested to represent an “if x is A then y is B” rule simply as a connection (for example as a t-norm, $T(A,B)$ or any conjunctive operator) between the so called rule premise: x is A and rule consequence: y is B. (Let x be from universe X, y from universe Y, and let x and y be linguistic variables. Fuzzy set A on $X \subset R$ finite universe is characterized by its membership function μA: $x \to [0,1]$, and fuzzy set B on Y universe is characterized by its membership function μB: $y \to [0,1]$). If the rule output B' in a fuzzy rule base for one rule if A then B and the system input A' is modeled and calculated by the expression in the form

$$B'(y) = \sup_{x \in X}\left(ConjunctiveOperator\left(A'(x), Imp\left(A(x), B(y)\right)\right)\right) \tag{11}$$

$$B'(y) = \sup_{x \in X}\left(ConjunctiveOperator\left(A'(x), Imp\left(A(x), B(y)\right)\right)\right)$$

Let we consider the uninom residuum-based approximate reasoning and inference mechanism. Hence, and because of the results from above we can consider the general rule consequence for *i*-th rule from a rule system as

$$B_i'(y) = \sup_{x \in X}\left(max_{0.5}^{min}\left(A'(x), Imp_{max_{0.5}^{min}}\left(A_i(x), B_i(y)\right)\right)\right) \tag{12}$$

$$B_i'(y) = \sup_{x \in X}\left(max_{0.5}^{min}\left(A'(x), Imp_{max_{0.5}^{min}}\left(A_i(x), B_i(y)\right)\right)\right),$$

or, using (10)

$$B_i'(y) = \sup_{x \in X}\begin{cases} max_{0.5}^{min}\left(A'(x), max\left(1 - A_i(x), B_i(y)\right)\right) & if \quad A_i(x) \le B_i(y) \\ max_{0.5}^{min}\left(A'(x), min\left(1 - A_i(x), B_i(y)\right)\right) & elsewhere \end{cases} \tag{13}$$

$$B_i'(y) = \sup_{x \in X}\begin{cases} max_{0.5}^{min}\left(A'(x), max\left(1 - A_i(x), B_i(y)\right)\right) & if \quad A_i(x) \le B_i(y) \\ max_{0.5}^{min}\left(A'(x), min\left(1 - A_i(x), B_i(y)\right)\right) & elsewhere \end{cases}$$

Details see in [12].

4 Conclusion

In fact the uninorms offer new possibilities in fuzzy approximate reasoning, because the low level of covering over of rule premise and rule input has measurable influence on rule output as well. The modified, uninorm-based Mamdani's approach, with similarity measures between rule premises and rule input, does not rely on the compositional rule inference any more, but still satisfies the basic conditions supposed for the approximate reasoning for a fuzzy rule base system.[11]

Residuum-based approximate reasoning focused on distance based operators violates needed practical axioms for the rule outputs in a fuzzy logic control system. In the cases when we have normal input *A'* the output is contained in all consequences if we have not "faired" rule. If $A \neq A'$, the rule output belongs not to the convex hull of all rule outputs B_i, ($i=1,n$).

References

1. Bellmann, R.E., Zadeh., L.A., (1977), *Local and fuzzy logic,* Modern Uses of Multiple-Valued Logic, edited by Dunn, J.M., Epstein, G., Reidel Publ., Dordrecht, The Netherlands, pp. 103-165.
2. Birkhoff, G., (1973), *Lattice theory*. American Mathematical Siciety, Providence
3. De Baets, B., Fodor, J., (1999), *Residual operators of uninorms*,, Soft Computing 3., (1999), pp. 89-100.

4. Fodor, J., Rubens, M., (1994), *Fuzzy Preference Modeling and Multi-criteria Decision Support*. Kluwer Academic Pub.,1994.
5. Fodor, J., (1996), *Fuzzy Implications,* Proc. Of International Panel Conference on Soft and Intelligent Computing, Budapest, ISBN 963 420 510 0, pp. 91-98. 21] J., Fodor, B., De Baets, T., Calvo, *Structure of uninorms with given continuous underlying t-norms and t-conorms*, Proc. of the 24th Linz Seminar on Fuzzy Sets, 2003.
6. Klement, E.P., Mesiar, R, Pap, E., '*Triangular Norms*', Kluwer Academic Publishers, 2000, ISBN 0-7923-6416-3
7. E., H., Mamdani, S., Assilian , *An experiment in linguistic syntesis with a fuzzy logic controller,* Intern.,. J. Man-Machine Stud. 7. 1-13., 1975.
8. Pap, E., *Triangular norms in modelling uncertainly, non-linearity and decision*, in Proceedings of the 2th International Symposium of Hungarian researchers Computational Intelligence, ISBN 963 7154 06 X, pp. 7-18.
9. Rudas, I., *Absorbing-Norms,* in Proceedings of the 3th International Symposium of Hungarian Researchers Computational Intelligence, Buadpest, Nov. 2002., pp. 25-43.
10. Takacs, M., Rudas I. J., (1999), *Generalized Mamdani Type Fuzzy Controllers,* in Proceedings of Eurofuse-SIC 99, Budapest, May, 1999. pp. 162-165.
11. Takacs, M., (2003) *Approximate reasoning with Distance-based Operators and degrees of coincidence*, in Principles of Fuzzy Preference Modelling and Decision Making, edited by Bernard de Baets, János Fodor, Academia Press Gent,., ISBN 90-382-0567-8
12. Takacs, M., Approximate Reasoning in Fuzzy Systems Based On Pseudo-Analysis, Phd Thesis, Univ. of Novi Sad, 2004.
13. Yager, R. R., Rybalov, A., (1996), *Uninorm aggregation operators*, Fuzzy sets and systems 80., pp. 111-120.
14. R. R. Yager, (2001), Uninorms in fuzzy system modeling, Fuzzy Ses and Systems, 122, pp:167-175.
15. Zadeh, L. A., *A Theory of approximate reasoning*, In Hayes, J., and editors, Mashine Intelligence, Vol. 9, Halstead Press, New York, 1979., pp. 149-194.
16. Zadeh, L. A., (1999), *From Computing with Numbers to Computing with Words – From Manipulation of Measurements to manipulation of Perceptions,* In Proc. Of EUROFUSE – SIC Conf. 1999., Budapest, pp. 1-3.
17. Zimmermann, H.J., (1991), *Fuzzy Sets, Decision Making and Expert Systems*. Kluwer, Boston, 1991.

Discrete Fuzzy Numbers Defined on a Subset of Natural Numbers*

Jaume Casasnovas and J. Vicente Riera

Dept. Mathematics and Computer Science, Univ. of the Balearic Islands,
07122 Palma de Mallorca, Spain
{jaume.casasnovas,jvicente.riera}@uib.es

Abstract. We introduce an alternative method to approach the addition of discrete fuzzy numbers when the application of the Zadeh's extension principle does not obtain a convex membership function.

1 Introduction

Many authors have studied Fuzzy Numbers and the fuzzy arithmetic operations since many years (see, for instance, [5,6,7]). More recently, the discrete fuzzy numbers, understood as fuzzy numbers with discrete support, are present in papers of several authors [12,13]. In both cases, several definitions and allowed shapes of the membership functions have been considered. For the fuzzy numbers, in this paper we will consider convex membership functions with an interval as "support", LR-definitions and a "core" that is also a subinterval of the support.

The discrete fuzzy numbers whose support is a subset of natural numbers arise mainly when a fuzzy cardinality of a fuzzy set [4] or a fuzzy multiset [3,8,14] is considered. In both cases, the membership function of the involved discrete fuzzy numbers is decreasing. However, we can consider a wider kind of discrete fuzzy numbers in order to implement generalizations of the multiset concept [1,8,9,10,11]. So, a discrete fuzzy number $\tilde{5}$ means "about 5" occurrences of x_1 in M but with the constraint: "it is a natural number". Even, we can suppose to have an information like: "about 5 occurrences, but they are not 4".

In general, the arithmetic operations on fuzzy numbers can be approached either by the direct use of the membership function(by the Zadeh's extension principle) or by the equivalent use of the *r-cuts* representation. Nevertheless, in the discrete case, this process can yield a fuzzy subset that does not satisfy the conditions to be a discrete fuzzy number [2,13].

In a previous work [2] we have presented an approach to a closed addition of discrete fuzzy numbers after associating suitable non-discrete fuzzy numbers, which can be used like a carrier to obtain the desired addition. In this paper we prove that a suitable carrier can be a discrete fuzzy number whose support is an arithmetic sequence and even a subset of consecutive natural numbers.

---

* This work has been partially supported by Spanish Government grants of the DGES, MTM2006-08322.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 573–582, 2007.
springerlink.com

2 Preliminaries

Definition 1 [7]. A fuzzy subset u of $\mathbb{R}$ with membership mapping $u:\mathbb{R}\to[0,1]$ is called *fuzzy number* if its support is an interval $[a,b]$ and there exist real numbers s,t with $a\le s\le t\le b$ and such that:

1. $u(x_i)=1$ with $s\le x\le t$
2. $u(x)\le u(y)$ with $a\le x\le y\le s$
3. $u(x)\ge u(y)$ with $t\le x\le y\le b$
4. $u(x)$ is upper semi-continuous.

We will denote the set of fuzzy numbers by FN.

Definition 2 [12]. A fuzzy subset u of $\mathbb{R}$ with membership mapping $u:\mathbb{R}\to[0,1]$ is called *discrete fuzzy number* if its support is finite, i.e., there are $x_1,...,x_n\in\mathbb{R}$ with $x_1<x_2<...<x_n$ such that $supp(u)=\{x_1,...,x_n\}$, and there are natural numbers s,t with $1\le s\le t\le n$ such that:

1. u(x_i)=1 for any natural number and i with $s\le i\le t$ (core)
2. $u(x_i)\le u(x_j)$ for each natural number i,j with $1\le i\le j\le s$
3. $u(x_i)\ge u(x_j)$ for each natural number i,j with $t\le i\le j\le n$

From now on, we will denote the set of discrete fuzzy numbers by DFN and a discrete fuzzy number by a d.f.n.

2.1 Addition of Fuzzy Numbers [5,6,7]

Given two fuzzy numbers $u,v\in FN$, according to the Zadeh's extension principle, their addition is the fuzzy number $u\oplus v$, pointwise defined as follows:

$$(u\oplus v)(z)=\sup_{z=x+y}(\min(u(x),v(y))),\forall z\in\mathbb{R}$$

Equivalently, their addition is the fuzzy number whose r-cuts are defined:

$$[u\oplus v]^r=[u]^r+[v]^r,\forall r\in(0,1]$$

and, consequently, their membership function is:

$$(u\oplus v)(z)=\sup\left\{r\in[0,1]\text{ such that } z\in[u\oplus v]^r\right\}$$

2.2 Addition of Discrete Fuzzy Numbers

Let $u,v\in DFN$ be two discrete fuzzy numbers. If we consider them as fuzzy subsets of $\mathbb{R}$ we can apply the extension principle to obtain their extended sum. But we can see in [2,13] that it is possible a result which does not fulfill the conditions stated above in the Definition 2. In order to overcome this drawback several authors [2,13] have proposed modifications of the process:

- [13] By means of the r-cuts for one of d.f.n. $u, v \in DFN$, the author defines the r-cuts for a new d.f.n., which will be denoted by $u \bigoplus_W v$, such that $[u \bigoplus_W v]^r = \{x \in supp(u) + supp(v) : \min([u]^r + [v]^r) \leq x \leq \max([u]^r + [v]^r)\}$ where $min([u]^r + [v]^r) = min\{x + y : x \in [u]^r, y \in [v]^r\}$, $max([u]^r + [v]^r) = max\{x + y : x \in [u]^r, y \in [v]^r\}$ and $(u \bigoplus_W v)(x) = \sup\{r \in [0,1] \text{ such that } x \in [u \bigoplus_W v]^r\}$. Moreover, this author proves that for each $u, v \in DFN$, if $u \oplus v \in DFN$ (obtained through the Zadeh's extension principle) then $u \oplus v = u \bigoplus_W v$.
- [2] The authors propose a method that can be seen as a generalization of Wang's one and it is based on the concept of association of a fuzzy number to a discrete fuzzy number, understanding by association a mapping $A : DFN \to FN$ that fulfills determined conditions. Thus, from an association A, for each couple $u, v \in DFN$, we define the discrete fuzzy number $u \bigoplus_A v$, such that:

$$supp(u \bigoplus_A v) = \left\{ \begin{matrix} z \in \mathbb{R} \text{ such that } z = x + y, \\ x \in supp(u), y \in supp(v) \end{matrix} \right\}$$

$$(u \bigoplus_A v)(z) = (A(u) \oplus A(v))(z) \ \forall \ z \in supp(u \bigoplus_A v)$$

where $A(u) \oplus A(v)$ represents the addition of the fuzzy numbers $A(u)$ and $A(v)$ defined in the subsection 2.1 .

3 Addition of d.f.n. Whose Support Is an Arithmetic Sequence of Natural Numbers

Let's call $DFN(\mathbb{N})$ to the set of d.f.n. whose support is a subset of the set of Natural Numbers.

Let $\mathcal{A}_r$ be the set { $f \in DFN(\mathbb{N})$, such that $supp(f)$ is the set of terms of an arithmetic sequence with r as common difference}.

Proposition 3. If $f, g \in \mathcal{A}_r$ .The following facts:

1. $f \oplus g \in DFN(\mathbb{N})$
2. $f \oplus g \in \mathcal{A}_r$

hold.

Proof: Let f, g two d.f.n. belonging to $\mathcal{A}_r$ such that:

$$supp(f) = \{a, a+r, a+2\cdot r, ..., a+p\cdot r\} \text{ and } core(f) = \{a+l\cdot r, ..., a+j\cdot r\}$$
$$supp(g) = \{b, b+r, b+2\cdot r, ..., b+q\cdot r\} \text{ and } core(g) = \{b+m\cdot r, ..., b+n\cdot r\}$$

It is easy to see that:

$$supp(f \oplus g) = \{a+b+i\cdot r, \text{ where } i = 0, ..., p+q\} \tag{1}$$

$$core(f \oplus g) = \{a+b+k\cdot r, \text{ where } k = l+m, ..., j+n\} \tag{2}$$

If we call $\tilde{i} = a+b+i\cdot r$ where $i \in \{0, 1, ..., p+q\}$ then:

$$(f \oplus g)(\tilde{i}) = \sup_{\substack{x+y=i \\ x\in\{0,...,p\} \\ y\in\{0,...,q\}}} (f(a+x\cdot r) \wedge g(b+y\cdot r)) = f(a+x^i\cdot r) \wedge g(b+y^i\cdot r) \tag{3}$$

with $x^i \in \{0, ..., p\}$, $y^i \in \{0, ..., q\}$, $x^i + y^i = i$
Without lack of generality, we can suppose that:

$$(f \oplus g)(\tilde{i}) = f(a+x^i\cdot r) \tag{4}$$

By means of a tedious but easy study of all possible cases for x^i and y^i we can prove that:

1. If $\tilde{i} \in \{a+b, ..., a+b+(l+m-1)\cdot r\}$ then $(f \oplus g)(\tilde{i}) \le (f \oplus g)(\widetilde{i+1})$
2. If $\tilde{i} \in \{a+b+(j+n+1)\cdot r, ..., a+b+(p+q)\cdot r\}$ then $(f \oplus g)(\tilde{i}) \ge (f \oplus g)(\widetilde{i+1})$
3. If $\tilde{i} \in \{a+b+(l+m)\cdot r, ..., a+b+(j+n)\cdot r\}$ then $(f \oplus g)(\tilde{i}) = 1$

Therefore, $f \oplus g \in DFN(\mathbb{N})$. On the other hand, if we consider the previous relations (1) and (2) we will obtain $f \oplus g \in \mathcal{A}_r$

Corollary 1. Let $\mathcal{C}$ be the set $\{ f \in DFN(\mathbb{N})$; such that for all $i, j \in supp(f)$, if $i \le k \le j$, with $k \in \mathbb{N}$ then $k \in supp(f) \}$.
If $f, g \in \mathcal{C}$, then $f \oplus g \in \mathcal{C}$.

Proof: It is a direct consequence of the previous Proposition, because $\mathcal{C} = \mathcal{A}_1$.

3.1 Remarks

1. The previous proposition does not hold if the supports of the considered d.f.n. f and g, as arithmetic progressions, do not have the same common difference. For instance, let $u = \{0.2/1, 0.4/3, 1/5, 0.7/7, 0.5/9\}$ and $v = \{0.2/6, 1/10, 0.8/14, 0.7/18\}$ be two d.f.n.. In both cases the points in their support form an arithmetic sequence with a common difference, namely 2 and 4 respectively. If we apply the extension

principle to calculate their sum, we obtain the fuzzy subset $u \oplus v = \{0.2/7, 0.2/9, 0.2/11, 0.4/13, 1/15, 0.7/17, 0.8/19, 0.7/21, 0.7/23, 0.7/25, 0.5/27\}$ that does not belong to DFN, because $(u \oplus v)(17) = 0.7 < (u \oplus v)(19) = 0,8$ and so for this reason the condition 3 of definition 2 doesn't hold .

2. The previous proposition cannot be generalized to the set:

   $\mathcal{G}_r = \{ f \in DFN(\mathbb{N})$ such that $supp(f)$ is a geometric sequence with common ratio r }

   For instance, if $u = \{0.5/1, 1/2, 0.7/4\}$ and $v = \{0.8/8, 1/16, 0.7/32\}$.

   Then $u, v \in$, $\mathcal{G}_r$ but the fuzzy subset

   $u \oplus v = \{0.5/9, 0.8/10, 0.7/12, 0.5/17, 1/18, 0.7/20, 0.5/33, 0.7/34, 0.7/36\}$ that does not belong to DFN, because $(u \oplus v)(10) = 0.8 > (u \oplus v)(12) = 0,7$

   and so for this reason the condition 2 of definition 2 doesn't hold.

4 Addition of Discrete Fuzzy Numbers Whose Support Is Any Subset of the Natural Numbers

The two examples included in the above remark and the following one show that the extension principle applied to a couple of discrete fuzzy numbers, whose support is any subset of the natural numbers, does not determine in general a d.f.n.

Let's consider $u = \{0.3/1, 1/2, 0.5/3\}$ and $v = \{0.4/4, 1/6, 0.8/8\}$. Then: $u \oplus v = \{0.3/5, 0.4/6, 0.4/7, 1/8, 0.5/9, 0.8/10, 0.5/11\}$ which does not satisfy the conditions stated in Definition 2.

In order to overcome this drawback, we will associate a d.f.n. belonging to the set $\mathcal{C}$ to each d.f.n..

This association will be a carrier to obtain an approach to the addition of two d.f.n. that will be a d.f.n.

Definition 4. Let $u \in DFN(\mathbb{N})$ be a discrete fuzzy number such that its support is $supp(f) = \{x_1, ..., x_s, ..., x_t, ..., x_n\}$ with $x_1 < \ldots < x_s < \ldots < x_t < \ldots < x_n$ and $u(x_p) = 1$ for all natural number p, $s \le p \le t$.

We will associate to u any d.f.n. $C(u) \in \mathcal{C}$ (defined in Corollary 1) fulfilling the following properties:

1. If $x_i \in supp(u)$ then $C(u)(x_i) = u(x_i)$ for each $i = 1, \ldots, n$
2. $u(x_i) \le C(u)(x) \le u(x_{i+1})$, $\forall x \in [x_i, x_{i+1}]$ with $1 \le i \le i+1 \le s$.
3. $C(u)(x_i) = 1$, $\forall x \in [x_i, x_{i+1}]$ with $s \le i \le i+1 \le t$.
4. $u(x_i) \ge C(u)(x_i) \ge u(x_{i+1})$, $\forall x \in [x_i, x_{i+1}]$ with $t \le i \le i+1 \le n$.

4.1 Examples

If $u \in DFN(\mathbb{N})$ we will call $\alpha-$association, to the d.f.n. $C_\alpha(u) \in \mathcal{C}$ defined in the following way

$$C_\alpha(u)(x) = \begin{cases} u(x_i) & \text{if} \quad x \in [x_i, x_{i+1}), \text{with } x_{i+1} < x_s \\ 1 & \text{if} \quad x \in [x_s, x_t] \\ u(x_{i+1}) & \text{if} \quad x \in (x_i, x_{i+1}], \text{with } x_i > x_t \end{cases}$$

For instance:

If $v = \{0.4/2, 1/5, 1/6, 0.8/9\}$. Its $\alpha-$associated will be the following discrete fuzzy number

$$C_\alpha(v)(x) = \begin{cases} 0.4 & \text{if} \quad x \in \{2,3,4\} \\ 1 & \text{if} \quad x \in \{5,6\} \\ 0.8 & \text{if} \quad x \in \{7,8,9\} \end{cases}$$

If $u = \{0.3/1, 1/3, 0.5/7\}$ its $\alpha-$associated will be the following discrete fuzzy number

$$C_\alpha(u)(x) = \begin{cases} 0.3 & \text{if} \quad x \in \{1,2\} \\ 1 & \text{if} \quad x \in \{3\} \\ 0.5 & \text{if} \quad x \in \{4,5,6,7\} \end{cases}$$

If $u \in DFN(\mathbb{N})$ we will call $\omega-$association, to the d.f.n. $C_\omega(u) \in \mathcal{C}$ defined in the following way:

$$C_\omega(u)(x) = \begin{cases} u(x) & \text{if} \quad x \in supp(u) \\ u(x_{i+1}) & \text{if} \quad x \in (x_i, x_{i+1}), \text{with } x_{i+1} \le x_s \\ 1 & \text{if} \quad x \in (x_s, x_t) \\ u(x_i) & \text{if} \quad x \in (x_i, x_{i+1}), \text{with } x_i \ge x_t \end{cases}$$

For instance:

If $v = \{0.4/2, 0.6/4, 1/5, 1/6, 0.9/7, 0.8/9\}$ its $\omega-$associated will be the d.f.n. defined in the following way: $C_\omega(v)(x) = \begin{cases} 0.4 & \text{if} \quad x \in \{2\} \\ 0.6 & \text{if} \quad x \in \{3,4\} \\ 1 & \text{if} \quad x \in \{5,6\} \\ 0.9 & \text{if} \quad x \in \{7,8\} \\ 0.8 & \text{if} \quad x \in \{9\} \end{cases}$

If $u \in DFN(\mathbb{N})$ we will call linear-association, to the d.f.n. $C_l(u) \in \mathcal{C}$ defined in the following way: $C_l(x) = \begin{cases} u(x) & \text{if} \quad x \in supp(u) \\ \frac{u(x_{i+1}) - u(x_i)}{x_{i+1} - x_i}(x - x_i) + u(x_i) & \text{if} \quad x \in (x_i, x_{i+1}) \end{cases}$

It is straightforward to prove that C_α, C_ω and C_l satisfy the properties of an association as have been established in Definition 4.

4.2 Remark

If $C : DFN(\mathbb{N}) \to \mathcal{C}$ is an association, then the relation $C_\alpha(u) \le C(u) \le C_\omega(u)$ holds for all $u \in DFN(\mathbb{N})$.

Proof :
Straightforward.

Definition 5. Let $u, v \in DFN(\mathbb{N})$ be a couple of d.f.n. and let $C : DFN(\mathbb{N}) \to \mathcal{C}$ be an association. Let's consider the fuzzy subset, that will be denoted by $u \bigoplus_C v$, defined as follows:

$$supp(u \bigoplus_C v) = \left\{ \begin{matrix} z \in \mathbb{N} \text{ such that } z = x + y, \\ x \in supp(u), y \in supp(v) \end{matrix} \right\}$$

$$(u \bigoplus_C v)(z) = (C(u) \oplus C(v))(z), \forall z \in supp(u \bigoplus_C v)$$

Then $u \bigoplus_C v \in DFN(\mathbb{N})$ and we will call it the C-addition of the couple u, v of discrete fuzzy numbers.

4.3 Remark

We will denote the C-addition by $u\bigoplus_{\alpha} v$ when the association C is the α-association C_{α} defined in 4.1 and $u\bigoplus_{\omega} v$ when the association is the ω-association C_{ω} also defined in 4.1.

4.4 Examples

Let $u = \{0.3/1, 1/3, 0.5/7\}$ and $v = \{0.4/2, 1/5, 1/6, 0.8/9\}$ be two d.f.n.. We have:

- According to the extension principle: $u \oplus v = \{0.3/3, 0.4/5, 0.3/6, 0.3/7, 1/8, 1/9, 0.3/10, 0.8/12, 0.5/13, 0.5/16\}$ and $u \oplus v \notin DFN$
- If we consider the α-associations which have been calculated in 4.1, then: $u\bigoplus_{\alpha} v = \{0.3/3, 0.4/5, 0.4/6, 0.4/7, 1/8, 1/9, 0.8/10, 0.8/12, 0.5/13, 0.5/16\}$
- If we consider the ω-associations:
  $C_{\omega}(u) = \{0.3/1, 1/2, 1/3, 1/4, 1/5, 1/6, 0.5/7\}$ and
  $C_{\omega}(v) = \{0.4/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 0.8/9\}$ then:
  $u\bigoplus_{\omega} v = \{0.3/3, 1/5, 1/6, 1/7, 1/8, 1/9, 1/10, 1/12, 1/13, 0.5/16\}$

Proposition 6. If $u, v \in DFN(\mathbb{N})$ are two d.f.n.. Then, $u\bigoplus_{W} v = u\bigoplus_{\alpha} v$ where $u\bigoplus_{W} v$ has been defined in 2.2 and $u\bigoplus_{\alpha} v$ has been defined in 4.3.

Proof: We will see that $\left[u\bigoplus_{W} v\right]^r = \left[u\bigoplus_{\alpha} v\right]^r, \forall r \in [0,1]$

Since the definition of $C_{\alpha}(u)$ and $C_{\alpha}(v)$, for each $r \in [0,1]$:

a) $[u]^r \subseteq [u_{\alpha}]^r$ and $[v]^r \subseteq [v_{\alpha}]^r$

b) $\min[u]^r = \min[u_{\alpha}]^r$ and $\min[v]^r = \min[v_{\alpha}]^r$

c) $\max[u]^r = \max[u_{\alpha}]^r$ and $\max[v]^r = \max[v_{\alpha}]^r$

d) Besides, from the definition of $u\bigoplus_{\alpha} v$, it can be deduced that:

$$\left[u\bigoplus_{\alpha} v\right]^r \subseteq \left[u_{\alpha} \oplus v_{\alpha}\right]^r$$

e) Since 2.2: $x \in \left[u \bigoplus_W v \right]^r \Leftrightarrow \begin{cases} x \in \operatorname{supp}\left(u \bigoplus_W v \right) = \operatorname{supp}(u) + \operatorname{supp}(v) \\ \min([u]^r + [v]^r) \leq x \leq \max([u]^r + [v]^r) \end{cases}$

Taking into account a), b), c), d), e), the proof can be completed like in [2], where an analogous proposition for non-discrete associations is proved.

5 Conclusion

This paper has shown an alternative method to approach the addition of discrete fuzzy numbers when the application of the Zadeh's extension principle does not obtain a d.f.n. Also, we have seen that in the particular case of addition of d.f.n., whose support is a set of terms of an arithmetic sequence with r as common difference, the Zadeh's extension principle yields a d.f.n.. But, this is not true when the d.f.n. have as support either a set of terms of a geometric sequence with common ratio r or a set of terms of arithmetic sequences with different differences.

References

1. W.D.Blizard, (1991) The development of multiset theory, Modern logic, pp. 319–352.
2. J.Casasnovas, J. Vicente Riera, (2006) On the addition of discrete fuzzy numbers", Wseas Transactions on Mathematics, Issue 5, Volume 5, pp. 549–554.
3. J. Casasnovas, F. Rosselló, "Scalar and fuzzy cardinalities of crisp and fuzzy multisets," Accepted in International Journal of Intelligent Systems.
4. J. Casasnovas, J.Torrens, (2003), An Axiomatic Approach to the fuzzy cardinality of finite fuzzy sets. Fuzzy Sets and Systems 133, 193–209.
5. D.Dubois, H.Prade, (1980) Fuzzy Sets and Systems: Theory and Applications. Academic Press. New York
6. D.Dubois, H.Prade, (2000) (Eds) Fundamentals of Fuzzy Sets. The handbooks of Fuzzy Sets Series, Kluwer, Boston.
7. Klir, George J., Yuan Bo ,(1995), Fuzzy sets and fuzzy logic (Theory and applications), Prentice Hall
8. S. Miyamoto, (2004), Generalizations of Multisets and Rough Approximations, International Journal of intel.Systems, Vol. 19, pp. 639–652.
9. S. Miyamoto, (2004), Multisets and Fuzzy Multisets as a Framework of Information Systems, Lecture Notes in Computer Science,3131, pp. 27-40
10. S. Miyamoto, (2001), Fuzzy Multisets and Their Generalizations, Multisets Processing LNCS 2235, pp. 225–235.
11. S. Miyamoto, (1996), Fuzzy multisets and applications to rough approximation of fuzzysets, in Proc. Fourth Inter. Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery, RSFD'96 pp. 255–260.

12. Voxman, W., (2001), Canonical representations of discrete fuzzy numbers, Fuzzy Sets and Systems 54 ,pp. 457–466.
13. Guixiang Wang,Cong Wu, Chunhui Zhao, (2005), Representation and Operations of discrete fuzzy numbers , Southeast Asian Bulletin of Mathematics,28:1003–1010.
14. R. Yager, (1986), On the theory of bags, Inter. J. General Systems 13, pp. 23–37.

Collaborative Recommending Based on Core-Concept Lattice

Kai Li, YaJun Du, and Dan Xiang

School of Mathematical and Computers Science, Xihua University, Chengdu, Sichuan, 610039, China
likai0148@hotmail.com,duyajun@xhu.edu.cn,jintianxd@163.com

Abstract. In this paper, two new notions called core-concept and core-concept lattice are proposed and applied to collaborative recommendation system. The core-concept lattice is constructed based on the core-concept, which is extracted from rating matrix between users and items in collaborative recommendation systems. Compared with traditional FCA, it is obviously that the extraction of core-concept very easy and fast. We present the improved nearest neighbors algorithm, it use core-concept lattice as an index to the recommendation's ratings matrix. The improved nearest neighbors algorithm could remarkably accelerate finding the nearest neighbors. Therefore, it could evidently improve efficiency of recommendation.

1 Introduction

As the continuously developing of internet technology, internet has become an important tool which be used to retrieve information. But the rapidly expanded World Wide Web also causes the overloading problem of information. So, how to help users find the information efficiently and conveniently has become the concerned research.

Recommendation system is a special class of personalized systems that aim at predicting a user's interest on available products and services by relying on previously rated items or item attributes [1]. Current common approaches for personalized recommendation systems are the content-based filtering (CBF) and collaborative filtering (CF) [2, 3, 4, 5].

Content-based filtering makes predictions upon the assumption that a user's previous preferences or interests are reliable indicators for his/her future behavior. CBF requires that items are described by attributes, and is typically applied upon text-based documents, or in domains with structured data [6, 7].

On the other hand, Collaborative filtering operates upon the assumption that if a user A and B rates some items similarly, they share similar tastes and hence will rate other items similarly. Collaborative filtering is applicable to any type of content [6], while it can also capture concepts that are hard to represent, such as quality and taste [8]. Additionally, collaborative filtering does not restrict the spectrum of recommendations to items similar to the ones that the user has previously evaluated. Collaborative filtering has been acknowledged as the most successful and most widely implemented recommendation technique to date [9, 10]. For these reasons, we will focus on the collaborative filtering strategy in this paper.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 583–592, 2007.
springerlink.com

Collaborative filtering approaches can be distinguished into two major classes: model-based and memory-based [3, 11]. Model-based methods develop a model, which is applied upon the target user's ratings to make predictions for unobserved items.

In contrast to model-based, memory-based methods operate upon the entire database of users to find the closest neighbors of the target user and weight their recommendation according to their similarities. The fundamental algorithm of the memory-based class is the nearest neighbors (denoted as NN, hereafter); it can be described as a process divided in three steps as follows, for more details, sees [1, 12]:

1. Measurement of similarities between the target and the remaining users. A typical measure of similarity is the Pearson correlation coefficient.
2. Selection of the neighbors who will serve as recommenders.
3. Prediction based on the weighted average of the neighbors' ratings, weighted by their similarity to the target user.

The efficiency of computing similarity between all users in huge data must very low. For solving low efficiency of finding the nearest neighbors, the method of Formal Concept Analysis (FCA) was put forward in [13]. It regard rating matrix as formal context, based context, concept lattice was established as index, therefore accelerate finding the nearest neighbors. Experiment in [13] has proved feasibility of this method, but extract formal concept in formal context and establish concept lattice are also time-consuming.

Based above questions, aiming at collaborative recommending systems, we present core-concept and core-concept lattice in this paper. It obviously that extraction of core-concept is easier and faster than formal concept. Then we can do the work of collaborative recommendation on the basis of this method, and solve the low efficiency of finding the nearest neighbors ultimately.

We introduce relevant knowledge of FCA and present the approach of constructing core-concept lattice in section 2. Then, in section 3 we present how to apply core-concept lattice to collaborative recommending systems and propose the improved nearest neighbors algorithm. We conclude in section 4 with a look into the future.

2 Formal Concept Analysis

Formal Concept Analysis (FCA) is a mathematical method for analyzing binary relations, it's a power tool which used to analyze data and extract knowledge from formal context by concept lattice. 1982, concept lattice was first introduced by Wille [14], it established on the basis of FCA in theory. In FCA, data are structured into formal concepts, which form a concept lattice, ordered by a subconcept–superconcept relation. At present, FCA has been extensively applied in several areas such as knowledge discovery [15], software engineering [16] and case-based reasoning [17].

2.1 Formal Context and Formal Concept

First, we recall some basic notions of FCA. The definitions and theorems in this subsection are quoted from [14, 18, 19, 20, 21].

Definition 1. A formal context is a triple K: = (G, M, I) where G and M are sets and I $\subseteq G \times M$ is a binary relation. The elements of G are called objects and the elements of M are called attributes. The inclusion (g, m) $\in$ I is read "object g has attribute m". For A $\subseteq$ G, we define

$$A' := \{m \in M \mid \forall g \in A: (g, m) \in I\};$$

and for B $\subseteq$ M, we define dually

$$B' := \{g \in G \mid \forall m \in B: (g, m) \in I\};$$

In this paper, we assume that all sets are finite, especially G and M.

Definition 2. A formal concept is a pair (*A, B*) with $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. (This is equivalent to $A \subseteq G$ and $B \subseteq M$ being maximal with $A \times B \subseteq I$.) *A* is called extent and *B* is called intent of the concept.

Definition 3. The set $\underline{\mathfrak{B}}(\mathbb{K})$ of all concepts of a formal context $\mathbb{K}$ together with the partial order $(A_1, B_1) \leq (A_2, B_2) :\Leftrightarrow A_1 \subseteq A_2$ (which is equivalent to $B_1 \supseteq B_2$) is called concept lattice of $\mathbb{K}$.

The fundamental theorem of FCA [14] shows that each concept lattice is a complete lattice, and that the set of its intents is a closure system [18].

Theorem 1. (Fundamental Theorem of FCA). Let $\mathbb{K}$: = (G, M, I) be a formal context. Then $\underline{\mathfrak{B}}(\mathbb{K})$ is a complete lattice in which infima and suprema can be described as follows:

$$\bigwedge_{j\in J}(A_j, B_j) = \left(\bigcap_{j\in J} A_j, \left(\bigcup_{j\in J} B_j\right)''\right), \quad \bigvee_{j\in J}(A_j, B_j) = \left(\left(\bigcup_{j\in J} A_j\right)'', \bigcap_{j\in J} B_j\right)$$

Table 1. A binary formal context

	a	b	c	d	e	f
1	×	×				×
2	×	×			×	
3	×	×	×			
4	×		×			×
5	×	×		×	×	×
6	×	×	×	×		×

Definition 4. if (A,B) and (C,D) are two concepts of $\mathbb{K}$, (A,B)≤(C,D) iff $A \subseteq C$ (or, equivalently, $D \subseteq B$). (A, B) is called sub-concept of (C, D), and (C, D) is called super-concept of (A, B).

Table 1 describes a binary formal context. *G*={1, 2, 3, 4, 5, 6}, *M*={a, b, c, d, e, f }, *I* depicts objects in *G* have attributes in *M*.

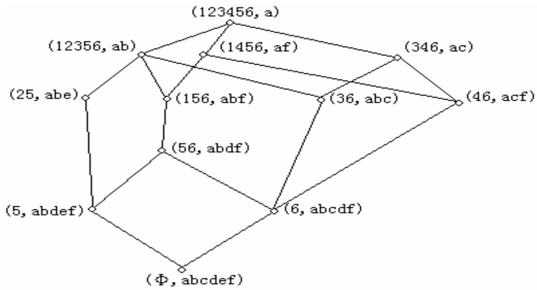


Fig. 1. The concept lattice that corresponds to the formal context in Table 1

2.2 Core-Concept and Core-Concept Lattice

In this subsection, we put forward definition of core-concept and core-concept lattice at first, then explain how to extract core-concepts and construct core-concept lattice.

Definition 5. A core-concept is a pair (A, B) with A$\subseteq$G, B$\subseteq$M. $\forall g \in G$, let E={g}, then $B = E'$ and $A = E''$. A is called extent and B is called intent of the core-concept.

Definition 6. The set $\Re(K)$ of all core-concepts of a formal context K together with the partial order (A1, B1) $\leq$ (A2, B2): $\Leftrightarrow$ A1$\subseteq$A2 (which is equivalent to B1$\supseteq$B2) is called core-concept lattice of K. Then $\Re(K)$ is a complete lattice after adding infima and suprema.

According to Definition 5, we could fast process the objects that correspond to the formal context in Table 1 one by one. The core-concept extracted from Table 1's formal context and the corresponding core-concept lattice is shown as follows:

(156, abf); (25, abe); (36, abc); (46, acf); (5, abdef); (6, abcdf).

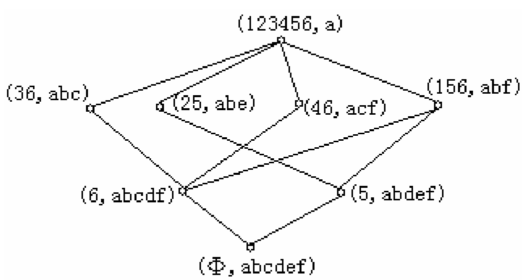


Fig. 2. The core-concept lattice that corresponds to the context in Table 1

The concise format [13, 21] of Fig. 2's core-concept lattice is shown in Fig. 3. In this lattice, the node labeled object denote a core-concept. Its extent is read from the descendants, and intent is read from the ancestors.

The number of core-concept is not better than the number of objects. It obviously that core-concept is formal concept all the same, but all core-concepts just a subset of whole formal concept.

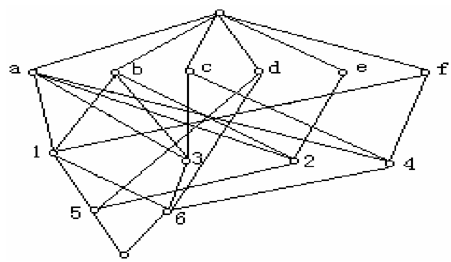


Fig. 3. The concise format core-concept lattice

3 Applying Core-Concept Lattice to Collaborative Recommendation

We apply core-concept lattice to collaborative recommendation for accelerating search for the nearest neighbors of target user. Aiming at collaborative recommendation, we propose the concept of core-concept and core-concept lattice. The core-concept lattice can then act as an index [13] to the ratings matrix to speed up the search for neighbors. Compared with traditional method of FCA, the all core-concepts just a subset of whole formal concept, however, based on Definition 5 and 6, each object must have a corresponding node in the concise format core-concept lattice, so it can completely predict and recommend for all users.

3.1 Obtaining Core-Concept and Core-Concept Lattice in Collaborative Recommendation

Core-concept requires a formal context, i.e., a binary relation between objects and attributes. An example of a ratings matrix for music is shown as Table 2. We take users to be objects and items to be attributes, the rating matrix shown in Table 2 correspond to a multi-value formal context.

Literature [13] states a Hypothesis: Users who rate the same items tend to rate items the same. For simplifying the computation and basing on the Hypothesis [13], we produce a formal context from the ratings matrix: a cell contains × iff $r_{u,i} \neq$^. The formal context that corresponds to Table 2 ratings matrix is shown as a cross-table in Table 3.

Table 2. A rating matrix for music

	Angel	Vincent	Incomplete	Ghetto
John	5	^	^	2
David	^	^	3	4
Nick	2	5	^	2
Anna	2	5	4	2
Christina	^	4	^	2
Billy	4	3	4	^
Jordan	4	5	^	3

Table 3. The context that corresponds to the rating matrix in Table 2

	Angel	Vincent	Incomplete	Ghetto
John	×			×
David			×	×
Nick	×	×		×
Anna	×	×	×	×
Christina		×		×
Billy	×	×	×	
Jordan	×	×		×

According to the section 2.2, the core-concept extracted from Table 3's formal context and the concise format core-concept lattice is shown as follows:

{Anna, (Incomplete, Ghetto, Angel, Vincent)};
{(Billy, Anna), (Incomplete, Angel, Vincent)};
{(Nick, Jordan, Anna), (Ghetto, Angel, Vincent)};
{(David, Anna), (Incomplete, Ghetto)};
{(John, Nick, Jordan, Anna), (Ghetto, Angel)};
{(Christina, Nick, Jordan, Anna), (Ghetto, Vincent)}.

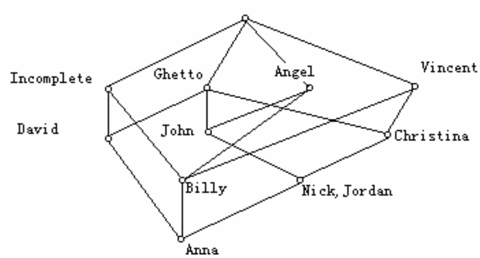


Fig. 4. The concise format core-concept lattice

3.2 Using Core-Concept Lattice to Find the Nearest Neighbors

Here we introduce how to use core-concept lattice to improve NN. Before computing the similarities between the target and the remaining users, we use core-concept lattice to reduce the number of remaining users. The improved nearest neighbors algorithm (denoted as INN, hereafter) shows as follows:

```
neighbor: The set of users who have not zero similarity with target
user.
nearest_neighbor: The set of users who will recommend to the target
user.
target_user.subset: The set of users who is sub-node in core-concept
lattice of the target user.
target_user.superset: The set of users who is super-node in core-concept
lattice of the target user.
target_user.accompanier: The set of users who share the same node in
core-concept lattice with the target user.
N: The number of nearest neighbors.
m: The number of all users.
Begin
  Input: target_user;
  Output: nearest_neighbor;
  neighbor={};
  nearest_neighbor={};
  for each user in target_user.subset do
  begin
    X ← user;
    X. similarity ← similarity (target_user,X);
    neighbor ← X;
  end;
  for i:1 → m do
  begin
    if ( remain_user(i) in neighbor ) or (remain_user (i)
       in target_user.superset) or (remain_user (i) in
              target_user.accompanier) then
      continue;
    else
    if has_parent(remain_user (i), target_user) then
    begin
      X ← remain_user (i);
      X. similarity ← similarity (target_user,X);
      neighbor ← X;
    end;
  end;
nearest_neighbor=Select(neighbor, N);
end.
```

Given the target user, INN first walks core-concept lattice to find the neighbors who could recommend. It gets rid of the users who have no effect on the target user, but doesn't debase the accuracy and coverage.

There are conclusions about who could or not recommend to the target user. By using INN, it can visit the users' node of core-concept lattice in a most-likely order. Here we introduce how to find the candidate nearest neighbors, and give an example of recommending to Nick in the core-concept lattice depicted in Fig. 4:

a) If users shares one and the same node with target user, they have no additional ratings and so can't be the nearest neighbors. For Nick, according to Fig. 4, Jordan can't recommend new music to him.

b) If users' node is the super-node of target user's node, the items they has rated is a subset of the items which target user has rated, so they can't recommend to the target user. In Fig. 4, Christina and John can't be used to make recommendation for Nick.
c) If users haven't a common super-node with the target user, they have no co-rated items with target user, so they can't recommend to the target user.
d) The sub-nod should firstly considered, because may be they have the most similarity with target user. Thus Anna should take into account above all.
e) The node used to recommend should considered from the lowest level up, because the lower level the users locate, the more similarity the users and target user have.

Firstly, INN could remarkably reduce the numbers of the candidate user. Then it just need to compute similarities between the target user and a very few remaining users to find the nearest neighbors. The nearest neighbors' rating can then be used either to make predictions in the case where an objective item is also supplied, or to recommend items rated by the neighbors but not yet rated by the target user.

Based on the core-concept lattice, INN distinctly speeds up the search for the nearest neighbors. INN does considerably less work than NN, but both guarantee accuracy and coverage results equal to NN.

4 Conclusions

In this paper, aiming at collaborative recommendation system, we present two new concepts—core-concept and core-concept lattice, and have shown how core-concept lattice can be applied to collaborative recommendation system. It's time-consuming and troubled to extract formal concept in traditional FCA, however, compared with traditional FCA, our approach can obviously reduce the time of extracting concept, the extraction of core-concept is fast and easy. Though the number of all core-concepts less than whole formal concepts, it is enough for collaborative recommendation.

The method of traditional collaborative recommendation is first acquiring users' similarity matrix from users' rating matrix. By contrast, we build a core-concept lattice from a cross-table that is derived from the original ratings matrix when finding the neighbors. On the basis of core-concept lattice, we propose INN to predict and recommend, it doesn't need to compute the similarity between target user and all the other users, after finding the neighbors who have effect on the target user by using the core-concept lattice, it just needs to compute the similarity between the target and a very few candidate users. INN distinctly reduce the cost of finding the nearest neighbors, and also guarantee accuracy and coverage results equal to NN.

In the future works, we will apply INN to practice in collaborative recommendation system and prove its efficiency ulteriorly.

Acknowledgements

This work is supported by The Application Foundation of Sichuan Province (Grant No.2006J13-056). This work forms part of projects supported by the cultivating foundation of the science and technology leader of Sichuan province and the science & technology foundation of Xihua University (Grant No.R2006-26).

References

1. George Lekakos , George M. Giaglis, Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors, in: Interacting with Computers 18 (2006) pp.410–431.
2. J. L. Herlocker, Understanding and Improving Automated Collaborative Filtering Systems, in: Ph.D. thesis, University of Minnesota, 2000.
3. D.M. Pennock, E. Horvitz, Collaborative filtering by personality diagnosis: a hybrid memory- and model-based approach, in: IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden, August 1999, International Joint Conference on Artificial Intelligence.
4. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. Using collaborative filtering to weave an information tapestry, in: Communications of the ACM, 35(12), 1992, pp.61–70.
5. Mooney, R., & Roy, L. Content-based book recommending using learning for text categorization, in: ACM Conference on Digital Libraries, 5, 2000, pp. 195–204
6. Balabanovic, M., Shoham, Y. Fab: content-based collaborative recommendation, in: Communications of the ACM 40 (3). 1997 , pp. 66–72.
7. Pazzani, M.. A framework for collaborative, content-based and demographic filtering, in: Artificial Intelligence Review 13 (5–6), 1999, pp. 393–408.
8. Herlocker, J., Konstan, J., Riedl, J.. An empirical analysis of design choices in neighborhood-base collaborative filtering algorithms. in: Information Retrieval 5, 2002, pp. 287–310.
9. Burke, R.. Hybrid recommender systems: survey and experiments, in: User Modeling and User Adapted Interaction 12, 2002, pp. 331–370.
10. Sarwar, B., Karypis, G., Konstan, J., Riedl, J. Analysis of recommendation algorithms for e-commerce, in: Proceedings of the ACM E-commerce Conference, 2000, pp. 158–167.
11. Breese, J.S., Heckerman, D., Kadie, C.. Empirical analysis of predictive algorithms for collaborative filtering, in: Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43–52.
12. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.. Grouplens: an open architecture for collaborative filtering of netnews, in: Proceedings of the ACM Conference on Computer Supported Cooperative Work, 1994, pp. 175–186.
13. Patrick du Boucher-Ryan, Derek Bridge, Collaborative Recommending using Formal Concept Analysis, in: Knowledge-Based System, Volume 19, Issue 5, September 2006, PP.309-315.
14. R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: I. Rival (Ed.), Ordered Sets, Reidel, Dordrecht, Boston, 1982, pp. 445–470.
15. G. Stumme, R. Wille, U. Wille, Conceptual Knowledge Discovery in Databases using Formal Concept Analysis Methods, in: Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery, 1998, pp. 450–458.
16. T. Tilley, R. Cole, P. Becker, P. Eklund, A Survey of Formal Concept Analysis Support for Software Engineering Activities, in: Proceedings of the First International Conference on Formal Concept Analysis, 2003.
17. B. Díaz-Agudo, P.A. González-Calero, Classification-Based Retrieval using Formal Concept Analysis, in: Proceedings of the 4th International Conference on Case-Based Reasoning, 2001, pp. 173–188.
18. B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, Berlin, 1999.

19. S. Wrobel, K. Morik, T. Joachims, Maschinelles lernen und data mining, in: G. Grz, C.-R. Rollinger, J.Schneeberger (Eds.), Handbuch der Knstlichen Intelligenz, vol. 3, Auflage, Oldenbourg, Munchen, Wien, 2000, pp. 517–597.
20. Sergei O. Kuznetsov, Complexity of learning in concept lattices from positive and negative examples, in: Discrete Applied Mathematics 142, 2004, pp. 111–125.
21. Gerd Stumme, Rafik Taouil, Yves Bastide, Nicolas Pasquier, Lotfi Lakhal, Computing iceberg concept lattices with TITANIC, in: Data & Knowledge Engineering 42, 2002, pp. 189–2

How to Construct Formal Systems for Fuzzy Logics*

San-min Wang[1] and Zhi-jun Lu[2]

[1] Dept. of Computer, Nanchang Univ., Nanchang 330029, China
[2] College of Science, Hehai Univ., Jiangsu 210024, China
wangsanmin@hotmail.com

Abstract. In this paper, we present a sufficient and necessary condition to decide whether a weakly implicative logic has the proof by cases property (PCP for short). This result gives a general method for constructing a weakly implicative fuzzy logic from any given weakly implicative logic.

Keywords: Weakly implicative logic, weakly implicative fuzzy logic, the proof by cases property, substructural fuzzy logics.

1 Introduction

Basic fuzzy logic (BL for short) is the many-valued residuated logic introduced by P.Hájek [3] to cope with the logic of continuous t-norms and their residua. Hájek's approach to fuzzy logics has been extended by F.Esteva and L.Godo in [2], where the monoidal t-norm based logic (MTL for short) was introduced for capturing the set of 1-tautologies common for all left-continuous t-norm based propositional calculi. For BL and MTL, we not only need to prove that they are complete w.r.t. the class of all linearly ordered BL-algebras(or MTL-algebras) but also w.r.t. algebras with lattice reduct $[0,1]$. However, Cintula posed the thesis that fuzzy logics are those which are complete w.r.t. totally ordered matrices and do not need to be complete w.r.t. matrices with lattice reduct $[0,1]$.

Cintula's thesis is embodied in his weakly implicative logic (WIL for short) and weakly implicative fuzzy logic (WIFL for short) in [1], where he presents many general theorems for both classes of logics and demonstrate their usefulness and importance. For our purposes it is interesting to remark the following properties of WIL and WIFL: (i) The matching rules for the implication $\to$ are wonderful because under such selection WIL are complete w.r.t. totally ordered matrices and WIFL admit several equivalent characterizations, for example, a finitary weakly implicative logic is a fuzzy logic if and only if it has Linear Extension Property (LEP for short) if and only if it has Prelinearity Property (PP for short) if and only if it has Subdirect Decomposition Property (SDP for short); (ii) The matching rules for the additive disjunction $\vee$

* This work is supported by the National Foundation of Natural Sciences of China (Grant No: 60663002) and the Grant Project of science and technology of The Education Department of Jiangxi Province under Grant No 200618.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 593–601, 2007.
springerlink.com

are supremality, maximality, commutativity and under such selection a finitary weakly implicative logic with $\vdash (\varphi \to \psi) \vee (\psi \to \varphi)$ is a weakly implicative fuzzy logic if and only if it has the Proof by Cases Property (PCP for short).

From the property (ii), (PCP) plays a key role in deciding whether a weakly implicative logic is a weakly implicative fuzzy logic. Then which logics admits (PCP) ? In Section 3 of this paper, we firstly present new matching rules for $\vee$ which are independent of the implication $\to$ and weaker than those given by Cintula. Then we give a sufficient and necessary condition to decide whether a weakly implicative logic has (PCP), which presents a general method for constructing a weakly implicative fuzzy logic from any given weakly implicative logic. In Section 4, we apply this method to substructural fuzzy logics.

2 Weakly Implicative Logics

Now we recall some of the concepts required in this paper, which are mainly from [1,6].

Definition 1 [1]**.** A propositional language L is a triple $(\mathbf{VAR}, \mathbf{C}, \mathbf{a})$, where $\mathbf{VAR}$ is a non-empty set of the (propositional) variables, $\mathbf{C}$ is a non-empty set of the (propositional) connectives, and $\mathbf{a}$ is a function assigning to each element of $\mathbf{C}$ a natural number, called its arity. A connective c for which $\mathbf{a}(c) = 0$ is called a truth constant. The set $\mathbf{VAR}$ is usually taken as a fixed countable set, and so we usually define the propositional language L as a pair $(\mathbf{C}, \mathbf{a})$. Later on we fix symbols for some basic connectives $(\to, \wedge, \vee, \bot)$ together with their arities and then we define propositional language just as a set of connectives.

Definition 2 [1]**.** Let L be a propositional language. The set of (propositional) formulae $\mathbf{FOR}_L$ is the smallest set which contains $\mathbf{VAR}$ and is closed under connectives from $\mathbf{C}$, i.e., for each $(c, n) \in L$ and each $\varphi_1, \varphi_2, \cdots, \varphi_n \in \mathbf{FOR}_L$ we have $c(\varphi_1, \varphi_2, \cdots, \varphi_n) \in \mathbf{FOR}_L$.

Definition 3 [1]**.** Let L be a propositional language. A substitution is a mapping $\sigma : \mathbf{FOR}_L \to \mathbf{FOR}_L$, such that σ is compatible with connectives, i.e., $\sigma(c(\varphi_1, \varphi_2, \cdots, \varphi_n)) = c(\sigma(\varphi_1), \sigma(\varphi_2), \cdots, \sigma(\varphi_n))$ holds for each $(c, n) \in L$, where (c, n) denotes that c is n-arity connective. The set of all substitutions will be denoted as $\mathbf{SUB}_L$.

Definition 4 [1]**.** A consecution in the propositional language L is a pair $< X, \varphi >$, where $X \subseteq \mathbf{FOR}_L$ and $\varphi \in \mathbf{FOR}_L$. The set of all consecutions will be denoted as CON_L. Instead of $< X, \varphi >$ we write $X \rhd \varphi$. Since CON_L =

$P(\mathbf{FOR}_L)\times\mathbf{FOR}_L$, each subset X of CON_L can be understood as a relation between the sets of formulae and formulae (we identify the set X and the relation $\vdash_X$ in the following way $X\vdash_X \varphi$ iff $X \rhd \varphi \in X$). For $X\subseteq\mathbf{FOR}_L$, $X\subseteq CON_L$ and substitution σ we fix the following convention: (i) By $\sigma(X)$ we understand the set $\{\sigma(\varphi)\,|\,\varphi\in X\}$; (ii) By $\sigma(X)$ we understand the set $\{\sigma(X)\sigma(\varphi)\,|\,(X\rhd\varphi)\in X\}$; (iii) By $SUB_L(X)$ we denote the set $\cup_{\sigma\in\mathbf{SUB}_L}\sigma(X)$.

Definition 5 [1]**.** Let L be a propositional language. A non-empty set $\mathbf{L}\subseteq CON_L$ is called a logic in language L when it satisfies:

(i) if $\varphi\in X$, then $X\vdash_{\mathbf{L}}\varphi$.

(ii) if $Y\vdash_{\mathbf{L}}\varphi$ for each $\varphi\in X$ and $X\vdash_{\mathbf{L}}\psi$ then $Y\vdash_{\mathbf{L}}\psi$.

(iii) if $X\vdash_{\mathbf{L}}\varphi$, then $\sigma(X)\vdash_{\mathbf{L}}\sigma(\psi)$ for each $\sigma\in\mathbf{SUB}_L$.

Definition 6 [1]**.** Let L be a propositional language. An axiomatic system AS in language L is a non-empty set $AS\subseteq CON_L$, which is closed under arbitrary substitution (i.e., $\mathbf{SUB}_L(AS)=AS$). The elements of AS of the form $X\rhd\varphi\in AS$ are called axioms for $X=\varnothing$, n-ary deduction rules for $|X|=n$, and infinitary deduction rules otherwise. The axiomatic system is said to be finite if there is a finite set $X\in AS$ such that $\mathbf{SUB}_L(X)=AS$. The axiomatic system is said to be finitary if all its deduction rules are finite.

Definition 7 [1]**.** Let L be a propositional language and AS an axiomatic system in L. By a theory we mean just a set of formulas. An AS-proof of the formula φ in a theory T is a well-founded tree labeled by formulae; the root is labeled by φ and the leaves by either axioms or elements of T; and if a node is labeled by ψ and its preceding nodes are labeled by $\psi_1,\psi_2,\cdots$ then $<\{\psi_1,\psi_2,\cdots\},\psi>\in AS$. We write $T\vdash^p_{AS}\varphi$ if there is a proof of φ in T. The assertion that a tree is well-founded just means that there is no infinitely long branch.

Definition 8 [1]**.** Let L be a propositional language, AS an axiomatic system in L, and $\mathbf{L}$ a logic in L. We say that AS is an axiomatic system for (a presentation of) the logic $\mathbf{L}$ iff $\mathbf{L}=\vdash^p_{AS}$. A logic is said to be finitely axiomatizable (finitary) if it has some finite (finitary) presentation. Let (R) be a symbol of a consecution and $\mathbf{L}$ a logic then **L**+(R) denotes the extension of **L** obtained by adding (R) to an axiomatic system of **L**.

Definition 9 [1]**.** Let L be a propositional language, such that $(\to, 2) \in L$ and let $\mathbf{L}$ be a logic in L. We say that $\mathbf{L}$ is a weakly implicative logic iff the following consecutions are elements of $\mathbf{L}$:

(Ref) $\vdash_{\mathbf{L}} \varphi \to \varphi$

(MP) $\varphi, \varphi \to \psi \vdash_{\mathbf{L}} \psi$

(WT) $\varphi \to \psi, \psi \to \chi \vdash_{\mathbf{L}} \varphi \to \chi$

(Cng_c) $\varphi_1 \leftrightarrow \psi_1, \cdots, \varphi_n \leftrightarrow \psi_n \vdash_{\mathbf{L}} c(\varphi_1, \cdots, \varphi_n) \to c(\psi_1, \cdots, \psi_n)$ for each $(c, n) \in L$, where $\varphi \leftrightarrow \psi$ be a shortcut for $\{\varphi \to \psi, \psi \to \varphi\}$. In Definition 9, the abbreviation (Ref) is for reflexivity, (MP) for modus ponens, (WT) for weak transitivity, and (Cng) for congruence.

Definition 10 [1]**.** A matrix $\mathbf{M} = (\mathbf{A}, D)$ for a language L is an L-algebra $\mathbf{A}$ plus its subset D of designated values. The interpretation of a connective c of the language L in an L-matrix $\mathbf{M}$ is denoted by $c_{\mathbf{M}}$. We say that an evaluation e in $\mathbf{M}$ validates a formula φ iff $e(\varphi) \in D$. The relation of semantic consequence $X \models_{\mathbf{M}} \varphi$ holds iff each evaluation in M that validates all formulae in X validates φ as well. For the logic $\mathbf{L}$ represented by the consequence relation $\vdash_{\mathbf{L}}$, $\mathbf{M}$ is an $\mathbf{L}$-matrix iff $X \vdash_{\mathbf{L}} \varphi$ implies $X \models_{\mathbf{M}} \varphi$ (for all X and φ).

Weakly implicative logics can be characterized as those which are complete w.r.t. a class of *(pre)ordered matrices* (in which the set D of designated values is upper), if the ordering of the elements of the matrix $\mathbf{M}$ is defined as $x \leq_{\mathbf{M}} y$ iff $x \to_{\mathbf{M}} y \in D$.

Definition 11 [1]**.** A weakly implicative logic $\mathbf{L}$ is called a fuzzy logic if it is complete w.r.t. the linearly ordered $\mathbf{L}$-matrices.

Definition 12 [1]**.** A weakly implicative logic $\mathbf{L}$ with the additive disjunction $\vee$ has the Proof by Cases Property (PCP) if for each theory T we get $T, \varphi \vee \psi \vdash \chi$ whenever $T, \varphi \vdash \chi$ and $T, \psi \vdash \chi$.

Definition 13 [1]**.** A weakly implicative logic L has Prelinearity Property (PP) if for each theory T we get $T \vdash \chi$ whenever $T, \varphi \to \psi \vdash \chi$ and $T, \psi \to \varphi \vdash \chi$.

Theorem 1 [1]**.** Let $\mathbf{L}$ be a finitary logic then $\mathbf{L}$ is a fuzzy logic iff $\mathbf{L}$ has (PP).

3 The Proof by Cases Property of Logics

Let $\mathbf{L}$ be a finitary logic in the sense of Definition 5 and its language has the additive disjunction $\vee$ (possibly without the implication $\rightarrow$) and the matching rules for $\vee$ are as follows:

$(\vee_1)$ $\varphi\vee\varphi\vdash_{\mathbf{L}}\varphi$

$(\vee_2)$ $\varphi\vdash_{\mathbf{L}}\varphi\vee\psi$

$(\vee_3)$ $\varphi\vee\psi\vdash_{\mathbf{L}}\psi\vee\varphi$

$(\vee_4)$ $(\varphi\vee\psi)\vee\chi\vdash_{\mathbf{L}}\varphi\vee(\psi\vee\chi)$.

Proposition 1. The followings are rules of $\mathbf{L}$.

(D_1) $\psi\vdash_{\mathbf{L}}\varphi\vee\psi$

(D_2) $\varphi\vee(\psi\vee\chi)\vdash_{\mathbf{L}}(\varphi\vee\psi)\vee\chi$

Proof. (D_1):

1° ψ assumption

2° $\psi\vee\varphi$ $1^\circ,\vee_2$

3° $\varphi\vee\psi$ $2^\circ,\vee_3$

(D_2):

1° $\varphi\vee(\psi\vee\chi)$ assumption

2° $(\psi\vee\chi)\vee\varphi$ $1^\circ,\vee_3$

3° $\psi\vee(\chi\vee\varphi)$ $2^\circ,\vee_4$

4° $(\chi\vee\varphi)\vee\psi$ $3^\circ,\vee_3$

5° $\chi\vee(\varphi\vee\psi)$ $4^\circ,\vee_4$

6° $(\varphi\vee\psi)\vee\chi$ $5^\circ,\vee_3$ □

Theorem 2. If $\mathbf{L}$ has (PCP) and $\varphi_1,\cdots,\varphi_n\vdash_{\mathbf{L}}\varphi$ then $\varphi_1\vee\psi,\cdots,\varphi_n\vee\psi\vdash_{\mathbf{L}}\varphi\vee\psi$.

Proof. $\varphi_1,\cdots,\varphi_{n-1},\varphi_n\vdash_{\mathbf{L}}\varphi\vee\psi$ holds by $\varphi_1,\cdots,\varphi_{n-1},\varphi_n\vdash_{\mathbf{L}}\varphi$ and $\varphi\vdash\varphi\vee\psi$. $\varphi_1,\cdots,\varphi_{n-1},\psi\vdash_{\mathbf{L}}\varphi\vee\psi$ holds by $\varphi_1,\cdots,\varphi_{n-1},\psi\vdash_{\mathbf{L}}\psi$ and $\psi\vdash_{\mathbf{L}}\varphi\vee\psi$. Thus $\varphi_1,\cdots,\varphi_{n-1},\varphi_n\vee\psi\vdash_{\mathbf{L}}\varphi\vee\psi$ holds by (PCP). Then $\varphi_1\vee\psi,\cdots,\varphi_n\vee\psi\vdash_{\mathbf{L}}\varphi\vee\psi$ holds by repeating this procedure. □

Definition 14. Let (R) be a rule of the form $\varphi_1, \cdots, \varphi_n \vdash_{\mathbf{L}} \varphi$ then $\varphi_1 \vee \phi, \cdots, \varphi_n \vee \phi \vdash_{\mathbf{L}} \varphi \vee \phi$ is called the $*$-rule corresponding to (R) and denoted by (R^*). For example, (MP^*) denotes $\varphi \vee \phi, (\varphi \to \psi) \vee \phi \vdash_{\mathbf{L}} \psi \vee \phi$.

Theorem 3. Let $\mathbf{L}^*$ denote the logic obtained by adding the $*$-rule of each rule of $\mathbf{L}$ to $\mathbf{L}$. Then

(i) If $T, \varphi \vdash_{\mathbf{L}^*} \chi$ and $T, \psi \vdash_{\mathbf{L}^*} \phi$ then $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi \vee \phi$.

(ii) If $T, \varphi \vdash_{\mathbf{L}^*} \chi$ and $T, \psi \vdash_{\mathbf{L}^*} \chi$ then $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi$, i.e., $\mathbf{L}^*$ has (PCP).

Proof. (i) Let the length of proofs for $T, \varphi \vdash_{\mathbf{L}^*} \chi$ and $T, \psi \vdash_{\mathbf{L}^*} \phi$ are h_1 and h_2, respectively. We proceed by induction on the combined length $h = h_1 + h_1$ of proofs for $T, \varphi \vdash_{\mathbf{L}^*} \chi$ and $T, \psi \vdash_{\mathbf{L}^*} \phi$. If $h = 0$, then we have several cases. If χ is φ and ψ is ϕ, then the result follows immediately. Otherwise, without loss of generality, $\chi \in T$ or χ is an axiom, and since $\chi \vdash_{\mathbf{L}^*} \chi \vee \phi$, it follows that $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi \vee \phi$.

Suppose that $h > 0$. Let $\varphi_1, \cdots, \varphi_n \vdash_{\mathbf{L}} \varphi_0$ be an arbitrary rule of $\mathbf{L}$ and denoted by (R). We only give two cases for, without loss of generality, $T, \varphi \vdash_{\mathbf{L}^*} \chi$:

Case 1. The proof of $T, \varphi \vdash_{\mathbf{L}^*} \chi$ ends with (R) then $T, \varphi \vdash_{\mathbf{L}^*} \varphi_1, \cdots, T, \varphi \vdash_{\mathbf{L}^*} \varphi_n$ and χ is φ_0. By the induction hypothesis, $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \varphi_1 \vee \phi, \cdots,$ $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \varphi_n \vee \phi$. Then $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi \vee \phi$ holds by (R^*).

Case 2. The proof of $T, \varphi \vdash_{\mathbf{L}^*} \chi$ ends with (R^*) then $T, \varphi \vdash_{\mathbf{L}^*} \varphi_1 \vee \gamma, \cdots, T, \varphi \vdash_{\mathbf{L}^*} \varphi_n \vee \gamma$ and χ is $\varphi_0 \vee \gamma$. By the induction hypothesis, $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} (\varphi_1 \vee \gamma) \vee \phi, \cdots,$ $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} (\varphi_n \vee \gamma) \vee \phi$. Then $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \varphi_1 \vee (\gamma \vee \phi), \cdots,$ $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \varphi_n \vee (\gamma \vee \phi)$ by $(\vee_4)$ and thus $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \varphi_0 \vee (\gamma \vee \phi)$ holds by (R^*). Hence $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi \vee \phi$ holds by (D_2).

(ii) Observe that if $T, \varphi \vdash_{\mathbf{L}^*} \chi$ and $T, \psi \vdash_{\mathbf{L}^*} \chi$ then $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi \vee \chi$ holds by (i) and thus $T, \varphi \vee \psi \vdash_{\mathbf{L}^*} \chi$ holds by $(\vee_1)$. □

Notice that (R^{**}) is provable from (R^{*}) and $(\vee_4)$ and thus $\mathbf{L}^{*}$ is finitely axiomatizable if $\mathbf{L}$ is finitely axiomatizable. Clearly, Theorem 2 and 4 give a sufficient and necessary condition for logics with rules $(\vee_1),(\vee_2),(\vee_3)$ and $(\vee_4)$ to admit (PCP).

The rules $\vee_1,\vee_2,\vee_3$ and $\vee_4$ for $\mathbf{L}$ are independent of the implication $\rightarrow$ while in [1] the simplest weakly implicative logic is independent of the additive disjunction $\vee$. In order to make a relation between the simplest weak implicative logic and $\mathbf{L}$, we add the following additional rule to the matching rules $\vee_1,\vee_2,\vee_3$ and $\vee_4$ for the additive disjunction $\vee$.

$$(\vee_5) \quad \varphi\rightarrow\psi,\varphi\vee\psi\vdash\psi$$

Remark 1. In [1], Cintula selects the following as the matching rules for $\vee$.

$$(Max) \quad \vdash\varphi\rightarrow\varphi\vee\psi$$
$$(Com) \quad \vdash\varphi\vee\psi\rightarrow\psi\vee\varphi$$
$$(Sup) \quad \varphi\rightarrow\chi,\psi\rightarrow\chi\vdash\varphi\vee\psi\rightarrow\chi$$

where the abbreviation (Max) is for maximality, (Com) for commutativity and (Sup) for supremality. Notice that this three rules heavily depend on the implication $\rightarrow$ and are stronger than $(\vee_1)\sim(\vee_5)$. Thus some results in [1] are not valid any more. But the following proposition holds by Lemma 16,17 of [1] and Theorem 1.

Proposition 2. Let $\mathbf{L}$ be a finitary weakly implicative logic with the additive disjunction $\vee$ for which the matching rules are $(\vee_1)\sim(\vee_5)$ and (PL) $\vdash_{\mathbf{L}}(\varphi\rightarrow\psi)\vee(\psi\rightarrow\varphi)$. Then $\mathbf{L}$ is a fuzzy logic iff $\mathbf{L}$ has (PCP).

Remark 2. By Theorem 3 and Proposition 2, a general method is presented for constructing the minimal weakly implicative fuzzy logic from any given finitary weakly implicative logic, which is the most important result in this section. By this general method, we can easily construct many weakly implicative fuzzy logics and further research has been done in Section 4.

4 Some Applications

Definition 15 [4]**.** Multiplicative additive intuitionistic linear logic **MAILL** consists of the following axioms and rules:

$$(A_1) \quad \vdash\varphi\rightarrow\varphi$$
$$(A_2) \quad \vdash(\varphi\rightarrow\psi)\rightarrow((\psi\rightarrow\chi)\rightarrow(\varphi\rightarrow\chi))$$
$$(A_3) \quad \vdash(\varphi\rightarrow(\psi\rightarrow\chi))\rightarrow(\psi\rightarrow(\varphi\rightarrow\chi))$$

(A_4) $\vdash(\varphi \& \psi \to \chi) \leftrightarrow (\varphi \to (\psi \to \chi))$

(A_5) $\vdash \varphi \wedge \psi \to \psi$

(A_6) $\vdash \varphi \wedge \psi \to \varphi$

(A_7) $\vdash (\chi \to \varphi) \wedge (\chi \to \psi) \to (\chi \to \varphi \wedge \psi)$

(A_8) $\vdash \varphi \to \varphi \vee \psi$

(A_9) $\vdash \psi \to \varphi \vee \psi$

(A_{10}) $\vdash (\varphi \to \chi) \wedge (\psi \to \chi) \to (\varphi \vee \psi \to \chi)$

(A_{11}) $\vdash \varphi \leftrightarrow (\top \to \varphi)$

(A_{12}) $\vdash \varphi \to \bar{1}$

(A_{13}) $\vdash \bar{0} \to \varphi$

(MP) $\varphi, \varphi \to \psi \vdash \psi$

(ADJ) $\varphi, \psi \vdash \varphi \wedge \psi$

where $\varphi \leftrightarrow \psi$ is an abbreviation for $(\varphi \to \psi) \wedge (\psi \to \varphi)$.

Definition 16 [4]**.** Uninorm logic **UL**, the most basic substructural fuzzy logic, is MAILL extended with the following prelinearity axiom.

$$(PRL)\ \vdash ((\varphi \to \psi) \wedge \top) \vee ((\psi \to \varphi) \wedge \top)$$

Clearly, $(\vee_1) \sim (\vee_5)$ hold in **MAILL** and then **MAILL** $+ (MP^*) + (ADJ^*)$ has (PCP) by Theorem 3. By Proposition 2, **MAILL** $+ (MP^*) + (ADJ^*) + (PL)$ is a fuzzy logic in Cintula's sense, where

(MP^*) $\varphi \vee \chi, (\varphi \to \psi) \vee \chi \vdash \psi \vee \chi$

(ADJ^*) $\varphi \vee \chi, \psi \vee \chi \vdash (\varphi \wedge \psi) \vee \chi$

(PL) $\vdash (\varphi \to \psi) \vee (\psi \to \varphi)$.

Furthermore, we have the following proposition.

Proposition 3. **MAILL** $+ (MP^*) + (ADJ^*) + (PL) =$ **UL**.

Proof. By Lemma 14, 25 of [3], (MP^*), (ADJ^*) and (PL) hold in **UL** and thus **MAILL** $+ (MP^*) + (ADJ^*) + (PL) \subseteq$ **UL**. By $\vdash_{\mathbf{MAILL}} \top$, $\varphi \to \psi \vdash_{\mathbf{MAILL}} (\varphi \to \psi) \wedge \top$ and $\psi \to \varphi \vdash_{\mathbf{MAILL}} (\psi \to \varphi) \wedge \top$ hold and thus $((\varphi \to \psi) \wedge \top) \vee ((\psi \to \varphi) \wedge \top)$ is a theorem of **MAILL** $+ (MP^*) + (ADJ^*) + (PL)$ by its proof by cases property. This completes the proof. □

5 Conclusion and Remark

In this paper, I select $(\vee_1) \sim (\vee_4)$ as the basic matching rules for $\vee$ and $(\vee_5)$ as the additional matching rule. In my opinion, such selection is wonderful because it embodies my viewpoint that the criterion of selection is the minimal conditions for logic to admit PCP. Thus a general method is presented by Theorem 3 for constructing a weakly implicative fuzzy logic from any given finitary weakly implicative logic. Some work about this research direction has been done in [6].

References

[1] Cintula, P.: Weakly Implicative (Fuzzy) logics I, Archive for Mathematical Logic, (2006)673-704.

[2] Esteva, F., Godo, L.: Monoidal t-norm based logic: towards a logic for left-continuous t-norms, Fuzzy Sets and Systems, (2001)271-288.

[3] Hájek, P.: Metamathematics of Fuzzy Logic, Dordrecht, Kluwer, 1998.

[4] Metcalfe, G., Montagna, F.: Substructural fuzzy logics, To appear in *Journal of Symbolic Logic*.

[5] Wang, S.M.,Qin F., Wang M.Y.: Solutions to Cintula's open problems Fuzzy Sets and Systems, (2006) 2091-2099.

[6] Wang, S.M., Cintula, P.: On the proof by cases property in weakly implicative logics, Fuzzy Sets and Systems, submitted.

A Logical Framework for Fuzzy Quantifiers Part I: Basic Properties*

San-min Wang[1,2], Bin Zhao[2], and Peng Wang[3]

[1] Dept. of Mathematics, Shaanxi Normal Univ., Xi'an 710062, China
[2] Dept. of Computer, Nanchang Univ., Nanchang 330029, China
[3] Dept. of Mathematics, Beijing Forestry Univ., Beijing 100083, China
wangsanmin@hotmail.com,b.zhao@snnu.edu.cn

Abstract. Fuzzy quantifiers have important applications in a great variety of fields such as database querying, data mining and knowledge discovering, inductive learning and so on. Recently, M.S.Ying introduces a novel fuzzy framework for linguistic quantifiers which are modeled by Sugeno integrals. Essentially, the conjunction and disjunction in Ying's framework are interpreted as the "min" and "max" operations, which restricts the application of this theory in some sense. In this paper, we extended Ying's framework by interpreting the conjunction and disjunction as t-norm and t-conorm respectively. And some elegant logical results for our framework have been obtained.

Keywords: Fuzzy logic, Fuzzy quantifier, Sugeno's integral, MTL.

1 Introduction

Since fuzzy quantifiers have important applications in a great variety of fields such as database querying, data mining and knowledge discovering, inductive learning and so on, they are studied by various authors and many interesting theories about them have been proposed [3,4]. Particularly, M.S.Ying [6] introduces a new framework for modeling quantifiers in natural languages.

In Ying's framework, a quantifier Q is seen as a family of fuzzy measures indexed by nonempty sets. When predicates in linguistically quantified statements are vague, the truth value of a quantified proposition is then evaluated using Sugeno's integral [5]. The advantage of this framework is that it allows us to have some elegant logical properties of linguistic quantifiers. For example, we are able to establish a prenex normal form theorem for linguistic quantifiers (see Corollary 34 of [6]). However, the conjunction and disjunction in Ying's framework are interpreted as the "min" and "max" operations (see Definition 23 of [6]), which restricts the application of this theory in some sense. In this paper, we extended Ying's framework by interpreting the conjunction and disjunction as t-norm and t-conorm respectively. Actually, we construct several logical systems which are based on the monoidal t-norm

* This work is supported by the National Foundation of Natural Sciences of China (Grant No: 60663002) and the Grant Project of science and technology of The Education Department of Jiangxi Province under Grant No 200618.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 602–611, 2007.
springerlink.com

based logic MTL and with linguistic quantifiers modelled by Sugeno integrals. And some elegant logical results for these logics have been obtained.

2 Sugeno Integrals and Ying's Definition of Fuzzy Quantifiers

Now we recall some of the concepts required in this paper, which mainly are from [5,6].

Definition 1. [5, page 10] Let X be a nonempty set. A Borel field over X is a subset $\wp$ of 2^X satisfying the next conditions:

(1) $\emptyset \in \wp$; (2) If $E \in \wp$, then $X - E \in \wp$; and (3) If $E_n \in \wp$ for $1 \leq n < \infty$, then $\bigcup_{n=1}^{\infty} E_n \in \wp$.

Definition 2. [5, Definition 2.3] If X is a nonempty set and $\wp$ is a Borel field over X, then $(X, \wp)$ is called a measurable space.

In order to define fuzzy measure, we need the notion of limit of set sequence. If $E_1 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$, then the sequence E_n is said to be increasing and we define $\lim_{n\to\infty} E_n = \bigcup_{n\to\infty}^{\infty} E_n$, and if $E_1 \supseteq \cdots \supseteq E_n \supseteq E_{n+1} \supseteq \cdots$, then the sequence E_n is said to be decreasing and we define $\lim_{n\to\infty} E_n = \bigcap_{n\to\infty}^{\infty} E_n$. Both increasing and decreasing sequences of sets are said to be monotone.

Definition 3. [5, Definitions 2.2 and 2.4] Let $(X, \wp)$ be a measurable space. If a set function $m : \wp \to [0,1]$ satisfies the following properties:

(1) $m(\emptyset) = 0$ and $m(X) = 1$;

(2) (Monotonicity) If $E, F \in \wp$ and $E \subseteq F$, then $m(E) \leq m(F)$; and

(3) (Continuity) If $E_n \in \wp$ for $1 \leq n < \infty$ and E_n is monotone, then

$m(\lim_{n\to\infty} E_n) = \lim_{n\to\infty} m(E_n)$, then m is called a fuzzy measure over $(X, \wp)$, and $(X, \wp, m)$ is called a fuzzy measure space.

Definition 4. [5, Definition 3.3] Let $(X, \wp)$ be a measurable space, and let $h : X \to [0,1]$ be a mapping from X into the unit interval. For any $\in [0,1]$, we write $H = \{x \in X : h(x) \geq \lambda\}$. If $H_\lambda \in \wp$ for all $\lambda \in [0,1]$, then h is said to be $\wp$-measurable. For any measurable space $(X, \wp)$, we write $M(X, \wp)$ for the set of all fuzzy measures on $(X, \wp)$.

Definition 5. [5, Definition 3.1 and page 19] Let $(X,\wp,m)$ be a fuzzy measure space. If $A\in\wp$, $h:X\to[0,1]$ is a $\wp$-measurable function, then the Sugeno's integral of h over A is defined by $\int_A h\circ m=\sup_{\in[0,1]}\min[,m(A\cap H)]$, where $H=\{x\in X:h(x)\geq\}$ for each $\in[0,1]$. In particular, $\int_A h\circ m$ will be abbreviated to $\int h\circ m$ whenever $A=X$.

Definition 6. [6, Definition 5] Let $(X,\wp,m)$ be a fuzzy measure space. Then the dual set function $m^*:\wp\to[0,1]$ of m is defined by $m^*(E)=1-m^*(X-E)$ for each $E\in\wp$. It is easy to see that m^* is a fuzzy measure over $(X,\wp)$ too.

Definition 7. [6, Definition 14] A fuzzy quantifier (or quantifier for short) consists of the following two items: (i) for each nonempty set X, a Borel field $\wp_X$ over X is equipped; and (ii) a choice function $Q:(X,\wp_X)\mapsto Q_{(X,\wp_X)}\in M(X,\wp_X)$ of the (proper) class $\{M(X,\wp_X):(X,\wp_X)$ is a measurable space$\}$. For simplicity, $Q_{(X,\wp_X)}$ is often abbreviated to Q_X whenever the Borel field $\wp_X$ can be recognized from the context.

Definition 8. [6, Definition 18] Let Q be quantifiers. Then the dual Q^* of Q is defined as follows: for any nonempty set X and for any $E\in\wp_X$, $Q_X^*(E)=1-Q_X(X-E)$.

3 MTL and Its Properties

In this section, we introduce F.Esteva and L.Godo's MTL system and its main properties.

Definition 9. [1] For any left-continuous t-norm $*$ and its residuum $\Rightarrow$, a propositional calculus QPC($*$) is defined as follows: The set F of well-formed formulas ($wfs.$ for short) of QPC($*$) is defined as usual from a countable set of propositional variables $p_1,p_2,\cdots$, three connectives $\&,\to,\wedge$ and the truth constant $\bar{0}$. Further definable connectives are: $\neg\varphi$ is $\varphi\to\bar{0}$, $\varphi\equiv\psi$ is $(\varphi\to\psi)\&(\psi\to\varphi)$ and $\varphi\vee\psi$ is $((\varphi\to\psi)\to\psi)\wedge((\psi\to\varphi)\to\varphi)$.

Precedence of connectives in decreasing order is negation $\neg$, the strong conjunction $\&$, conjunction $\wedge$, disjunction $\vee$, implication $\to$ and equivalence $\equiv$. Sometimes parentheses are omitted in a formula.

A truth evaluation e is a mapping from F to [0,1] such that $e(\varphi \& \psi) = e(\varphi) * e(\psi)$, $e(\varphi \to \psi) = e(\varphi) \Rightarrow e(\psi)$ and $e(\varphi \wedge \psi) = min(e(\varphi), e(\psi))$.

The Monoidal t-norm based logic (MTL for short) which aims at capturing the set of 1-tautologies which are common for all left-continuous t-norm based propositional calculi QPC($*$), is defined as follows:

Definition 10. [1] Axioms of MTL are:

$$(A_1)\quad (\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi)) \qquad (A_2)\quad \varphi \& \psi \to \varphi$$
$$(A_3)\quad \varphi \& \psi \to \psi \& \varphi \qquad (A_4)\quad \varphi \wedge \psi \to \varphi$$
$$(A_5)\quad \varphi \wedge \psi \to \psi \wedge \varphi \quad (A_6)\quad \varphi \& (\varphi \to \psi) \to \varphi \wedge \psi$$
$$(A_{7a})\quad (\varphi \to (\psi \to \chi)) \to (\varphi \& \psi \to \chi)$$
$$(A_{7b})\quad (\varphi \& \psi \to \chi) \to (\varphi \to (\psi \to \chi))$$
$$(A_8)\quad ((\varphi \to \psi) \to \chi) \to (((\psi \to \varphi) \to \chi) \to \chi) \quad (A_9)\quad \bar{0} \to \varphi$$

The rule of inference of MTL is modus ponens. A theory of MTL is a set of $wfs.$. $T \vdash \varphi$ denotes that φ is provable in the theory T. $\bar{1}$ stands for $\bar{0} \to \bar{0}$.

Proposition 1. [1] The following are theorems or rules of MTL:

$$(D_1)\quad \{\varphi \to \psi, \psi \to \chi\} \vdash \varphi \to \chi$$
$$(D_2)\quad \varphi \to (\psi \to \chi) \equiv \varphi \& \psi \to \chi \qquad (D_3)\quad \varphi \& \psi \to \varphi^2 \vee \psi^2$$
$$(D_4)\quad \varphi \vee \varphi \equiv \varphi \qquad (D_5)\quad \varphi \to (\psi \to \chi) \equiv \psi \to (\varphi \to \chi)$$

Definition 11. [1] An MTL-algebra is a bounded residuated lattice $\mathbf{L} = (\mathbf{M}, \wedge, \vee, *, \Rightarrow, 0, 1)$, where $\wedge$ and $\vee$ are the lattice meet and join operations and $(*, \Rightarrow)$ is a residuated pair, satisfying the pre-linearity equation, i.e., $(x \Rightarrow y) \vee (y \Rightarrow x) = 1$. An MTL-chain is a linearly ordered MTL-algebra. If $*$ is a left-continuous t-norm and $\Rightarrow$ is its residuum then $([0,1], \wedge, \vee, *, \Rightarrow, 0, 1)$ is an MTL-chain and called standard MTL-algebra. As usual, evaluations of F over $\mathbf{L}$ and $\mathbf{L}$-tautologies are defined.

Definition 12. [1] A logical calculus C is a schematic extension of MTL if it results from MTL by adding some (finitely or infinitely many) axiom schemes. Moreover, an MTL-algebra M is said to be a $C-$algebra if it is an MTL-algebra such that all axioms of C are $\mathbf{L}$-tautologies.

4 L_Q and Its Syntax

In this section, we introduce the logical systems L_Q, which are based on the monoidal t-norm based logic MTL and with linguistic quantifiers modelled by Sugeno integrals. We can regard it as a generalization of Ying's first order logical language L_q [6]. Let $Q \subseteq \{\forall, \exists, \vartriangle, \triangledown\}$. The language L_Q has the following symbols:

(i) The sequence $\{c_n\}$ of object constants, denoted by L_c; (ii) The sequence $\{x_n\}$ of object variables, denoted by L_v; (iii) The sequence $\{P_j^k\}$ of predicate letters, denoted by L_p, where P_j^k denotes the j-th k-ary predicate; (iv) The set $\{\&, \rightarrow, \wedge, \overline{0}\}$ of connectives; (v)The set $\{(,),,\}$ of punctuation symbols; (vi) the quantifier set Q. As usual we can define term, atomic formula, well-formed Q-formula (Q-$wf.$ for short), etc.. The set of terms is denoted by L_t and the set of Q-$wfs.$ is denoted by L_w^Q.

Let φ be a $wf.$ and x, y variables. If $y = x$ then y is substitutable for x in φ. If $y \neq x$ then y is substitutable for x in φ if y does not occur free in φ and no subformula of φ of the form $(\vartriangle y)\psi$ contains an occurrence of x free in φ. A constant is substitutable for any variable in any formula.

Definition 15. The result $\varphi(x/t)$ of substitution of a term t for a variable x in φ is defined as follows: (i) If φ is atomic then $\varphi(x/t)$ results from φ by replacing all occurrence of x in φ by t; (ii) $\overline{0}(x/t) = \overline{0}$, $\overline{1}(x/t) = \overline{1}, (\varphi \rightarrow \psi)(x/t) = \varphi(x/t) \rightarrow \psi(x/t), (\varphi \,\&\, \psi)(x/t) = \varphi(x/t) \,\&\, \psi(x/t)$ and $(\varphi \wedge \psi)(x/t) = \varphi(x/t) \wedge \psi(x/t)$; (iii) $[(\delta x)\varphi](x/t) = (\delta x)\varphi$ (no change) and $[(\delta y)\varphi](x/t) = (\delta y)[\varphi(x/t)]$, where $\delta \in Q$ and $y \neq x$.

We present the axiomatic schemata for quantifiers as follows:

(i) Axiomatic schemata for $\{\forall\}$:

$(\forall_t)$ $(\forall x)\varphi(x) \rightarrow \varphi(t)$ (t substitutable for x in $\varphi(x)$)

$(\forall_\rightarrow)$ $(\forall x)(\chi \rightarrow \varphi) \rightarrow (\chi \rightarrow (\forall x)\varphi)$ (x not free in χ)

$(\forall_\vee)$ $(\forall x)(\chi \vee \varphi) \rightarrow (\chi \vee (\forall x)\varphi)$ (x is not free in χ)

(ii) Axiomatic schemata for $\{\exists\}$:

$(\exists_t)$ $\varphi(t) \rightarrow (\exists x)\varphi(x)$ (t substitutable for x in $\varphi(x)$

$(\exists_\rightarrow)$ $(\exists x)(\chi \rightarrow \varphi) \rightarrow (\chi \rightarrow (\exists x)\varphi)$ (x not free in χ)

$(\exists_{\vee})$ $(\exists x)(\chi \vee \varphi) \to (\chi \vee (\exists x)\varphi)$ (x is not free in χ)

(iii) Axiomatic schemata for $\{\forall, \exists\}$:

$(\overrightarrow{\forall\exists})$ $(\forall x)(\varphi \to \chi) \to ((\exists x)\varphi \to \chi)$ (x not free in χ)

(iv) Axiomatic schemata for $\{\triangle\}$:

$(\triangle_{\bar{1}})$ $(\triangle x)\bar{1}$; $(\triangle_r)$ $(\triangle x)\chi \to \chi$ (x is not free in χ)

$(\triangle_s)$ $(\triangle x)\varphi(x) \to (\triangle y)\varphi(y)$ (y substitutable for x in $\varphi(x)$)

$(\triangle_{\to})$ $(\triangle x)(\chi \to \varphi) \to (\chi \to (\triangle x)\varphi)$ (x not free in χ)

$(\triangle_{\vee})$ $(\triangle x)(\chi \vee \varphi) \to (\chi \vee (\triangle x)\varphi)$ (x is not free in χ)

(v) Each axiomatic schemata for $\{\nabla\}$ is obtained by replacing any occurrence of $\triangle$ in any given axiomatic schema for $\triangle$ with ∇.

(vi) Axiomatic schemata for $\{\triangle, \nabla\}$:

$(\overrightarrow{_{\triangle\nabla}})$ $\vdash (\triangle x)(\varphi \to \chi) \to ((\nabla x)\varphi \to \chi)$ (x not free in χ)

$(\overrightarrow{_{\nabla\triangle}})$ $\vdash (\nabla x)(\varphi \to \chi) \to ((\triangle x)\varphi \to \chi)$ (x not free in χ)

(vii) Axiomatic schemata for $\{\forall, \triangle\}$:

$(\overrightarrow{_{\forall\triangle}})$: $(\forall x)(\varphi(x) \to \psi(x)) \to ((\triangle x)\varphi(x) \to (\triangle x)\psi(x))$

(viii) The rule of inference in common to any quantifier in question is modus ponens (MP for short), i.e., from φ and $\varphi \to \psi$ infer ψ.

The rules for $\triangle$ are $\triangle$-Generalization ($\triangle$-G for short) which from $\varphi \to \psi$ infer $(\triangle x)\varphi \to (\triangle x)\psi$, and $\triangle$-Objectification ($\triangle$-O for short) which from $(\triangle x)\varphi \to (\triangle x)\psi$ infer $\varphi \to \psi$.

The rule for ∇ is ∇-Generalization (∇-G for short) which from $\varphi \to \psi$ infer $(\nabla x)\varphi \to (\nabla x)\psi$, and ∇-Objectification (∇-O for short) which from $(\nabla x)\varphi \to (\nabla x)\psi$ infer $\varphi \to \psi$.

The rule for $\forall$ is $\forall$-Generalization ($\forall$-G for short) which from φ infer $(\forall x)\varphi$. For arbitrary theory T, $\triangle$-O or ∇-O can not apply to the case of $T \neq \varnothing$, i.e., we can not infer $\varphi \to \psi$ from $T \vdash (\triangle x)\varphi \to (\triangle x)\psi$ or $T \vdash (\nabla x)\varphi \to (\nabla x)\psi$ if $T \neq \varnothing$.

Definition 16. Let C be a schematic extension of MTL and $Q \subseteq \{\forall, \exists, \triangle, \nabla\}$. We associate with C the corresponding predicate calculus C_Q over a given predicate language L_Q by taking all formulas resulting from the axioms of C by substituting arbitrary formulas of C_Q for propositional variables and the axioms and rules for quantifiers from Q as logical axioms or rules.

From above, we define $\mathrm{MTL}_{\{\forall\}}$, $\mathrm{MTL}_{\{\exists\}}$, $MTL_{\{\triangle\}}$, $\mathrm{MTL}_{\{\triangledown\}}$, $\mathrm{MTL}_{\{\forall,\exists\}}$, $\mathrm{MTL}_{\{\triangle,\triangledown\}}$, $\mathrm{MTL}_{\{\forall,\exists,\triangle\}}$, $\mathrm{MTL}_{\{\forall,\exists,\triangledown\}}$, $\mathrm{MTL}_{\{\forall,\triangledown,\triangle\}}$, $\mathrm{MTL}_{\{\exists,\triangle,\triangledown\}}$ and $\mathrm{MTL}_{\{\forall,\exists,\triangle,\triangledown\}}$ so on.

Theorem 1. MTL-$\{\triangle\}$ proves the followings:

($\triangle_1$) $(\triangle x)\varphi(x) \equiv (\triangle y)\varphi(y)$ (y substitutable for x in $\varphi(x)$),

($\triangle_2$) $\varphi \vdash (\triangle x)\varphi$, ($\triangle_3$) $(\triangle x)\chi \equiv \chi$ (x is not free in χ)

($\triangle_4$) $(\triangle x)(\chi \vee \varphi) \equiv (\chi \vee (\triangle x)\varphi)$ (x is not free in χ)

($\triangle_5$) $(\triangle x)(\chi \wedge \varphi) \rightarrow \chi \wedge (\triangle x)\varphi$ (x is not free in χ)

($\triangle_6$) $(\triangle x)\varphi \,\&\, \chi \rightarrow (\triangle x)(\varphi \,\&\, \chi)$ (x is not free in χ)

($\triangle_7$) $((\triangle x)(\chi \rightarrow \varphi))^n \rightarrow (\triangle x)(\chi \rightarrow \varphi)^n$ (x not free in χ)

Proof. the proof is omitted by the limitation on length of papers. □

Remark 1. ($\triangle_7$) shows the followings are theorems of MTL$\forall$:

(i) $((\forall x)(\chi \rightarrow \varphi))^n \rightarrow (\forall x)(\chi \rightarrow \varphi)^n$ (x not free in χ)

(ii) $((\exists x)(\chi \rightarrow \varphi))^n \rightarrow (\exists x)(\chi \rightarrow \varphi)^n$ (x not free in χ)

5 Semantics of L_Q

In this section, let C be a schematic extension of MTL and $\mathbf{L}$ be a linearly ordered C-algebra, $Q \subseteq \{\forall,\exists,\triangle,\triangledown\}$ and L_Q be a language and T be a theory over C_Q. Notice that the proofs of propositions are omitted by the limitation on length of papers.

Definition 17. An $\mathbf{L}$-structure $\mathbf{I}$ of L_Q is a quadruple $(D_\mathbf{I}, P_\mathbf{I}, C_\mathbf{I}, \triangle_\mathbf{I})$, where (i) $D_\mathbf{I} \neq \emptyset$; (ii) $P_\mathbf{I} : L_p \rightarrow \bigcup_{n=1}^{\infty} \mathbf{L}^{D^n}$, where for all $P_j^k \in L_p$,, $P_\mathbf{I}(P_j^k) \in \mathbf{L}^{D^k}$ (iii) $C_\mathbf{I} : L_c \rightarrow D$, (iv) $\triangle_\mathbf{I} = (D, \wp, m)$ is an $\mathbf{L}$-fuzzy measure space. For simplicity, we use $\triangle_\mathbf{I}$ to denote m in $\triangle_\mathbf{I}$, $P_\mathbf{I}$ to denote $P_\mathbf{I}(P)$ for each predicate P and $c_\mathbf{I}$ to denote $C_\mathbf{I}(c)$ for each constant c.

Definition 18. An $\mathbf{I}$-evaluation v in an $\mathbf{L}$-structure $\mathbf{I}$ is a mapping from L_t to $D_\mathbf{I}$ such that $v(c) = c_\mathbf{I}$ for all $c \in L_c$. Two $\mathbf{I}$-evaluations v and v' are x-equivalent (denoted by $v \equiv_x v'$) if $v(y) = v'(y)$ for each variable y distinct from x.

The value of a term under an $\mathbf{I}$-evaluation v is defined by $|x|_{\mathbf{I},v} = v(x)$ and $|c|_{\mathbf{I},v} = c_{\mathbf{I}}$. The truth value $|\varphi|^{\mathbf{L}}_{\mathbf{I},v}$ of a $wf.\varphi$ under an $\mathbf{I}$-evaluation v is defined as follows:

(i) $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = r_{P^k_j}(v(t_1),\cdots,v(t_k))$ if φ is atomic formula $P^k_j(t_1,\cdots,t_k)$

(ii) If $\varphi = \psi \to \chi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = |\psi|^{\mathbf{L}}_{\mathbf{I},v} \Rightarrow |\chi|^{\mathbf{L}}_{\mathbf{I},v}$

(iii) If $\varphi = \psi \,\&\, \chi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = |\psi|^{\mathbf{L}}_{\mathbf{I},v} * |\chi|^{\mathbf{L}}_{\mathbf{I},v}$

(iv) If $\varphi = \psi \wedge \chi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = \min(|\psi|^{\mathbf{L}}_{\mathbf{I},v}, |\chi|^{\mathbf{L}}_{\mathbf{I},v})$

(v) $|\overline{0}|^{\mathbf{L}}_{\mathbf{I},v} = 0$, $|\overline{1}|^{\mathbf{L}}_{\mathbf{I},v} = 1$,

(vi) If $\varphi = (\forall x)\psi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = \inf_{w \equiv_x v} |\psi|^{\mathbf{L}}_{\mathbf{I},w}$

(vii) If $\varphi = (\exists x)\psi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = \sup_{w \equiv_x v} |\psi|^{\mathbf{L}}_{\mathbf{I},w}$

(viii) If $\varphi = (\vartriangle x)\psi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = \sup_{\in \mathbf{L}} * \overline{\vartriangle}(\{w(x) : w \equiv_x v, |\psi|^{\mathbf{L}}_{\mathbf{I},w} \geq \})$

(ix) If $\varphi = (\triangledown x)\psi$ then $|\varphi|^{\mathbf{L}}_{\mathbf{I},v} = \sup_{\in \mathbf{L}} * \neg\overline{\vartriangle}(\{w(x) : w \equiv_x v, |\psi|^{\mathbf{L}}_{\mathbf{I},w} < \})$.

Definition 19. An $\mathbf{L}$-structure $\mathbf{I} = (D_{\mathbf{I}}, P_{\mathbf{I}}, C_{\mathbf{I}}, \vartriangle_{\mathbf{I}})$ of L_Q is safe if $|\varphi|^{\mathbf{L}}_{\mathbf{I},v}$ is defined for all φ and v. The truth value of a $wf.$ φ in a safe $\mathbf{L}$-structure $\mathbf{I}$ is $|\varphi|^{\mathbf{L}}_{\mathbf{I}} = \inf\{|\varphi|^{\mathbf{L}}_{\mathbf{I},v} : v$ is an $\mathbf{I}$-evaluation in $\mathbf{I}$ }.

Definition 20. Let $\mathbf{I} = (D_{\mathbf{I}}, P_{\mathbf{I}}, C_{\mathbf{I}}, \vartriangle_{\mathbf{I}})$ be an arbitrary safe $\mathbf{L}$-structure of L_Q and T be a theory of C_Q. $\mathbf{I}$ is an $\mathbf{L}$-model of T if $|\varphi|^{\mathbf{L}}_{\mathbf{I}} = 1$ for each φ in T. A $wf.$ φ is an $\mathbf{I}$-tautology of T in an $\mathbf{L}$-model $\mathbf{I}$ of T (denoted by $T \models_{\mathbf{I}} \varphi$) if $|\varphi|^{\mathbf{L}}_{\mathbf{I}} = 1$. A $wf.$ φ is an $\mathbf{L}$-tautology of T (denoted by $T \models_{\mathbf{L}} \varphi$) if $T \models_{\mathbf{I}} \varphi$ for every $\mathbf{L}$-model $\mathbf{I}$ of T. A $wf.$ φ is a tautology of T (denoted by $T \models \varphi$) if $T \models_{\mathbf{L}} \varphi$ for every linearly ordered C-algebra $\mathbf{L}$.

Proposition 2. Let $\{x_i\}_{i\in I}, \{x_j\}_{j\in J} \subseteq \mathbf{L}$ and $a \in \mathbf{L}$. then

(1) $a \wedge (\vee_{i\in I} x_i) = \vee_{i\in I}(a \wedge x_i)$, (2) $a \vee (\vee_{i\in I} x_i) = \vee_{i\in I}(a \vee x_i)$,

(3 $a \wedge (\wedge_{i\in I} x_i) = \wedge_{i\in I}(a \wedge x_i)$, (4) $a \wedge (\vee_{i\in I} x_i) = \vee_{i\in I}(a \wedge x_i)$,

(5) $a * (\vee_{i\in I} x_i) = \vee_{i\in I}(a * x_i)$, (6) $(\vee_{i\in I} x_i) \Rightarrow a = \wedge_{i\in I}(x_i \Rightarrow a)$,

(7) $(\vee_{j\in J} y_j) * (\vee_{i\in I} x_i) = \vee_{i\in I, j\in J}(x_i * y_j)$.

Proposition 3. Let $\varphi(x)$ be a wf. in which x appears free and y be other than x and substitutable for x in $\varphi(x)$. $\mathbf{I}=(D_\mathbf{I},P_\mathbf{I},C_\mathbf{I},\triangle_\mathbf{I})$ be an $\mathbf{L}$-structure of L_Q and v,v',v'' be $\mathbf{I}$-evaluations and $v'\equiv_x v, v''\equiv_y v$ and $v''(y)=v'(x)$. Then $|\varphi(x)|^\mathbf{L}_{\mathbf{I},v'}=|\varphi(y)|^\mathbf{L}_{\mathbf{I},v''}$.

Proposition 4. Let φ be a wf. in which x does not appear free, $\mathbf{I}=(D_\mathbf{I},P_\mathbf{I},C_\mathbf{I},\triangle_\mathbf{I})$ be an $\mathbf{I}$-structure of L_Q and v,v' be $\mathbf{I}$-evaluations and $v\equiv_x w$. Then $|\varphi|^\mathbf{L}_{\mathbf{I},v}=|\varphi|^\mathbf{L}_{\mathbf{I},w}$.

Proposition 5. Axiom schemes $(\triangle_{\bar{1}}),(\triangle_r),(\triangle_s),(\triangle_\rightarrow),(\triangle_\vee),(\overline{\triangle\nabla}),(\overline{\nabla\triangle})$ and $(\overline{\forall\triangle})$ are tautologies.

Proposition 6. Let v be an $\mathbf{I}$-evaluation in an arbitrary safe M-structure $\mathbf{I}=(D_\mathbf{I},P_\mathbf{I},C_\mathbf{I},\triangle_\mathbf{I})$ and φ,ψ be wfs.. Then **(i)** $|\psi|^\mathbf{L}_{\mathbf{I},v}\geq|\varphi|^\mathbf{L}_{\mathbf{I},v}*|\varphi\rightarrow\psi|^\mathbf{L}_{\mathbf{I},v}$. Especially, $|\psi|^\mathbf{L}_{\mathbf{I},v}=1$ if $|\varphi|^\mathbf{L}_{\mathbf{I},v}=|\varphi\rightarrow\psi|^\mathbf{L}_{\mathbf{I},v}=1$;**(ii)** $|\psi|^\mathbf{L}_\mathbf{I}\geq|\varphi|^\mathbf{L}_\mathbf{I}*|\varphi\rightarrow\psi|^\mathbf{L}_\mathbf{I}$. Especially, $|\psi|^\mathbf{L}_\mathbf{I}=1$ if $|\varphi|^\mathbf{L}_\mathbf{I}=|\varphi\rightarrow\psi|^\mathbf{L}_\mathbf{I}=1$; **(iii)** $\{\varphi,\varphi\rightarrow\psi\}\models_\mathbf{I}\psi$;
(iv) $\{\varphi\rightarrow\psi\}\models_\mathbf{I}(\triangle x)\varphi\rightarrow(\triangle x)\psi$; **(v)** $\models_\mathbf{L}\varphi\rightarrow\psi$ if $\models_\mathbf{L}(\triangle x)\varphi\rightarrow(\triangle x)\psi$;
(vi) $\models_\mathbf{L}\varphi\rightarrow\psi$ if $\models_\mathbf{L}(\nabla x)\varphi\rightarrow(\nabla x)\psi$.

Theorem 2. (Soundness of provability) Let C be a schematic extension of MTL and $Q\subseteq\{\forall,\exists,\triangle,\nabla\}$ and T be a theory over C_Q, let φ be a formula of T. If $T\vdash\varphi$ (φ is provable in T) then $|\varphi|^\mathbf{L}_\mathbf{I}=1$ for each linearly ordered C-algebra and each $\mathbf{L}$-model $\mathbf{I}$ of T, i.e., $T\models\varphi$.

Proof. This follows by the induction on the length of a proof and Proposition 5 and 6. □

6 Conclusion

Theorem 2 shows that the soundness of provability for C_Q w.r.t each linearly ordered C-algebra, then whether does the completeness of provability for C_Q w.r.t each linearly ordered C-algebra hold? This is a very difficult problem.

References

1. Esteva, F., Godo, L. (2001): Monoidal t-norm based logic: towards a logic for left-continuous t-norms, Fuzzy Sets and Systems, 271-288.
2. Hájek, P. (1998): Metamathematics of Fuzzy Logic, Dordrecht, Kluwer.
3. Liu, Y., Kerre, E.E. (1998): An overview of fuzzy quantifiers (I): Interpretations, Fuzzy Sets and Systems, 1–21.
4. Glockner, I. (2006): Fuzzy Quantifiers: A Computational Theory, Springer,.
5. Sugeno, M., (1974): Theory of fuzzy integrals and its applications, PhD thesis, Tokyo Institute of Technology.
6. Ying, M.S. (2006): Linguistic quantifiers modelled by Sugeno integrals, Artificial Intelligence, 581–606.

Solving Planning Under Uncertainty: Quantitative and Qualitative Approach

Minghao Yin, Jianan Wang, and Wenxiang Gu

School of Computer of Northeast Normal University, Changchun,
China, 130024

Abstract. Classical decision-theoretic planning methods assume that the probabilistic model of the domain is always accurate. We present two algorithms rLAO* and qLAO* in this paper. rLAO* and qLAO* can solve uncertainty Markov decision problems and qualitative Markov decision problems respectively. We prove that given an admissible heuristic function, both rLAO* and qLAO* can find an optimal solution. Experimental results also show that rLAO* and qLAO* inherit the merits of excellent performance of LAO* for solving uncertainty problems.

1 Introduction

In the field of decision-theoretic planning, the theory of Markov decision processes has received much attention as a nature work for modeling and solving complex decision problems [1]. But up to now, researchers have focused on the "classical" models of MDP approach, in which uncertainty in the consequences of the actions are represented with probabilities, and the satisfaction of agents are represented by a numerical, additive utility function. However, when planning modeling experts are modeling the real world problems, transition probabilities for representing the consequences of the actions are not always accurate [5]. Only incomplete quantitative information can be obtained for modeling the uncertainty. Existing work shows that uncertainty is sometimes represented as a set of possible models, each assigned a model probability [17]. This representation can be simplified by assigning each model an equal probability [2, 18]. In this paper, we focus on the method used in [5], representing model uncertainty by allowing each probability in a single model to lie in an interval. On the other hand, sometimes we can only obtain ordinal, qualitative information rather than quantitative information about uncertainty. That's why researchers have advocated several qualitative versions of decision theory [8, 9, 10]. In this paper, we focus on qualitative decision theory frame work based on possibility theory, which gave rise to the definition of the possibilistic Markov decision processes framework [19].

Over the past ten years, approaches to solving MDPs without evaluating complete states have been developed. LAO* algorithm has been proved to be one of the most efficient methods among them [13,15]. Our aim is to extend LAO*'s capability to

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 612–620, 2007.
springerlink.com

solve planning problems under uncertainty with incomplete information. rLAO* and qLAO* are algorithms that can solve MDPs with uncertainty probability and possibilistic MDPs respectively. We prove that given an admissible heuristic function, both rLAO* and qLAO* can find an optimal solution. Experimental results also show that rLAO* and qLAO* inherit the merits of excellent performance of LAO* for solving uncertainty problems.

2 Background

Decision-theoretic planning problems can be formalized into a special class of MDPs called stochastic shortest-path problems [3]. It can be defined as a turple <S,s_0,G,A,T,c>, where S is a finite set of state space; $s_0 \in S$ is an initial state; $G \subseteq S$ is a set of goal states; A is the set of available actions; To each action $a \in A$ applied in state s is assigned a probability distribution p(.|a). Formally, the system's dynamics can be described by the transition function T, defined as: T(s,a,s') = Pr(s_{t+1} = s'| s_t = s, a_t = a); the rewards or the cost of taking action a in state s are denoted by c(s,a). A policy π, applied in the initial state s_0, defines a Markov chain that can be regarded a solution to SSPs. The aim for solving SSPs amounts to finding an optimal, stationary policy, i.e., a function π: $S \rightarrow \prod(A)$ that minimizes the expected cost J* incurred to reach a goal state. Optimal policy can be obtained as the unique solution of the fixed optimal Bellman equation:

$$J(s) = \min_{a \in A} \sum_{s \in S'} T(s,a,s')[c(s,a,s') + J(s')] \tag{1}$$

Now we turn back to Markov decision problems with uncertainty probability. Sometimes, we can only obtain incomplete quantitative information for transition probabilities. We follow the representation method used in [5], thus considering interval-based uncertainty MDPs. MDPs with uncertainty probability can also be defined as a turple <S,s_0,G,A,T,c>, where for lack of information, transition probability in this framework is only known to be in an interval. This leads to two extended version of Bellman function to get the solution to a MDP with uncertainty probability, in the pessimistic and optimistic case respectively.:

$$J(s) = \min_{a \in A} \sum_{s' \in S} T_m(s,a,s')[c(s,a) + J(s')] \,|\, \text{m is the worst model} \tag{2}$$

$$J(s) = \min_{a \in A} \sum_{s' \in S} T_m(s,a,s')[c(s,a) + J(s')] \,|\, \text{m is the best model} \tag{3}$$

In practical problems, especially in Artificial Intelligence applications, sometimes we could obtain ordinal, qualitative information rather than quantitative information about uncertainty. This gave rise to the introduction of qualitative decision theory. Possibilistic Markov decision problems might be the one used most widely among them [8, 9, 10]. In possibilistic MDPs, uncertainty about the effects of an action a is represented by a possibility distribution π: S→(L,>), where L is a bounded, linearly ordered valuation set. [8] have proposed two qualitative decision criteria. In [11] possibilistic qualitative decision theory has been extended to finite horizon, multi-stage decision procedure. [19] extends the framework by admitting stationary, optimal policies in the infinite horizon case. The possibilistic counterpart of Bellman Equation is described as follows equation (4) and (5). Equation (4) indeed corresponds to an optimistic attitude in front of uncertainty, whereas equation (5) is pessimistic (cautious).

$$u_*(s) = \max_{a\in A} \min_{s'\in S} \max\{n(\pi(s,a,s')), u_*(s')\} \tag{4}$$

$$u^*(s) = \max_{a\in A} \max_{s'\in S} \min\{\pi(s,a,s'), u^*(s')\} \tag{5}$$

3 rLAO*

We first discuss how to extend LAO*'s ability for solving MDPs with uncertainty probability. According to equation (2) and (3), the only difference between solving MDPs and MDPs with uncertainty probability is just to find the best/worst model. Algorithm 1 has shown how to find a worst model or best model in equation (2) and (3), namely that how to calculate the value of the state. To make things clear, we shall use an example to illustrate the problem. Suppose the current state is s, via executing action a, the consequent state is s_0, s_1, s_2, with uncertainty probabilities [0.3, 0.4], [0.2, 0.3], [0.4, 0.5] respectively. Suppose the cost of a is 1 and the values of s_0, s_1, s_2 are 5, 10, 9. Note that the probability in this example is uncertainty and the combinations of probabilities of these three transition satisfying constraint (4) are infinite, for example (0.3, 0.3, 0.4), (0.31, 0.29, 0.4). So in our algorithm, we only consider the best case where, in this model, the probability combination will minimize the value of the policy; while in the worst case, the probability combination we adopt will maximize the value. Since J(s) = p_1*5 + p_2*10 + p_3*9 + 1, satisfying p_1+ p_2+ p_3=1. The idea of algorithm 1 is plain and direct, for example, in order to maximize J, we should make p_2 (which value is biggest) big enough, then p_3 (which is only smaller than p_2), then p_1. In order to minimize J, things are just reverse. So in this example, in the best model, the probabilities for p_1, p_2, p_3 are (0.4,0.1,0.5), while in the worst case, they are (0.3, 0.5, 0.2).

Algorithm 1 BestModel (WorstModel)

1. Suppose $R = (s_1^{'},...,s_k^{'})$ is the set of reachable states set by apply a in s. Sort R by its value (topdown for worst model, downtop for best model).bound =1, i=1, let $p_i^{\min}$ and $p_i^{\max}$ denote the smallest and biggest probability for p_i ;
2. While $(bound - p_i^{\min} + p_i^{\max} < 1)$)do
3. $bound \leftarrow bound - p_i^{\min} + p_i^{\max}$;
4. $P_r(s_i^{'}) \leftarrow p_i^{\max}$; i = i+1
5. end while
6. $r = i$, $P_r(s_r^{'}) \leftarrow 1-(bound - p_r^{\min})$

7. for all $i \in \{r+1,...,k\}$ do
8. $P_r(s_r^{'}) \leftarrow p_r^{\min}$
9. end for

To avoid missing relevant states, we adopt the method introduced in (Buffet 2005), to make sure that each state should be assigned a positive probability. Now we describe rLAO* algorithm for solving MDPs with uncertainty probability.

Algorithm 2 rLAO*
1. The explicit graph G initially consists of the initial state S.
2. While the best solution graph has some non-terminal tip state:
   - Expand best partial solution: Expand some non-terminal tip state n of the best partial solution graph and add any new successor states to G. For each new state s' added to G by expanding n, if s' is a goal state then J(s') = 0, else J(s')= h(s)
   - Update state costs and mark best actions:
   - Create a set Z that contains the expanded state and all of its ancestors in the explicit graph along marked action arcs. (i.e., only include ancestor states from which the expanded state can be reached by following the current best solution.)
   - %For optimistic case: for all the states in set Z, Perform value iteration using backup of equation (2) to update their state costs and determine the best action for each state. %For pessimistic case: for all the states in set Z, Perform value iteration using backup of equation (3) to update their state costs and determine the best action for each state.
3. Convergence test: perform value iteration on the states in the best solution graph. Continue until one of the following two conditions is met. (i) If the error bound falls below ε, go to step 4. (ii) If the best current solution graph changes so that it has an unexpanded tip state, go to step 2.
4. Return an optimal solution graph.

Theorem 1. If the heuristic evaluation function h is admissible and Pessimistic value iteration is used to perform the cost revision step of rLAO*, then:

(1) J(s)≤J*(s)for every state s at every point in the algorithm;
(2) J(s) converges to within ε of J*(s) for every state s of the best solution graph, after a finite number of iterations.

Proof. (1) The proof is by induction. Every state $i \in G$ is assigned an initial heuristic cost estimate and J(s)= h(s) ≤J*(s) by the admissibility of the heuristic evaluation function. We make the inductive hypothesis that at some point in the algorithm, J(s)≤J*(s) for every state. If a backup is performed for any state s,

$$J(s) = \max_{m\in \mathrm{M}} \min_{a\in A} \sum_{s'\in S} T_m(s,a,s')[c(s,a) + J(s')]$$

$$\leq \max_{m\in \mathrm{M}} \min_{a\in A} \sum_{s'\in S} T_m(s,a,s')[c(s,a) + J^*(s')] = J^*(s)$$

where the last equation restates the equation (2).

(2)It's obvious that rLAO* terminates after a finite number of iterations if the implicit graph G is finite, or equivalently, the number of states in the MDP with uncertainty probability is finite. Because the graph is finite, rLAO* must eventually find a solution graph that has no non-terminal tip states. Performing Pessimistic value iteration on the states in this solution graph makes the error bound of the solution arbitrary small after a finite number of iterations, by the convergence proof of pessimistic value iteration.

Theorem 2. If the heuristic evaluation function h is admissible and optimistic value iteration is used to perform the cost revision step of rLAO*, then:

(1) J(s)≤J*(s) for every state s at every point in the algorithm;
(2) J(s) converges to within ε of J*(s) for every state s of the best solution graph, after a finite number of iterations.

Proof. The proving procedure is similar to theorem 1.

4 qLAO*

Now we discuss how to extend LAO*'s ability for solving possibilistic MDPs. Notice that quotation (4) and (5) are aim to find an optimal, stationary policy that will maximize (not minimize) the expected cost J* incurred to reach a goal state. Although we can easily change them by exchange the aggregate operator min and max to obtain a possibilistic counterpart of equation (1), we don't do that in order to be compatible with [19]. We slightly change the definition of the value of states, i.e. for s∈G, $J(s)=1_L$, and $J(s) = 0_L$ for other cases. They can be represented as follows:

$$N[Goals \mid \pi_{init:}(a_i)_{i=0}^{N-1}] = \min_{s_0 \in S_i s_N \in Goals} \max(1-\pi_{init}(s_0), 1-\pi[s_N \mid s_0, (a_i)_{i=0}^{N-1}]) \tag{6}$$

$$\Pi[Goals \mid \pi_{init:}(a_i)_{i=0}^{N-1}] = \max_{s_0 \in S_i s_N \in Goals} \min(\pi[s_N \mid s_0, (a_i)_{i=0}^{N-1}], \pi_{init}(s_0)) \tag{7}$$

This falls into possibilistic planning framework introduced in [6] have introduced counterpart versions of Value Iteration algorithm for solving possibilistic MDPs and proved both of the algorithms converge to Q* in a finite number of steps. For the limits of pages, we shall not present Algorithm qLAO* here. Instead we introduce the main idea. qLAO* algorithm differs with LAO* and rLAO* mainly in step 2 and step 3. Because qLAO* only knows qualitative possibilistic information, it relies on a "possibilistic" dynamic programming algorithm, which have been introduced as algorithm 7 and algorithm 8. In this sense, qLAO* also assume that a utility function u on S is given, that express the preference of the agent on the states that the system shall reach and stay in. And qLAO* converges only when the residual is zero. Now we prove qLAO* shares the properties of AO*, LAO*. Given an admissible heuristic function, all state costs in the explicit graph are admissible after each step and qLAO* converges to an optimal solution both in optimistic case and in pessimistic case.

Theorem 3. If the heuristic evaluation function h is admissible and Possibilistic value iteration is used to perform the cost revision step of qLAO*, then:

(1) J(s)≤J*(s) for every state s at every point in the algorithm;
(2) J(s) converges to J*(s) for every state s of the best solution graph, after a finite number of iterations.

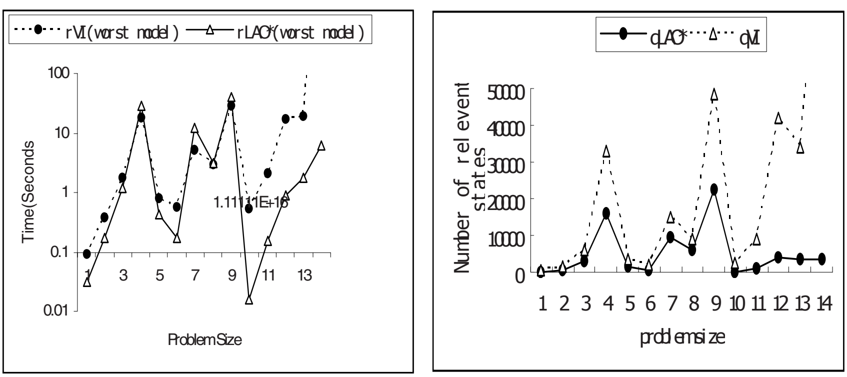
Proof. We only prove the optimal case, for pessimistic case, things are similar.
(1) The proof is by induction. Every state $s \in G$ is assigned an initial heuristic cost estimate and J(s)=h(s) ≤J*(s) by the admissibility of the heuristic evaluation function. We make the inductive hypothesis that at some point in the algorithm, J(s)≤J*(s) for every state. If a backup is performed for any state i,

$$J(s) = \max_{a \in A} \min_{s' \in S} \max\{n(\pi(s,a,s')), J(s')\}$$
$$\leq \max_{a \in A} \min_{s' \in S} \max\{n(\pi(s,a,s')), J^*(s')\} = J^*(s)$$

(2)It's obvious that qLAO* terminates after a finite number of iterations if the implicit graph G is finite, or equivalently, the number of states in the possibilistic MDP is finite. Because the graph is finite, qLAO* must eventually find a solution graph that has no non-terminal tip states. Performing Pessimistic value iteration on the states in this solution graph makes the error bound of the solution to zero after a finite number of iterations, by the convergence proof of possibilistic value iteration.

5 Experimental Results of rLAO* and qLAO*

To evaluate the performance of rLAO* and qLAO*, we integrated them into the LAO* code that is also used in [4]. We examine the performance of LAO* on the racetrack problem used in [1].



(a) (b)

Fig. 1. Comparision of rLAO*, qLAO* and rVI, qVI

We test rLAO* on different kinds of maps. When the race car is driven optimally, it avoids large parts of the track as well as dangerous velocities. Table 1 has shown the comparison results of pessimistic rLAO*, optimistic qLAO*, standard LAO* in terms of running time, convergence expected cost value and number of relevant states. The results have shown that rLAO* inherit the merits of excellent performance of LAO* for solving uncertainty problems. We compare rLAO* with a robust value iteration (VI) algorithm. Note that our algorithm is orders of magnitude faster than VI. Figure 1(a) has shown the experimental results. We also implemented a qualitative version of LAO*. For convenience, we set the preference degree of each state to be 1_L. This means this falls into the frame work of possibilistic planning. To our surprise, it runs very fast, and most of the problem can be solved within 0.1 seconds. The reason is the times for iteration is rather small. In figure 1(b), we can see that qLAO* can remove most nodes compared with qualitative value iteration.

Table 1. Comparison of LAO*, wLAO*, bLAO*. Value, RS, TM denote expected cost value, numbers of relevant states and running time respectively.

	LAO*			wLAO*(worst model)			bLAO*(best model)		
problem	value	RS	TM	value	RS	TM	value	RS	TM
Ring-1	5.43	221	0.03	5.68	221	0.031	5.21	221	0.031
Ring-2	7.70	631	0.172	8.06	631	0.172	7.34	631	0.188
Ring-3	10.38	2814	1.204	10.68	2867	1.218	10.15	2704	1.002
Ring-4	14.96	14593	13.798	15.51	15857	28.473	14.46	14573	13.716
Racetrack1	5.40	1435	0.328	5.62	1429	0.423	5.20	1435	0.391
Racetack3	8.22	658	0.109	8.94	734	0.174	7.57	689	0.11
Racetack4	14.54	8941	10.455	15.28	9529	11.875	13.79	9186	9.906
Racetack5	9.95	892	0.627	11.07	6167	3.128	9.02	1567	0.313
Racetack6	13.66	21285	33.376	14.21	22376	39.389	13.24	21545	34.076
Square-1	4.31	121	0.015	4.48	121	0.016	4.15	121	0.015
Square-2	5.41	810	0.124	5.62	810	0.157	5.20	810	0.125
Square-3	7.51	4041	0.875	7.78	4206	0.894	7.25	3994	0.875
Y-Y	13.76	3280	1.486	14.37	3506	1.717	13.30	3329	1.737
Y-1	13.76	3280	4.611	14.37	3506	6.077	13.30	3329	4.084

6 Conclusions

In this paper, we discuss how to extend LAO*'s ability to solve uncertainty Markov decision problems and qualitative Markov decision problems. We propose two algorithms, namely, rLAO* and qLAO*. Both of these algorithms can find the optimal solution, given an admissible heuristic function. Preliminary results show that these algorithms inherit the merits of excellent performance of LAO* for solving uncertainty problems. Indeed LAO*, rLAO*, qLAO* are indeed complementary rather than opposite. For example, when the agent is put into a totally strange environment, it may only have qualitative information, thus it calls the qLAO* algorithm. After more information is gathered, it may have incomplete quantitative information, then rLAO* can be called. After the agent have total knowledge about the environment, it can use LAO* to guide its navigation. In this sense, rLAO* and qLAO* play important roles to make LAO* applicable.

References

1. Barto, A. G. et al (1995) Learning to act using real-time dynamic programming. Artificial Intelligence *72:81-138*.
2. Bagnell, J. A. et al (2001) Solving uncertainty Markov decision problems. Technical Report CMU-RI-TR-01-25, Robotics Institute, Carnegie Mellon University.
3. Bertesekas, D.P. and Tsitsiklis (1996), J.N. Neurodynamic Programming, Athena Scientific.
4. Blai Bonet and Hector Geffner (2003) Labeled RTDP: Improving the Convergence of Real-Time Dynamic Programming. In Proc. of ICAPS-03. Trento, Italy, AAAI Press, pp 12-21.
5. Buffet, O. (2005), Planning with robust (L)RTDP. In: Proc. of IJCAI-05.
6. C. da Costa Pereira, F. Garcia, J. Lang and R. Martin-Clouaire (1997) Planning with graded nondeterministic actions: a possibilistic approach, International Journal of Intelligent Systems, 12 (11/12), 935-962.
7. C. da Costa Pereira, F. Garcia, J. Lang and R. Martin-Clouaire (1999) Possibilistic planning: representation and complexity. Recent Advances in Planning (Proceedings of ECP'99), Lectures Notes in Artificial Intelligence, Springer Verlag, 1997, 143-155.
8. Dubois, D. and Prade H. (1995) Possibility theory as a basis of qualitative decision theory. In: Proc. of the IJCAI-95
9. Dubois, D. etc (1998) Logical representation and computation of optimal decisions in a qualitative setting. In: proc. of AAAI-98, California, AAAI press.
10. Dubois, D. etc. (2003) Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach, Artificial Intelligence 148: 219-260.
11. Fargier, H. etc. (1998) Towards qualitative approaches to multi-state decision making. International Journal of Approximate Reasoning 19, 441-471.
12. Fargier, H. etc.(2005) Qualitative decision under uncertainty: back to expected utility, Artificial Intelligence 164:245-280.
13. Feng zhengzhu, Hansen, Eric A.(2002) Symbolic Heuristic Search for factored Markov decision processes, In: proceedings of AAAI 2002.
14. Givan, R., etc. (2000) Bounded parameter markov decision processes. Artificial Intelligence, 122(1-2): 71-109.
15. Hansen, Eric A., Zilberstein, S.(2001) LAO*: A heuristic search algorithm that finds solutions with loops, Artificial Intelligence 129: 35-62.
16. Hosaka, M. (2001) etc.: Controlled markov set-chains under average criteria. Applied Mathematics and Computation, 120(1-3):195-209
17. Munos, R.(2001) Efficient resources allocation for markov decision processes. In: Proc. of NIPS 2001.
18. Nilim, A. and Ghaoui, L. El. (2004) Robustness in markov decision problems with uncertainty transition matrices. In: Proc. of NIPS 2004.
19. Sabbadin. R. (2001) Possibilistic markov decision processes. Engineering of Artificial Intelligence 14:287-300.

20. Tan, S. W. etc. (1994) Qualitative decision theory. In: proc of AAAI-94, California, AAAI press.
21. Russell S. and Norvig P (2003) Artificial Intelligence: a modern approach 2nd Edition, Prentice Hall.
22. Blai Bonet and Hector Geffner (2005) An Algorithm Better than AO*? In Proc. of AAAI-05 Pittsburgh, Pennsylvania. AAAI Press. Pages 1343-1348.

The Compositional Rule of Inference and Zadeh's Extension Principle for Non-normal Fuzzy Sets

Pim van den Broek and Joost Noppen

Dept. of Computer Science, University of Twente P.O. Box 217, 7500 AE Enschede, The Netherlands
{pimvdb,noppen}@ewi.utwente.nl

Abstract. Defining the standard Boolean operations on fuzzy Booleans with the compositional rule of inference (CRI) or Zadeh's extension principle gives counter-intuitive results. We introduce and motivate a slight adaptation of the CRI, which only effects the results for non-normal fuzzy sets. It is shown that the adapted CRI gives the expected results for the standard Boolean operations on fuzzy Booleans. As a second application, we show that the adapted CRI enables a don't-care value in approximate reasoning. From the close connection between the CRI and Zadeh's extension principle, we derive an adaptation of the extension principle, which, like the modified CRI, also gives the expected Boolean operations on fuzzy Booleans.

Keywords: compositional rule of inference, extension principle, fuzzy Booleans.

1 Introduction

Fuzzy Booleans are introduced in [1], in analogy with the concept of fuzzy numbers [2], as fuzzy sets over the domain of truth-values {true, false}. A fuzzy Boolean is denoted as (a,b), where a and b are numbers from the interval [0,1], a shorthand for the conventional notation "a/true + b/false". The truth-values 'true' and 'false' are represented by (1,0) and (0,1) respectively. The advantage of fuzzy booleans is that they allow fuzzy reasoning with concepts 'contradiction' and 'undefined'. For instance, when interpreted as possibilities, (1,1) is 'undefined', and (0,0) is 'contradiction'; when interpreted as necessities, this is just the other way around.

Let AND and OR be the fuzzy equivalents of the crisp Boolean operators and and or, respectively. They may be defined by means of approximate reasoning. For instance, AND is defined by the following four fuzzy rules:

$$\text{IF } X_1=(1,0) \text{ AND } X_2=(1,0) \text{ THEN } Y=(1,0) \quad (1a)$$
$$\text{IF } X_1=(1,0) \text{ AND } X_2=(0,1) \text{ THEN } Y=(0,1) \quad (1b)$$
$$\text{IF } X_1=(0,1) \text{ AND } X_2=(1,0) \text{ THEN } Y=(0,1) \quad (1c)$$
$$\text{IF } X_1=(0,1) \text{ AND } X_2=(0,1) \text{ THEN } Y=(0,1)\ . \quad (1d)$$

With approximate reasoning, using the CRI, we obtain

$$\text{AND } ((a,b),(c,d)) = (\min(a,c), \max\{\min(a,d), \min(b,c), \min(b,d)\})\ . \quad (2)$$

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 621–628, 2007.
springerlink.com

Using the abbreviations T = (1,0) (True), F = (0,1) (False), C = (1,1) (Contradiction) and U = (0,0) (Unknown), we obtain for AND:

AND	T	F	C	U
T	T	F	C	U
F	F	F	F	U
C	C	F	C	U
U	U	U	U	U .

This is not in accordance with our intuitive understanding of the AND-operation. For instance, AND (F,U) = U, where we would expect AND (F,U) = F. Indeed, since U is the empty set, the CRI will give U whenever one of the arguments of AND is U. The same thing happens when we define AND with Zadeh's extension principle: whenever one of the arguments of AND is U, the result is U.

The aim of this paper is to introduce and motivate a slight adaptation of the CRI, which only effects the results for non-normal fuzzy sets, and show that the adapted CRI gives the expected results for the standard Boolean operations on fuzzy Booleans. As a second application, we will show that the adapted CRI enables a don't-care value in approximate reasoning. From the close connection between the CRI and Zadeh's extension principle, we will derive an adaptation of the extension principle, which, like the modified CRI, also gives the expected Boolean operations on fuzzy Booleans.

This paper is organised as follows. In the next section, we will introduce the adaptation of the CRI. In section 3 we describe approximate reasoning with the adapted CRI. In section 4 we show that the adapted CRI gives results for the standard Boolean operations on fuzzy Booleans which are in accordance with our intuitive understanding of the AND-operation. In section 5 we show that the adapted CRI enables a don't-care value in approximate reasoning. In section 6 we derive a adaptation of the extension principle, and show that with the adapted extension principle we also obtain the expected Boolean operations on fuzzy Booleans. Section 7 concludes the paper.

2 Adaptation of the Compositional Rule of Inference

Given a fuzzy set A on a domain U and a fuzzy relation R on the domain U⊗V, the CRI gives the fuzzy set B on V which is given by

$$B(v) = \sup_u \min (A(u), R(u,v)) . \tag{3}$$

If A is non-normal, B is non-normal as well. If A is empty, B is empty as well. If A is the crisp singleton set containing only u_0, then $B(v) = R(u_0,v)$. So, for each crisp singleton set A

$$B(v) >= \inf_u R(u,v) . \tag{4}$$

Our adaptation of the CRI is such that eq. (4) holds for every fuzzy set A. Instead of eq. (3), we thus propose

$$B(v) = \sup_u \max(\inf_u R(u,v), \min (A(u), R(u,v)) . \tag{5}$$

This adaptation can only give a different result when A is non-normal. Indeed, if $A(u_0) = 1$ for some u_0 in U then $\sup_u \min(A(u), R(u,v)) >= \min(A(u_0), R(u_0,v)) = R(u_0,v) >= \inf_u R(u,v)$, and so eq. (5) reduces to eq. (3). Also, when $R(u,v) = 0$ for some u in U there is no difference between eq. (5) and eq. (3).

Next we will define how to adapt a composition of two applications of the CRI as in

$$B(v) = \sup_{(u1,u2)} \min(A_1(u_1), \min(A_2(u_2), R((u_1,u_2),v))) . \tag{6}$$

Here it is not appropriate to adapt this in a single step by first writing this as

$$B(v) = \sup_u \min(A_1 \otimes A_2 (u), R(u,v))) \tag{7}$$

where $A_1 \otimes A_2$ is the Cartesean product of A_1 and A_2 and $u = (u_1,u_2)$. Instead, the adaptation should be applied twice, which leads to

$$B(v) = \sup_{(u1,u2)} \max(\inf_{u1} \max(\inf_{u2} R((u_1,u_2),v), \min(A_2(u_2), R((u_1,u_2),v))), \min(A_1(u_1), \max(\inf_{u2} R((u_1,u_2),v), \min(A_2 (u_2), R((u_1,u_2),v))))) . \tag{8}$$

3 Approximate Reasoning with the Adapted CRI

Consider the fuzzy rule

$$\text{IF } X = A' \text{ THEN } Y = B' . \tag{9}$$

Given the fact X = A, approximate reasoning with the standard CRI gives Y = B, where B is given by eq. (3), and R(u,v) is given by

$$R(u,v) = Q (A'(u), B'(v)) . \tag{10}$$

In case of approximate reasoning with the interpolation method [4], the operator Q is a t-norm; the most commonly used t-norm is the minimum operator. In case of approximate reasoning with the implication method [3], the operator Q is an implication operator.

In case of multiple fuzzy rules, one calculates a relation as in eq. (10) for each fuzzy rule, and then aggregates the results. Aggregation is done with the maximum operator in the interpolation method, and with the minimum operator in the implication method. The resulting aggregated relation is then used to calculate the inference results with eq. (3). Using the adapted CRI means that eq. (5) should be used instead of eq. (3).

Consider next the fuzzy rule with two antecedents

$$\text{IF } X_1 = A'_1 \text{ AND } X_2 = A'_2 \text{ THEN } Y = B' . \tag{11}$$

With the standard CRI we have, given the input $X_1 = A_1$ AND $X_2 = A_2$, the output Y = B, where B is given by eq. (6), and

$$R((u_1,u_2),v) = Q (\min (A'_1(u_1), A'_2(u_2)), B'(v)) \tag{12}$$

In case of multiple fuzzy rules, one calculates a relation as in eq. (12) for each fuzzy rule, and then aggregates the results. The resulting aggregated relation is then used to calculate the inference results with eq. (6). Using the adapted CRI means that eq. (8) should be used instead of eq. (6). Generalisation to three or more antecedents is straightforward.

Note that we described here the FATI (first aggregate, then inference) approach. In Mamdani's original approach [4], the interpolation method with the minimum operator as t-norm, the FITA (first inference, then aggregate) approach is used. Indeed, it happens that the inference results of FATI and FITA are the same in this case. This is no longer true when the adapted CRI is used. So, with the adapted CRI, one should always adopt the FATI approach.

4 Application to Fuzzy Booleans

In this section we will use the results of the previous section to compute the inference results for the four fuzzy rules of eq. (1) with the adapted CRI.

First we compute the four relations R_1, R_2, R_3 and R_4 for the fuzzy rules of eqs. (1a,1b,1c,1d) respectively, and their aggregation R in the interpolation method, using eq. (12):

	((t,t),t)	((t,f),t)	((f,t),t)	((f,f),t)	(t,t),f)	((t,f),f)	((f,t),f)	((f,f),f)
R_1	1	0	0	0	0	0	0	0
R_2	0	0	0	0	0	1	0	0
R_3	0	0	0	0	0	0	1	0
R_4	0	0	0	0	0	0	0	1
R	1	0	0	0	0	1	1	1

Here we used the abbreviations t and f for true and false, respectively.

Next we compute the four relations R_1, R_2, R_3 and R_4 and their aggregation R in the implication method:

	((t,t),t)	((t,f),t)	((f,t),t)	((f,f),t)	(t,t),f)	((t,f),f)	((f,t),f)	((f,f),f)
R_1	1	1	1	1	0	1	1	1
R_2	1	0	1	1	1	1	1	1
R_3	1	1	0	1	1	1	1	1
R_4	1	1	1	0	1	1	1	1
R	1	0	0	0	0	1	1	1

Note that the aggregated relation R is the same for both methods, and is independent of Q. This is in accordance with in general result in [1], where it is proved that this is always the case when the set of fuzzy rules is a complete set of fuzzy rules with crisp antecedents.

Substituting this relation in eq. (8) now gives

$$\text{AND}\ ((a,b),(c,d)) = (\min(a,c),\ \max\ (b,d)) \tag{13}$$

whereas with the standard CRI (eq. (6)) we would have obtained eq.(2). We can verify that eq. (13) is in accordance with our intuitive understanding of the AND-operation. Indeed, eq. (13) just says that the "trueness" of AND P Q is the trueness of both P and Q, and the "falseness" of AND P Q is the falseness of either P or Q. This should be compared with eq. (2), where the falseness of P does not imply the falseness of AND P Q; the falseness of AND P Q follows only if in addition to the falseness of P we also have either the trueness or the falseness of Q. The table in the introduction is replaced by

AND	T	F	C	U
T	T	F	C	U
F	F	F	F	F
C	C	F	C	F
U	U	F	F	U .

In the same way, we obtain the expression

$$\text{OR}\ ((a,b),(c,d)) = (\max\ (a,c), \min\ (b,d)) \tag{14}$$

and the table

OR	T	F	C	U
T	T	T	T	T
F	T	F	C	U
C	T	C	C	T
U	T	U	T	U

which are in accordance with our intuitive understanding of the OR-operation. Finally, the NOT-operation, given by NOT (a,b) = (b,a), is not affected by our adaptation of the CRI.

5 Don't-Care Value in Approximate Reasoning

As a second application of the adapted CRI, we will show in this section that with the adapted CRI there exists a don't-care value in approximate reasoning. Consider first the fuzzy rule of eq. (9). Given the fact X = A, approximate reasoning with the CRI gives Y = B, where B is given by

$$B(v) = \sup_u \min\ (A(u), Q\ (A'(u),B'(v)))\ . \tag{15}$$

We will consider first the interpolation method, i.e. Q is a t-norm. The fuzzy set A' is a don't-care value if B(v) = B'(v) for all v in V and all fuzzy sets A. Since B(v) <= B'(v) for all v in V, and B(v) increases if A' increases, the best value for A' is the universe U itself, i.e A'(u) = 1 for all u in U. Then eq. (15) becomes

$$B(v) = \sup_u \min\ (A(u), B'(v)) \tag{16}$$

which means that we have the desired property only if we restrict the input A to be normal. So, a don't-care value does not exist.

The inference result of the fuzzy rule in eq. (9) with the adapted CRI is

$$B(v) = \sup_u \max (\inf_u (Q (A'(u),B'(v))), \min (A(u), Q (A'(u),B'(v)))) . \quad (17)$$

Substituting A'(u) = 1 for all u in U gives B(v) = B'(v), which shows that with the adapted CRI A' has the required property, even for non-normal input A. Therefore, the universe U can be taken as don't-care value.

Consider next the following two fuzzy rules:

$$\text{IF } X_1 = A_1 \text{ THEN } Y = B_1 \quad (18a)$$

$$\text{IF } X_2 = A_2 \text{ THEN } Y = B_2 \quad (18b)$$

where the domains of X_1,X_2 and Y are U_1,U_2 and V respectively. For instance when one wants to compile a single relation for both fuzzy rules, one would like to write this as

$$\text{IF } X_1 = A_1 \text{ AND } X_2 = DC_2 \text{ THEN } Y = B_1 \quad (19a)$$

$$\text{IF } X_1 = DC_1 \text{ AND } X_2 = A_2 \text{ THEN } Y = B_2 \quad (19a)$$

where DC denotes a don't-care fuzzy set, i.e. the results for the fuzzy rules in eq. (19) should be the same as the results for the fuzzy rules in eq. (18). As above, such a don't-care fuzzy set does not exist. It is however a straightforward exercise to verify, by using the inference results with the adapted CRI (eqs. (5,8)), that the fuzzy rule

$$\text{IF } X_1 = A_1 \text{ AND } X_2 = U_2 \text{ THEN } Y = B \quad (20)$$

gives the same result as the rule

$$\text{IF } X_1 = A_1 \text{ THEN } Y = B \quad (21)$$

and that the rule

$$\text{IF } X_1 = U_1 \text{ AND } X_2 = A_2 \text{ THEN } Y = B \quad (22)$$

gives the same result as

$$\text{IF } X_2 = A_2 \text{ THEN } Y = B \quad (23)$$

showing that in case of fuzzy rules with two antecedents the universe can be taken as a don't-care fuzzy set. Generalisation to three or more antecedents is straightforward.

Next we consider the implication method. Then Q, in eq. (15), is an implication operator. Analogously to the reasoning above, we find that the fuzzy set A' with A'(u) = 0 for all u in U, i.e. the empty set, is a don't-care value.

So the conclusion of this section is that with the adapted CRI there exists a don't-care value for approximate reasoning; this don't-care value is the universe in case of the interpolation method and it is the empty set in case of the implication method.

6 Adaptation of Zadeh's Extension Principle

Zadeh’s extension principle, developed by Zadeh [6] and elaborated by Yager [5], extends functions from their domain to fuzzy sets on their domain. Let f be a function from the universe U onto the universe V. By Zadeh's extension principle, f maps each fuzzy set A on U onto a fuzzy set B on V which is given by

$$B(v) = \sup_{u:f(u)=v} A(u) \ . \tag{24}$$

There exists an intimate relation between the extension principle and the CRI: When the relation R is defined by

$$\forall u \in U : R\ (u,f(u)) = 1 \tag{25a}$$

$$\forall u \in U\ \forall v \in V : v \neq f(u) \Rightarrow R\ (u,v) = 0 \tag{25b}$$

then eq. (24) is an immediate consequence of eq. (3).

We derive the adapted extension principle in the same way from the adapted CRI. When the relation of eq. (25) is substituted in eq. (5) we find

$$B(v) = 1, \text{ if } \forall u \in U\text{: } f(u) = v \tag{26a}$$

$$B(v) = \sup_{u:f(u)=v} A(u), \text{ otherwise .} \tag{26b}$$

So there is a difference with the standard extension principle only in case f is a constant function. Then the membership value of f(u) is equal to 1, while it is equal to $\sup_u A(u)$ according to the standard extension principle. So, if A is normal there is no difference. We feel that this adaptation makes sense; indeed, where f maps each crisp element u of U to the same element v of V, there is no doubt that a fuzzy set on U should be mapped to the singleton set containing v.

This adaptation holds for functions with a single argument. In the case where f is a function with two arguments, whose domains are U_1 and U_2 respectively, we derive the adapted extension principle from eq.(8). We find that

$$f(A_1,A_2)(v) = \sup_{(u1,u2):f(u1,u2)=v} C(u_1,u_2) \tag{27a}$$

$$C(u_1,u_2) = 1, \text{ if } \forall u_3 \in U_1,\ \forall u_4 \in U_2\text{: } f(u_3,\ u_4) = v \tag{27b}$$

$$C(u_1,u_2) = \max(A_1(u1),\ A_2(u_2)), \quad \text{if } \forall u_4 \in U_2\text{: } f(u_1,u_4) = v\ \&\ \forall u_3 \in U_1\text{: } f(u_3,u_2) = v \tag{27c}$$

$$C(u_1,u_2) = A_1(u_1), \text{ if } \forall u_4 \in U_2\text{: } f(u_1,u_4) = v \tag{27d}$$

$$C(u_1,u_2) = A_2(u_2), \text{ if } \forall u_3 \in U_1\text{: } f(u_3,u_2) = v \tag{27e}$$

$$C(u_1,u_2) = \max(A_1(u_1),\ A_2(u_2)), \text{ otherwise .} \tag{27f}$$

Here the conditions of the five clauses for $C(u_1,u_2)$ should be checked from above.

As an example, let us again compute AND ((a,b),(c,d)). From eq. (25) we find that

$$\text{AND } ((a,b),(c,d))\ (\text{true}) = C(\text{true},\text{true}) \tag{28a}$$

$$\text{AND } ((a,b),(c,d))\ (\text{false}) = \max\ (C(\text{true},\text{false}),\ C(\text{false},\text{true}),\ C(\text{false},\text{false})) \tag{28b}$$

where C is given by

$$C(\text{true},\text{true}) = \min\ ((a,b)(\text{true}),\ (c,d)(\text{true})) = \min\ (a,c) \tag{29a}$$

$$C(\text{true},\text{false}) = (c,d)\ (\text{false}) = d \tag{29b}$$

$$C(false,true) = (a,b)\ (false) = b \tag{29c}$$

$$C(false,false) = max((a,b)(false), (c,d)(false) = max\ (b,d) \tag{29d}$$

which leads to

$$AND\ ((a,b),(c,d)) = (min\ (a,c), max\ (b,d)), \tag{30}$$

which is, of course, the same result as the one obtained in section 4.

7 Conclusion

We have defined adaptations for the compositional rule of inference and Zadeh's extension principle, which have effect in case of non-normal fuzzy sets. We have demonstrated the usefulness of the adaptations by means of two applications. Firstly, we have shown that we obtain the expected results for the standard Boolean operations on fuzzy Booleans. Secondly, we have shown that we obtain a don't-care value in approximate reasoning. Of course, from these two applications it cannot (yet) be concluded that our adaptations should replace the standard CRI and extension principle. Therefore, it is interesting to examine our adaptations in other applications where non-normal fuzzy sets are used.

References

1. van den Broek, P.M., Noppen, J.A.R.: Approximate Reasoning with Fuzzy Booleans, IPMU 2004, 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems 1429–1434. Editrice Universita La Sapienza, Italy (2004). ISBN 88-87242-54-2
2. Dubois, D., Prade, H.: (1987). Fuzzy numbers: an overview. In: Bezdek, J.C. (ed.): Analysis of Fuzzy Information - Vol. 1 : Mathematics and Logic. CRC Press, Boca Raton, FL 3-39.
3. Klir, G.J., Yuan, B. (1995): Fuzzy sets and fuzzy logic, theory and applications. Prentice Hall
4. Mamdani, E.H., Assilian, S. (1975): An experiment in linguistic synthesis with a fuzzy logic controller. Intern. J. of Man-Machine Studies 7 1–13.
5. Yager, R.R. (1986): A characterization of the extension principle. Fuzzy Sets and Systems 18 205-217.
6. Zadeh, L. (1975): The concept of a linguistic variable and its application to approximate reasoning. Information Sciences 8 199-251.

Satisfiability in a Linguistic-Valued Logic and Its Quasi-horn Clause Inference Framework

Jun Liu[1], Luis Martinez[2], Yang Xu[3], and Zhirui Lu[1]

[1] School of Computing and Mathematics, University of Ulster at Jordanstown, Newtownabbey BT37 0QB, Northern Ireland, UK
{j.liu, lu-z}@ulster.ac.uk
[2] Deparment of Computer Science, University of Jaén, E-23071 Jaén, Spain
martin@ujaen.es
[3] Department of Applied Mathematics, Southwest Jiaotong University, Chengdu 610031, Si chuan, P.R. China
xuyang@swjtu.edu.cn

Abstract. In this paper, we focus on the linguistic-valued logic system with truth-values in the lattice-ordered linguistic truth-valued algebra, then investigate its satisfiability problem and its corresponding Quasi-Horn-clause logic framework, while their soundness and completeness theorems are provided. The present framework reflects the symbolic approach acts by direct reasoning on linguistic truth values, i.e., reasoning with words, and provides a theoretical support for natural-language based reasoning and decision making system.

1 Introduction

Resolution principle is a single rule of inference for a test of unsatisfiability of a logical formula. Since its introduction in 1965 [1], automated reasoning based on Robinson's resolution rule has been extensively studied in the context of finding natural and efficient proof systems to support a wide spectrum of computational tasks. As the use of non-classical logics becomes increasingly important in computer science and artificial intelligence, the development of efficient automated reasoning on non-classical logic is currently an active area of research. e.g., on fuzzy logic, among others, see [2]-[7] and on many-valued logic, among others, see [8]-[21].

Lattice-valued logic is an important many-valued logic. In this paper, we extend the resolution principle from two-valued logic into lattice-valued logic with truth-value in a lattice-ordered logical algebraic structure - Residuated Lattice (RL). RL structure is a very popular algebra for inexact concepts as shown by Goguen in [22]. There have been considerable efforts about many-valued logic based on a RL, where Pavelka [23] and Novak [24] systematically discussed propositional and first-order calculi with truth values in an enriched RL. Although there are some important investigations in [17], [19-21] among others, up to now, the study of proving systems and automated reasoning based on RL, has not been extensively reported.

In this paper, we focus on a lattice-valued logic [19-20], [25] with truth-value in a LIA [25-26], which is a kind of RLs established by combining the lattice and

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 629–639, 2007.
springerlink.com

implication algebra. This kind of lattice-valued logics are an extension of classical logic in several aspects such as connectives, truth-valued field and inference rules, which includes Łukasiewicz logic $L[0, 1]$ with truth-values in [0, 1] as a special case. Especially their implication connectives are more general and not reducible to the other classical connectives (like ~, ∧ and ∨), unlike the Kleene implication ($p \to q = \sim p \vee q$, ~ is the negation, it implies that the formulae of its logic are syntactically equivalent to those in classical logic). This irreducibility, though semantically justifiable, complicates the calculus.

On the other hand, we all know that as human beings we are bound to express ourselves in a natural language that uses words. The meanings of words are inherently, imprecise, vague, and fuzzy, mostly can be very qualitative in nature. Hence, it is necessary to investigate natural language based reasoning within the realm of AI. A nice feature of linguistic variables is that their values are structured, which makes it possible to compute the representations of composed linguistic values from those of their composing parts. In order for linguistic variables to be useful tools of analysis, one ought to be able to manipulate them through various operations so that one can directly symbolically manipulate the linguistic variables themselves.

Accordingly, in the present work, we characterize the set of linguistic values by a lattice-valued algebraic structure (i.e., LIA) and investigate the corresponding logic systems with linguistic truth-value LIA. Furthermore, we address the satisfiability problems of the logical formula with respect to a certain linguistic truth-value level and the corresponding Quasi-Horn clause logic framework by establishing the resolution principles. A key idea behind these approaches is to directly manipulate the available linguistic information and knowledge, i.e., the symbolic approach acts by direct computation on linguistic truth values.

The paper is organized as follows. In Section 2, we describe and define the linguistic truth-value algebra and recall some basic concepts about LIAs, then define the lattice-valued propositional logic system with truth-valued in a linguistic-valued LIA. In Section 3, the corresponding resolution principle as well as its theorems of soundness and completeness are given on proving the satisfiability of logical formulae with respect to a certain linguistic truth-value level. Section 4 establishes a calculus for linguistic-valued Quasi-Horn clause and claims its soundness and completeness. The conclusion is included in Section 5.

2 Linguistic Truth-Value Logics

2.1 Linguistic-Valued Logical Algebra

2.1.1 Linguistic Assessment Instead of Numerical Assessment

Human beings cannot be seen as a precision mechanism. They usually express their knowledge about the world using linguistic variable in natural language with full of vague and imprecise concepts. The linguistic approach is an approximate technique appropriate for dealing with qualitative aspects of problems. Since words are less precise than numbers, the concept of a linguistic variable serves the purpose of providing a measure for an approximate characterization of the phenomena which are too complex or ill-defined to be amenable to their description by conventional quantitative terms [28].

In order for linguistic variables to be useful tools of analysis, one ought to be able to manipulate them through various operations so that one can directly symbolically manipulate the linguistic variables themselves, i.e., symbolic approach acting by direct reasoning on words, different from approximation approach which uses the associated membership functions, avoid the burden steps implying the investigation of human factor, semantics of the linguistic terms, subjective beliefs etc.

2.1.2 Lattice Structure and Lattice Implication Algebras

A nice feature of linguistic variables is that their values are structured. Symbolic linguistic approach assumes that the linguistic term set is an ordered structure. In general, lattice structures apply whenever ordinal information must be represented. Lattice structures actually provide one possible solution for model the linguistic value structure. The question of the appropriate operation and lattice structure has generated much literature [22-25]. One of most important work is by Goguen [22] who established *L*-fuzzy logic of which truth value set is a complete lattice-ordered monoid, which is also called a complete residuated lattice in Pavelka and Novak´s *L*-fuzzy logic [23-24].

Definition 2.1 [23]**.** A *residuated lattice* (RL) is a structure $<L, \otimes, \rightarrow>$, where

(1) **L**=$<L, \leq, \vee, \wedge, \mathrm{O}, \mathrm{I}>$ is a bounded lattice with the least element O and the greatest element I.

(2) $<\otimes, \rightarrow>$ is an adjoint couple on *L,* i.e.,

(a) $\otimes$ is istone (ordering preserving) on $L\times L$; (b) $\rightarrow$ is antione (order reversing) in the first and isotone in the second variable on $L\times L$; (c) for all $x, y, z\in L$ hold the adjointness condition or Galois correspondence: $x\otimes y\leq c$ iff $x\leq y\rightarrow z$

(3) $<L, \otimes, \mathrm{I}>$ is a commutative monoid.

The operation $\otimes$ is called multiplication and $\rightarrow$ is called residuation.

Since this algebraic structure is quite general, it is relevant to ask whether one can specify the structure. In this note, we specify the algebraic structure to lattice implication algebras introduced by Xu [25-26].

Definition 2.2 (LIA)**.** Let $(L, \vee, \wedge, ')$ be a bounded lattice with an order-reversing involution " $'$ " and the universal bounds O, I, $\rightarrow: L\times L\rightarrow L$ be a mapping. $(L, \vee, \wedge, ', \rightarrow)$ is called a *lattice implication algebra* (LIA) if the following axioms hold for all $x, y, z\in L$:

(A_1) $x\rightarrow(y\rightarrow z)=y\rightarrow(x\rightarrow z)$, (exchange property)
(A_2) $x\rightarrow x=\mathrm{I}$, (identity)
(A_3) $x\rightarrow y=y'\rightarrow x'$, (contraposition or contrapositive symmetry)
(A_4) $x\rightarrow y=y\rightarrow x=\mathrm{I}$ implies $x=y$, (equivalency)
(A_5) $(x\rightarrow y)\rightarrow y=(y\rightarrow x)\rightarrow x$,
(A_6) $(x\vee y)\rightarrow z=(x\rightarrow z)\wedge(y\rightarrow z)$,
(A_7) $(x\wedge y)\rightarrow z=(x\rightarrow z)\vee(y\rightarrow z)$.

Some basic concepts and properties of LIAs can be seen in [25].

2.1.3 Lattice-Valued Linguistic Algebra

A linguistic term differs from a numerical one in that its values are not numbers, but words or sentences in a natural or artificial language. On the other hand, these words, in different natural language, seem difficult to distinguish their boundary sometime, but their meaning of common usage can be understood. Moreover, there are some "vague overlap districts" among some words which cannot be strictly linearly ordered, as given in Fig. 1 of Example 1.1.

Example 1.1. The ordering relationships in some linguistic terms:

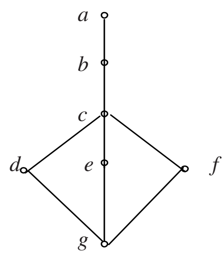


Fig. 1. Linguistic terms in an ordering a=very true, b=more true, c=true, d=approximately true e=possibly true, f=more or less true, g=little true

Note that *d, e*, and *f* are incomparable. One can not collapse that structure into a linearly ordered structure, because then one would impose an ordering on *d*, *e*, and *f* which was originally not present. This means the set of linguistic values may not be strictly linearly ordered. It is shown that linguistic term can be ordered by their meanings in natural language. Naturally, it should be suitable to represent the linguistic values by a partially ordered set or lattice. To attain this goal we characterize the set of linguistic truth-values by a LIA structure, i.e., use the LIA to construct the structure of value sets of linguistic variables.

In general, the value of a linguistic variable can be a linguistic expression involving a set of linguistic terms such as "*high*," "*middle*," and "*low*," modifiers such as "*very*," "*more or less*" (called hedges [27]) and connectives (e.g., "*and*," "*or*"). Let us consider the domain of the linguistic variable "*truth*": domain (*truth*)={*true, false, very true, more or less true, possibly true, very false, possibly false*, ...}, which can be regarded as a partially ordered set whose elements are ordered by their meanings and also regarded as an algebraically generated set from the generators *G*={*true, false*} by means of a set of linguistic modifiers *M*={*very, more or less, possibly*, ...}. The generators *G* can be regarded as the prime term, different prime terms correspond to the different linguistic variables.

Taking into account the above remarks, construction of an appropriate set of linguistic values for an application can be carried out step by step. Consider a set of linguistic hedges, e.g. H^{+}={*very, more or plus*}, H={*approximately, possibly, more or less, little*}, where H^{+} consists of hedges which strengthen the meanings of "true" and the hedges in H weaken it. Put $H=H^{+}\cup H$. H^{+}, H can be ordered by the degree of strengthening or weakening. We say that $a \leq b$ if and only if $a(True) \leq b(True)$ in the natural language, where a and b are linguistic hedges.

Applying the hedges of H to the primary term "*true*" or "*false*" we obtain a partially ordered set or lattice. For example, we can obtain a lattice generated from "*true*" or "*false*" by means of operations in H. We add these three special elements I, M, O called "*absolutely true*," "*medium*," and "*absolutely false*" to the obtaining set so that they have natural ordering relationship with the linguistic truth values. The set of linguistic truth-values obtained by the above procedure is a lattice with the boundary. Moreover, one can define $\wedge$, $\vee$, implication $\rightarrow$ and complement operation $'$ on this lattice according to the LIA structure.

Consider a totally ordered linguistic term set as an example, let n=5, $L_n=\{s_0=Poor, s_1=Low, s_2=Average, s_3=High, s_4=Good\}$. Any label, s_i, represents a possible value for a linguistic variable, and has the following characteristics [28]:

1. The set is ordered: $s_i \leq s_j$ if $i \leq j$.
2. There is the negation operator: $\text{Neg}(s_i) = s_j$ such that $j = n\text{-}1 \text{ -}i$.
3. There is the maximization operator: $\text{Max}(s_i, s_j) = s_i$ if $s_j \leq s_i$.
4. There is the minimization operator: $\text{Min}(s_i, s_j) = s_i$ if $s_i \leq s_j$.

Moreover, implication ($\rightarrow$) based on the L_n structure can be further defined.

For example, let $L_n=\{s_i\,;\ 1 \leq i \leq n\}$ is a totally ordered linguistic term set ($O=s_1<s_2<\ldots<s_n=I$). If the negation operator $'$: $L_n \longrightarrow L_n$ is defined by $(s_i)'=a_{n+1-i}$, and $\longrightarrow: L \times L \longrightarrow L$ is defined by $s_i \rightarrow s_j = s_{n+j-i}$ ($i, j \in \{1,\ldots, n\}$). Then (L_n, $\vee$, $\wedge$, ', $\longrightarrow$) is a residuated lattice, also a LIA, denoted by $\boldsymbol{L_n}$.

2.2 Linguistic-Valued Propositional Logic

In the following sections, L always represents the linguistic truth-value LIA.

Definition 2.3. Let X be the set of propositional variables, $T=L \cup \{', \otimes, \rightarrow\}$ be a type with $ar(') = 1$, $ar(\otimes) = ar(\rightarrow) = 2$ and $ar(a)=0$ for every $a \in L$. The propositional algebra of the linguistic-valued propositional calculus on the set of propositional variables is the free T algebra on X and is denoted by $LP(X)$.

Proposition 2.1. $LP(X)$ is the minimal set Y which satisfies the following conditions: (1) $X \bigcup L \subseteq Y$; (2) If $p, q \in Y$, then $p \otimes q$, $p', p \rightarrow q \in Y$.

Note that L and $LP(X)$ are the algebras with the same type T, where $T = L \bigcup \{', \otimes, \rightarrow\}$.

Definition 2.4 A valuation of $LP(X)$ is a propositional algebra homomorphism γ: $LP(X) \rightarrow L$.

Definition 2.5. Let $p \in LP(X)$, $\alpha \in L$. If $\gamma(p) \geq \alpha$ for every valuation γ of $LP(X)$, we say that p is valid by truth-value level α. If there exists a valuation γ of $LP(X)$ such that $\gamma(p) \geq \alpha$, then p is called α-satisfiable.

Beginning from the normal form is the usual way to discuss the satisfiability of the formula in classical logic. As a first step towards a variant resolution, it is important to deal with implication connectives and consider the generalized normal form.

Definition 2.6. An L-valued propositional logical formula f is called an extremely simple form, in short ESF, if a logical formula f^* obtained by deleting any constant or literal or implication term appearing in f is not equivalent to f.

Definition 2.7. if f is an ESF containing no connectives other than implication connectives, then f is called an indecomposable extremely simple form, in short IESF.

Definition 2.8. All the constants, literals and IESF´s are called generalized literals.

Definition 2.9. An L-valued propositional logical formula G is called a generalized clause (phrase), if G is a formula of the form:

$$G=g_1 \vee \ldots \vee g_i \vee \ldots \vee g_n \quad (G=g_1 \wedge \ldots \wedge g_i \wedge \ldots \wedge g_n)$$

where g_i (i=1,..., n) are generalized literals.

A conjunction (disjunction) of finite generalized clauses (phrases) is called a generalized conjunctive (conjunctive) normal form.

3 α-Satisfiability of Linguistic-Valued Propositional Logic

The following concepts and theorems can be obtained based on the work in [19]. Due to the limited space, the proofs are omitted.

Definition 3.1. (α-Resolution) Let $\alpha \in L$, and G_1 and G_2 be two generalized clauses of the form: $G_1=g_1 \vee \ldots \vee g_i \vee \ldots \vee g_m$ and $G_2=h_1 \vee \ldots \vee h_j \vee \ldots \vee h_n$. If $g_i \wedge h_j \leq \alpha$, then $G=g_1 \vee \ldots \vee g_{i-1} \vee g_{i+1} \vee \ldots \vee g_m \vee h_1 \vee \ldots \vee h_{j-1} \vee h_{j+1} \vee \ldots \vee h_n$ is called an α-resolvent of G_1 and G_2, denoted by $G=R_\alpha(G_1, G_2)$, and g_i and h_j form an α-resolution pair, denoted by (g_i, h_j)-α. It can be regarded as the complemented pair in the sense of α-false.

Definition 3.2. Suppose a generalized conjunctive normal form $S=C_1 \wedge C_2 \wedge \ldots \wedge C_n$, $\alpha \in L$, $\omega=\{D_1, D_2,..., D_m\}$ is called an α-resolution deduction from S to generalized clause D_m, if

(1) $D_i \in \{C_1, C_2, ..., C_n\}$ or (2) there exist j, $k<i$, such that $D_i = R_\alpha(D_j, D_k)$.

If there exists an α-resolution deduction from S to the empty clause $\varnothing$ (denoted by α-false), then ω is called an α-refutation.

Theorem 3.1. (Soundness) Suppose a generalized conjunctive normal form $S=C_1 \wedge C_2 \wedge \ldots \wedge C_n$, $\alpha \in L$, $\{D_1, D_2,..., D_m\}$ is an α-resolution deduction from S to a generalized clause D_m. If D_m=α-false, then $S \leq \alpha$, that is, if $D_m \leq \alpha$, then $S \leq \alpha$

Theorem 3.2. (**Completeness**) Let S be a regular generalized conjunctive normal form, $\alpha \in L$, $\alpha <$I, α be a dual numerator in L. And suppose that there exists $\beta \in L$ such that $\beta \wedge (\beta \rightarrow \beta') > \alpha$. If $S \leq \alpha$, then there exists an α-resolution deduction from S to α-false.

4 Linguistic-Valued Quasi-horn Clause Logic

In [20], [25], a lattice-valued first-order logic LF(X) based on LIA is established. In this section, we consider the restrict set $\Im$ of LF(X), called lattice-valued Quasi-Horn clause class as an extension of classical Horn clause. Some concepts about symbols, terms, well-formed formulas, and interpretation can be referred to [20], [25].

Definition 4.1. A linguistic-valued Quasi-Horn clause (in short, *L*-type Q-Horn clause) is a well-formed formula without free variables as the following form:

$$(\forall x_1)\cdots(\forall x_k)(p \to q) \tag{1}$$

or

$$(\forall x_1)\cdots(\forall x_k)(q) \tag{2}$$

where q is an atom formula with the variables $x_1, \cdots, x_k$ only, p is a formula without restriction of qualifiers and including only connective $\vee$ or $\wedge$.

In classical logic, a clause with at most one positive literal is called a Horn clause. Prolog programming problem in knowledge engineering, as the direct application of Horn, generally consists of three clauses as follows:

(1) Facts (or asserts) expressing the related objects and the relations among these objects, *P*.

(2) Rules defining the relationship among some objects. P : - P_1, P_2, …, P_n.

(3) Problems (or objectives). ? - Q_1, …, Q_m.

Here P in (1) is obviously a Horn clause. (2) can be represented as

$$(P_1 \wedge P_2 \wedge \cdots \wedge P_n) \to P \, .$$

That is, $\neg P_1 \vee \cdots \vee \neg P_n \vee P$, also a Horn clause. (3) represents if $Q_1 \wedge \cdots \wedge Q_m$ can be inferred from (1) and (2). From the resolution point of view, the negation of conclusion of objective, i.e., $\neg Q_1 \vee \cdots \vee \neg Q_m$, is again a Horn clause. Prolog programming is based on this kind of Horn clause logic.

The *L*-type Q-Horn clause in Definition 3.1 is an extension of classical Horn clause in order to establish an extended linguistic-valued or called a *L*-type Prolog programming. *L*-type Q-Horn clause looks similar as that in classical logic, however, here "→" has been different and been generalized, i.e., not any more a Kleene implication, but the more general implication form in LIA. At the same time, the rules and facts in classical Prolog can be interpreted as the set of axioms which are always true. The *L*-type Prolog which will be established in the following is an extension of this framework, i.e., the rules and facts are extended into L-type fuzzy rules and fuzzy facts, i.e., truth-value is extended from {0, 1} into lattice-ordered linguistic truth-values, and the reasoning process is associated with the change of truth-values.

Suppose that F is the set of all *L*-type Q-Horn clauses in $\Im$, and $\boldsymbol{F}_L(F)$ represents the set of all the *L*-type fuzzy set on F. Let $A \in \boldsymbol{F}_L(F)$, A is called a non-logical fuzzy axiom set. For any logical formula $\varphi \in F$, always associated with a value $A(\varphi) \in L$. In the real-world practices, one may suppose that $A(\varphi)$ is the minimal truth-value degree of a proposition φ or possibility degree, or credibility degree (based on the application context). It is expected that during the reasoning process, every inferred formula $\psi\in LP(X)$, whose associated (truth) value should be larger than $A(\psi)$. Thus, we need to know the minimal value $A(\psi)$ of the involved formula ψ, in addition, the associated truth values may continuously improved during the deduction process.

In the following, we always assume that for any L-type Q-Horn clause φ, $A(\varphi) > 0$ or A is said to be regular, and φ is called a non-logical axiom of A.

Definition 4.2. Let D be an interpretation of the language $\Im$, $A\in \boldsymbol{F}_L(\mathrm{F})$. D is called a model of A or D satisfies A, if for arbitrary $\varphi\in F$, $A(\varphi)\le\gamma(\varphi)_D$ holds, where $\gamma(\varphi)_D$ is the truth-value of φ under the interpretation D.

Definition 4.3. Let $A\in \boldsymbol{F}_L(F)$, $\varphi\in F$, $\alpha\in L$. φ is said to be α-true in A, denoted as $A\|_\alpha = \varphi$ if $|\alpha = \wedge\{\gamma(\varphi)_D;\ D \text{ is an interpretation satisfying } A\}$.

Set $Con(A)(\varphi) = \alpha = \wedge\{\gamma(\varphi)_D;\ D \text{ is an interpretation satisfying } A\}$

Now we consider the syntax in the following part. For arbitrary p, $q\in F$, set $p\otimes q=:(p\to q')'$.

Theorem 4.1. For any $p, q, r, \alpha\in L$, $m, n\in N$, the following statements hold:

(1) $\|\!\!= p \to \mathrm{I}$; (2) $\| = p \to p$; (3) $\| = p\wedge q \to p$; (4) $\| = p\wedge q \to q$

(5) $\| = (p')' \to p$; (6) $\| = \forall xp \to p$

Definition 4.4. Axioms in lattice-valued first-order Q-Horn clause logic system are L-type fuzzy set A_L ($A_L\in \boldsymbol{F}_L(F)$) in the following forms:

$$A_L(\varphi) = \begin{cases} \mathrm{I}, & \varphi \text{ is a formula in the form of } (1) \sim (6) \text{ in Theorem 4.1;} \\ \alpha, & \varphi = \alpha, \alpha\in L; \\ \mathrm{O}, & \text{Otherwise.} \end{cases}$$

Definition 4.5. Let $A\in \boldsymbol{F}_L(\mathrm{F})$, $\varphi\in F$. A formal proof ω of φ from A is a finite sequence in the following form: $(\varphi_1,\alpha_1),\cdots,(\varphi_n,\alpha_n)$, where $\varphi_n = \varphi$. For any i, $1\le i\le n$, $(\varphi_i,\alpha_i)\in F\times L$, and

(i) $A_L(\varphi_i) = \alpha_i$ or (ii) $A(\varphi_i) = \alpha_i$ or (iii) There exists $j, k<i$, $\varphi_j = (\forall x_1)\cdots(\forall x_m)(p)$, $\varphi_k = (\forall x_1)\cdots(\forall x_m)(p\to q)$, $\varphi_i = (\forall x_1)\cdots(\forall x_m)(q)$,
and $\alpha_i = [\alpha_j \otimes A[(\forall x_1)\cdots(\forall x_m)(p\to q)]]\vee A[(\forall x_1)\cdots(\forall x_m)(q)]$.

(iv) There exist $k<i$, such that $\varphi_k = (\forall x_1)\cdots(\forall x_k)(q)$ and there exists a term $t_{i_1},\cdots,t_{i_r}$ ($i_1,\cdots,i_r\in\{1,\ldots,k\}$) without free variables, where φ_k is obtained from φ_i by replacing the variables x_{i_j} ($j=1,\cdots,r$) in q as t_{i_j}, the qualifiers only bound to the remaining variables. In addition, $\alpha_i = A[(\forall x_1)\cdots(\forall x_k)(q)]\vee A(\varphi_i)$.

Here n is called the length of the proof ω, denoted as $l(\omega)$; α_n is called the value of the proof ω, denoted as $val(\omega)$.

Definition 4.6. Let $A \in \boldsymbol{F}_L(F)$, $\varphi \in F$, $\alpha \in L$. φ is called an α-theorem of A, denoted as $A \Vdash\!\!-_{\alpha} \varphi$, if $\alpha = \vee\{val(\omega);\ \omega \text{ is a proof of } \varphi \text{ from A}\}$.

Denote $Ded(A): F \to L$ as $Ded(A)(\varphi) = \vee\{val(\omega);\ \omega \text{ is a proof of } \varphi \text{ from A}\}$.

More generally, since it is reasonable that the uncertain premises will infer the uncertain conclusion, so starting from L-type fuzzy premise set $A \in \boldsymbol{F}_L(F)$, the inferred output would be also L-type fuzzy set A, $B \in F_L(F)$. We can generalize Definition 3.6 into the following form:

Definition 4.7. Let A, $B \in \boldsymbol{F}_L(F)$ be regular. B is said to be syntactically inferred from A, if there exists a finite sequence of the form $A_0, \ldots, A_n \in \boldsymbol{F}_L(F)$, where $B \le A_n$, and for each $k \in \{0, \ldots, n\text{-}1\}$, A_{k+1} can be obtained from A_k as the following steps:

(1) There exists a Q-Horn clause $(\forall x_1)\cdots(\forall x_m)(p \to q)$ in F such that

(i) if $q_0 \in F$, where the free variables are $x_1, \ldots, x_r$ and $q_0 \neq q$, then

$$A[\,(\forall x_1)\cdots(\forall x_m)(q_0)\,] = A_{k+1}[\,(\forall x_1)\cdots(\forall x_r)(q_0)\,];$$

or (ii) $A_{k+1}[(\forall x_1)\cdots(\forall x_m)(q)]$

$$= [A_k[(\forall x_1)\cdots(\forall x_m)(p)] \otimes A[(\forall x_1)\cdots(\forall x_m)(p \to q)]] \vee A[(\forall x_1)\cdots(\forall x_m)(q)].$$

(2) There exists a Q-Horn clause $\varphi = (\forall x_1) \cdots (\forall x_k)\ (q)$ and a term $t_{i_1}, \cdots, t_{i_r}$ ($i_1, \cdots, i_r \in \{1, \ldots, k\}$) without free variables, such that φ is transformed into φ^* by replacing the variables x_{i_j} $(j = 1, \cdots, r)$ in q as t_{i_j}, where the remaining variables are bounded by the qualifiers, and $A_{k+1}(\varphi^*) = A[(\forall x_1)\cdots(\forall x_k)(q)] \vee A(\varphi^*)$.

The mapping $Ded(A): F \to L$ is given as follows:

$$Ded(A)(\varphi) = \vee\{B(\varphi);\ B \text{ is syntactically inferred from } A\},$$

where φ is the Q-Horn clause in the form of (2) in Definition 4.1.

For an universal qualifier, if $q \in F$, x is an individual variable, then $\gamma(\forall xq)_D = \bigwedge_{d \in D} \gamma(q)_{D(x/d)}$, where $D(x/d)$ is an interpretation from the interpretation D by replaced x as d.

Lemma 4.1. Let p, $q \in F$, x does not appear in p as a free variable. Then

$$\forall x(p \to q) = p \to \forall xq.$$

Corollary 4.1. Let p, $q \in F$, both x and y does not appear in p as a free variable. Then $\forall x \forall y(p \to q) = p \to \forall x \forall yq$.

Corollary 4.2. Let p, $q \in F$, $x_1, \ldots, x_k$ do not appear in p as the free variables. Then $(\forall x_1)\ldots(\forall x_k)(p \to q) = p \to (\forall x_1)\ldots(\forall x_k)(q)$.

Lemma 4.2. Let p, $q \in F$, x does not appear in p as a free variable. Then

$$\gamma(\forall xq)_D \ge \gamma(\forall xp)_D \otimes \gamma(\forall x(p \to q))_D$$

holds for any interpretation D.

Corollary 4.3. Let p, $q\in F$, x_1, …, x_k do not appear in p as the free variables. Then $\chi((\forall x_1)\dots(\forall x_k)(q))_D \geq \chi((\forall x_1)\dots(\forall x_k)(p))_D \otimes \chi((\forall x_1)\dots(\forall x_k)(p\to q))_D$
holds for any interpretation D.

Theorem 4.2. (Soundness) Let $A\in \boldsymbol{F}_L(F)$ be regular. Then

$$Ded(A)(\varphi) \leq Con(A)(\varphi)$$

holds for any lattice-valued Q-Horn clause φ in the form of (2) in Definition 4.1.

Theorem 4.3. (**Completeness**) Let $A\in \boldsymbol{F}_L(F)$ be regular. Then

$$Ded(A)(\varphi) \geq Con(A)(\varphi)$$

holds for any lattice-valued Q-Horn clause φ in the form of (2) in Definition 4.1.

It follows from Theorems 4.2 and 4.3 that if $A\in \boldsymbol{F}_L(F)$ is regular, then $Con(A)=Ded(A)$. That means the established linguistic-valued Q-Horn clause logic is sound and complete, which will provide a support and theoretical foundation for further establishing lattice-ordered linguistic-valued Prolog language.

4 Conclusions

Based on the key idea of the symbolic approach acts by direct reasoning on linguistic truth values, we characterized the set of linguistic values by a lattice-valued algebraic structure (lattice implication algebra) and investigated the corresponding logic systems with linguistic truth values, and furthermore, investigated the automated reasoning scheme based on linguistic truth-valued logic system, while the resolution method as well as its theorems of soundness and completeness were given on proving the satisfiability of logical formulae with respect to a certain linguistic truth-value level, and a linguistic-valued quasi-Horn clause logic were investigated with its soundness and completeness theorem being given.

Acknowledgements

The work was supported by the Research Project TIN2006-02121, and partially supported by the National Natural Science Foundation of China (Grant No. 60474022).

References

1. J.P. Robinson, A machine-oriented logic based on the resolution principle, *J. of A.C.M.*, 12: 23-41, 1965.
2. R.C.T. Lee, Fuzzy logic and the resolution principle, *Journal of A.C.M.*, 19: 109-119, 1972.
3. X.H. Liu and H. Xiao, Operator fuzzy logic and fuzzy resolution, *Proc. of the 5th IEEE Inter. Symp. on Multiple-Valued Logic* (*ISMVL´85*), Kingstion, Canada, pages: 68-75, 1985.
4. M. Mukaidono, Fuzzy inference of resolution style, in: *Fuzzy Sets and Possibility Theory*, R.R. Yager (Ed.), Pergamon Press, New York, pages: 224-231, 1982.

5. D. Dubois and H. Prade, Resolution principle in possibilistic logic, *Int. J. of Approximate Reasoning* 4 (1): 1-21, 1990.
6. T.J. Weigert, J.P. Tsai and X.H. Liu, Fuzzy operator logic and fuzzy resolution, *Journal of Automated Reasoning*, 10(1): 59 – 78, 1993.
7. C.S. Kim, D.S. Kim, and J.S. Park, A new fuzzy resolution principle based on the antonym, *Fuzzy Sets and Systems*, 113 (2): 299-307, 2000.
8. C.G. Morgan, Resolution for many-valued logics, *Logique et Analyses*, 19 (74-76): 311-339, 1976.
9. E. Orlowska, Mechanical proof methods for Post logics, *Logique et Analyse*, 28(110): 173-192, 1985.
10. P.H. Schmitt, Computational aspects of three-valued logic, in*: Proc. of the 8th Inter. Conf. on Automated Deduction*, J.H. Siekmann (Ed.), Springer, LNCS, pages: 190-198, 1985.
11. R. Hahnle, *Automated Deduction in Multiple-Valued L*ogics, Oxford University Press, 1993.
12. H.A. Blair and V.S. Subrahamnian, Paraconsistent logic programming, *Theoretical Computer Science*, 68:135-154, 1989.
13. M. Kifer and V.S. Subrahamnian, Theory of generalized annotated logic programming and its application, *Journal of Logic Programming* 12: 335-367, (1992).
14. J.J. Lu and L.J. Henschen, The completeness of gp-resolution for annotated logic, *Inform. Process. Lett.* 44: 135-140, 1992.
15. M. Baaz, C.G. Fermuller, Resolution for many valued logics. In *Proc. of Logic Programming and Automated Reasoning* (LPAR'92), Voronkv, A. (Ed.). Springer, LNAI 624, page 107-118, 1992.
16. Z. Stachniak, *Resolution Proof Systems: An Algebraic Theory*, Kluwer Academic Publisher, 1996.
17. S. Lehmke. A resolution-based axiomatisation of `bold' propositional fuzzy logic. In: *Fuzzy Sets, Logics, and Reasoning about Knowledge* (Eds. by D. Dubois, E. P. Klement, and H. Prade), Kluwer Academic Publishers, Applied Logic, 1999.
18. V. Sofronie-Stokkermans, Chaining techniques for automated theorem proving in finitely-valued logics. In: *Proc. of the 2000 ISMVL.* Portland, Oregon, pages: 337-344, 2000.
19. Y. Xu, D. Ruan, E.E. Kerre., J. Liu, α-resolution principle based on lattice-valued propositional logic LP(X). *Information Sciences*, 130: 195-223, 2000.
20. Y. Xu, D. Ruan, E.E. Kerre., J. Liu, α-Resolution principle based on first-order lattice-valued logic LF(X), *Information Sciences*, 132: 221-239, 2001
21. J. Liu, D. Ruan, Y. Xu, Z.M. Song, A resolution-like strategy based on a lattice-valued logic, *IEEE Transactions on Fuzzy Systems*, 11 (4): 560-567, 2003.
22. J.A. Goguen, The logic of inexact concepts, *Synthese*, 19: 325-373, 1969.
23. J. Pavelka, On fuzzy logic I: Many-valued rules of inference, II: Enriched residuated lattices and semantics of propositional calculi, III: Semantical completeness of some many-valued propositional calculi, *Zeitschr. F. Math. Logik und Grundlagend. Math.*, 25: 45-52, 119-134, 447-464, 1979.
24. Novak, First-order fuzzy logic, *Studia Logica* 46 (1) (1982) 87-109.
25. Y. Xu, D. Ruan, K.Y. Qin, J. Liu, *Lattice-Valued Logic: An Alternative Approach to Treat Fuzziness and Incomparability*, Springer-Verlag, Heidelberg, July, 2003, 390 pages.
26. Y. Xu, Lattice implication algebras, *J. of Southwest Jiaotong Univ.* (in Chinese), 89(1)(1993) 20-27.
27. N.C. Ho, W. Wechler, Hedge algebras: an algebraic approach to structure of sets of linguistic truth values. *Fuzzy Sets and Systems* 35: 281–293, 1990.
28. F. Herrera, L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Systems*, 8 (6): 746–752 2000.

Part XIV

Fuzzy Logic Applications

Fuzzy Flip-Flops Revisited

László T. Kóczy[1] and Rita Lovassy[2]

[1] Inst. of Information Technology, Mechanical and Electrical Eng., Széchenyi István University, Győr, Hungary. Dept. of Telecommunication and Media Informatics, Budapest University of Technology and Economics, Hungary
koczy@sze.hu, koczy@tmit.bme.hu
[2] Inst. of Microelectronics and Technology, Budapest Tech, Hungary
lovassy.rita@kvk.bmf.hu

Abstract. J-K flip-flops are elementary digital units providing sequential features/memory functions. Their definitive equation is used both in the minimal disjunctive and conjunctive forms. Fuzzy connectives do not satisfy all Boolean axioms, thus the fuzzy equivalents of these equations result in two non-equivalent definitions, "reset and set type" fuzzy flip-flops (F^3) by Hirota *& al.* when introducing the concept of F^3. There are many alternatives for "fuzzifying" digital flip-flops, using standard, algebraic or other connectives. The paper gives an overview of some of the most famous F^3-s by presenting their definitions and presenting graphs of the inner state for a typical state value situation. Then a pair of non-associative operators is introduced, and the properties of the respective F^3 are discussed. The investigation of possible fuzzy flip-flops is continued by examining Türkşen's IVFS, its midpoint values, and by introducing "minimized IVFS" (MIVFS), along with the MIVFS midpoints.

Keywords: Fuzzy logic, J-K flip-flop, logical normal forms, IVFS.

1 Introduction

As a possible starting point for defining fuzzy logical sequential circuits and for effective processing of fuzzy information, [11] introduced the notion of the fuzzy J-K flip-flop. They introduced two types of fundamental characteristic equations of fuzzy flip-flops (F^3), "reset type and set type" equations, both being fuzzy extensions of the binary J-K flip-flop. While they proposed standard operations, the algebraic fuzzy flip-flop was introduced later by [10], as an example of the general F^3 concept also defined in the same paper.

While briefly reviewing these definitions, several new examples will be given, using further norms. Next, the concept will be extended for non-associative fuzzy operations (cf. [2]). We prove the surprising result, that comparing the equations of the next state of the set and reset type [4] version of a modified version of the Fodor fuzzy flip-flop (F^4) [2, 8], i.e. there is only one single F^4 and the two formulas are equivalent. Equivalence does not hold for any other F^3 defined as far in the literature.

Eventually interval valued fuzzy flip-flops will be overviewed both in sense of Türkşen's *IVFS* and according to a new concept of minterm-maxterm *IVFS* (*MIVFS*).

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 643–652, 2007.
springerlink.com

2 Binary (Boolean) and Fuzzy Flip-Flops

2.1 Binary Logic J-K Flip-Flops

All types of traditional binary flip-flop circuits, such as the most general J-K flip-flop store a single bit of information. These circuits are also the elementary components of synchronous sequential digital circuits. The next state $Q(t+1)$ of a J-K flip-flop is characterized as a function of both the present state $Q(t)$ and the two present inputs $J(t)$ and $K(t)$. In the next, (t) will be omitted. The minterm expression (Disjunctive Normal Form, DNF) of $Q(t+1)$ can be written as

$$Q(t+1) = \overline{J}\,\overline{K}Q + J\overline{K}\,\overline{Q} + J\overline{K}Q + JK\overline{Q} \ , \tag{1}$$

and this can be simplified to the minimal disjunctive form

$$Q(t+1) = J\overline{Q} + \overline{K}Q \ . \tag{2}$$

This latter is well-known as the characteristic equation of J-K flip-flops. The equivalent maxterm expression (Conjunctive Normal Form, CNF) is given by

$$Q(t+1) = (J+K+Q)(J+\overline{K}+Q)\left(J+\overline{K}+\overline{Q}\right)\left(\overline{J}+\overline{K}+\overline{Q}\right) \ , \tag{3}$$

in a similar way the minimized conjunctive form is

$$Q(t+1) = (J+Q)(\overline{K}+\overline{Q}) \ . \tag{4}$$

All four expressions, (1), (2), (3) and (4), are equivalent in Boolean logic, but there are no such fuzzy operation triplets where these forms are necessarily equivalent. Which of these four, or any other, should be considered as the most proper fuzzy extension of the definitive equation of the very fundamental concept of fuzzy J-K flip-flop? There is no justifiable argumentation that prefers any of these four to the other and there is no unambiguous way to introduce the concept of fuzzy J-K flip-flop. While normal forms are especially important for theoretical reasons both minimal forms are equally important in the practice. Thus [4, 5] proposed two dual definitions of fuzzy flip-flops (F^3). (2) was interpreted as the definition of “reset type F^3”:

$$Q_R(t+1) = (J \wedge \neg Q) \vee (\neg K \wedge Q) \ , \tag{5}$$

where the logic operation symbols stand for standard fuzzy conjunction, disjunction and negation. In a similar way the definition of “set type F^3” was obtained by reinterpreting (4) with fuzzy operations:

$$Q_S(t+1) = (J \vee Q) \wedge (\neg K \vee \neg Q) \ . \tag{6}$$

Of course, it is possible to substitute standard operations by any other fuzzy operation triplet (preferably De Morgan triplet), this way obtaining a multitude of various fuzzy flip-flop (F^3) pairs, such as the algebraic F^3 introduced in [10,11].

3 Behavior of F^3s Based on Various Fuzzy Operations

Many different norms play important roles in applications or have nice mathematical properties. Only few have been investigated from the point of view of the F^3-s generated by deploying them in (5) and (6): the standard and algebraic norms and a pair of non-associative operations. (For the three see [4, 5], [10, 11] and [2], respectively.) Reset and set type F^3-s sometimes do have rather different behavior even though they are both extensions of the same binary circuit. They are identical at the borders (*J, K* = 0, 1), thus the pairs are generalizations, but they are disturbingly non-dual. The negation used in the whole paper will be the standard negation.

3.1 Standard Fuzzy Flip-Flops

Equations (5) (reset type F^3) and (6) (set type F^3) can be expressed respectively as

$$Q_R(t+1) = \max\left[\min(J,1-Q), \min(1-K,Q)\right] \tag{7}$$

and

$$Q_S(t+1) = \min\left[\max(J,Q), \max(1-K,1-Q)\right] \ . \tag{8}$$

Standard F^3-s are characterized by piecewise linear functions with breakpoints in the projections, consequently break lines in three dimensions. Calculation with standard F^3-s is fast and easy, but characteristic functions are not smooth (see Fig. 1).

3.2 Algebraic Fuzzy Flip-Flops

If deploying standard complementation, algebraic product and algebraic sum for negation, t-norm and t-conorm, respectively, equations (5) and (6) can be rewritten as

$$Q_R(t+1) = J + Q - 2JQ - KQ + JQ^2 + JQK - JQK^2 \ , \tag{9}$$

$$Q_S(t+1) = J + Q - JQ - JKQ - KQ^2 + JKQ^2 \ . \tag{10}$$

These equations show the results of transformation into a simplified form by using the definitions of algebraic product and sum. Combining equations (9) and (10) we obtain the unified equation of reset and set type:

$$Q(t+1) = J + Q - JQ - KQ \ . \tag{11}$$

This equation is considered the *fundamental equation of the algebraic* F^3. It is remarkable how simple this combined equation is. In addition to its simplicity it represents a symmetrical, dual solution.

Evaluating the curves in Fig. 1, they clearly demonstrate the relation between $Q_R(t+1)$ and $Q_S(t+1)$:

$$Q_S(t+1) - Q_R(t+1) = \left[J(1-K) + K(1-J)\right]Q(1-Q) \geq 0 \tag{12}$$

$$Q_R(t+1) \le Q_S(t+1). \tag{13}$$

Reset type curves always go below set ones. Algebraic operations produce smooth (differentiable) curves and surfaces with no breakpoints or lines at all. This is true also for the combined F^3 which is not shown here.

3.3 Drastic Fuzzy Flip-Flops

Fuzzy unions and intersections satisfying the axiomatic skeleton of boundary conditions, commutativity, monotony, and associativity are bounded by the inequalities (cf. [6])

$$\max(a,b) \le u(a,b) \le u_{\max}(a,b), \tag{14}$$

$$u_{\max}(a,b) = \begin{cases} a & \text{when } b = 0, \\ b & \text{when } a = 0, \\ 1 & \text{otherwise.} \end{cases} \tag{15}$$

$$i_{\min}(a,b) \le i(a,b) \le \min(a,b), \tag{16}$$

$$i_{\max}(a,b) = \begin{cases} a & \text{when } b = 1, \\ b & \text{when } a = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Operations $u_{\max}(a,b)$ and $i_{\min}(a,b)$ are called “drastic sum” and “drastic product”. Together they represent a pair of fuzzy union and fuzzy intersection which are extremal in the sense that for all $a,b \in [0,1]$, inequality

$$u_{\max}(a,b) - i_{\min}(a,b) \ge u(a,b) - i(a,b) \tag{18}$$

is satisfied for an arbitrary pair of fuzzy union u and intersection i.

Fig. 1 illustrates also the behavior of these two extremal F^3-s. The characteristic lines are piecewise linear again, often following the 0 or 1 line. Obviously these lines (and the surfaces they represent) realize extremal cases even for the possible F^3-s.

3.4 Łukasiewicz Fuzzy Flip-Flops

The Łukasiewicz norm and co-norm are defined as follows:

$$i_L(a,b) = \max[a+b-1, 0], \tag{19}$$

$$u_L(a,b) = \min[a+b, 1]. \tag{20}$$

Based on them, the two Łukasiewicz F^3-s are given by

$$Q_R(t+1) = \min\left[\max(J-Q,0) + \max(Q-K,0),1\right] , \tag{21}$$

$$Q_S\left(t+1\right) = \max\left[\min(J+Q,1) + \min(2-K-Q,1) - 1,0\right] . \tag{22}$$

3.5 Non-associative Fuzzy Flip-Flops

In [2] a pair of non-associative (non-dual) operations for a new class of fuzzy flip-flops was proposed. It was stated there that any F^3 satisfying:

P1: F (0, 0, Q) = Q,
P2: F (0, 1, Q) = 0,
P3: F (1, 0, Q) = 1,
P4: F (1, 1, Q) = n (Q),
P5: F (*e*, *e*, Q) =*e*,
P6: F (D, n (D), Q) = D for D$\in\left(0,1\right)$,

where *e* = n (*e*) is the equilibrium belonging to n and n is a strong negation. P2, P3 and P6 can be merged into the single property P6':

P6': F (D, n (D), Q) =D, D$\in\left[0,1\right]$.

[2] states a Theorem according to which $Q_R\left(t+1\right)$ fulfils P6' if and only if there exists an automorphism ϕ of the unit interval such that

$$Q_R\left(t+1\right) = \phi^{-1}\left[\phi(J)(1-\phi(Q)) + \phi(Q)(1-\phi(K))\right] . \tag{23}$$

Similarly, $Q_S\left(t+1\right)$ fulfils the same if and only if there exists an automorphism ψ satisfying

$$Q_S\left(t+1\right) = \psi^{-1}\left[\psi(J)(1-\psi(Q)) + \psi(Q)\left(1-\psi\left(K\right)\right)\right] . \tag{24}$$

The following equation satisfies all P_i-s (i = 1, 4, 5, 6')

$$\begin{aligned} Q_R\left(t+1\right) &= \min\left[T\left(J,1-Q\right) + T\left(1-K,Q\right),1\right] = \\ &= T(J,1-Q) + T(1-K,Q) . \end{aligned} \tag{25}$$

The formula for the set type F^4 in [2] is however not the proper dual pair of (25), moreover, it is problematic in the sense of closeness for the unit interval. Thus in [8] we proposed a corrected definition for the set type formula as follows:

$$\begin{aligned} Q_S\left(t+1\right) &= \max\left[S\left(J,Q\right) + S\left(1-K,1-Q\right) - 1,0\right] = \\ &= S(J,Q) + S(1-K,1-Q) - 1 . \end{aligned} \tag{26}$$

This correction does not change the validity of the above referred Theorem concerning ψ. In (25) and (26) *T* and *S* denote the so called Łukasiewicz norms. From here:

$$Q_R(t+1)=\frac{\min\left(J,1-Q\right)+\max\left(J-Q,0\right)}{2}+\frac{\min\left(Q,1-K\right)+\max\left(Q-K,0\right)}{2}, \tag{27}$$

$$Q_S\left(t+1\right)=\frac{\max(J,Q)+\max\left(1-K,1-Q\right)-1}{2}+\frac{\min\left(J+Q,1\right)+\min\left(2-Q-K,1\right)-1}{2}. \tag{28}$$

These equations were obtained by combining the standard and the Łukasiewicz norms by the arithmetic mean in the inner part of the formula. The other parts use Łukasiewicz operations. Let us briefly denote this corrected pair of [2] type fuzzy flip-flops (referring to the name of the first author J. Fodor) by F^4.

Comparing the modified form of the set type F^4 we come to the surprising result that there is only one F^4 in this particular case, as the two formulas are equivalent.

Statement 3.1

$Q_R^{F^4}=Q_S^{F^4}$, thus there is *only one* (symmetrical, corrected) F^4.
In [9] we proved that

$$Q_R\left(t+1\right)=Q_S\left(t+1\right) \tag{29}$$

We found no straightforward technique for the proof other than analyzing the equations theoretically for every possible combination of J,K,Q. For the 6 values *J, K, Q,* $\neg J$, $\neg K$, $\neg Q$, the total number of all possible sequence combinations is $3!*2^3$ = 48 as any variable and its negated are symmetrical to the equilibrium e = 0.5. These 48 cases are, however, not all essentially different. We succeeded to identify 13 essentially different cases and we could prove the equality for each of them separately. The deduction of these proofs is omitted here for the sake of saving space.

Thus the new F^4 proposed in [9] is indeed a single F^3 with nice dual and symmetrical behavior (see also Fig. 1).

4 The DNF-CNF Interval

The Disjunctive and Conjunctive Normal Forms (DNF and CNF) play very special roles in classical Boolean logic. They represent those standard forms, which do not contain any redundancy in the sense that they cannot be further reduced by applying the idempotence law; however they consist of complete members containing all variables in question. Thus usually they can be simplified by merging and eliminating, this way obtaining the corresponding minimal forms (DMF and CMF). In fuzzy logic there are no standard forms in this sense as idempotence itself does usually not hold. Any repeated member would usually change the value of the expression.

Despite this latter fact several authors consider the fuzzy extensions of DNF and CNF as having special significance. Especially, Türkşen [12] examined the special properties of the extended normal forms and came to the very convenient conclusion that in fuzzy logic the equivalents of the CNF and DNF forms represent the extremes of all possible expressions that correspond to forms being equivalent in binary logic. Thus he showed that for any fuzzy connective the value of every other form lies in the interval formed by the CNF and DNF values. In [13], this result was proven for logical operations with an arbitrary number of variables.

4.1 Interval Valued Fuzzy Sets

While the theory of the interval valued fuzzy sets (*IVFS*) is much more general and can be considered as a special case of L-fuzzy sets [3], Türkşen proposed the interval determined by the Disjunctive and Conjunctive Normal Forms as the interval associated with the value belonging to an expression obtained by the fuzzy extension of some classic binary concept. Indeed, any theoretically possible formulation of the same concept would result in a value lying within the interval thus proposed.

According to Türkşen's statement DNF is always included in the corresponding CNF, i.e., $DNF(\cdot) \subseteq CNF(\cdot)$ where $(\cdot)$ represents any particular expression [12]. The fundamental result that every DNF is contained in its corresponding CNF is true for min-max operators and for algebraic triplets as well. Türkşen [12] proposed the definition of the interval-valued fuzzy set representing a Boolean expression as follows:

$$IVFS(\cdot) = \left[DNF(\cdot), CNF(\cdot) \right] . \tag{30}$$

4.2 Fuzzy Flip-Flops Based on Türkşen's *IVFS* and *MIVFS*

The DNF and CNF forms for J-K flip-flops are expressed in 2.1. By applying the usual denotations for fuzzy negation, t-norm and t-conorm, (1) is re-written as follows

$$Q_{DNF}(t+1) = \left((1-J) \wedge (1-K) \wedge Q\right) \vee \left(J \wedge (1-K) \wedge (1-Q)\right) \vee \left(J \wedge (1-K) \wedge Q\right) \vee \left(J \wedge K \wedge (1-Q)\right), \tag{31}$$

further, in the same way, (3) becomes

$$Q_{CNF}(t+1) = (J \vee K \vee Q) \wedge \left(J \vee (1-K) \vee Q\right) \wedge \left(J \vee (1-K) \vee (1-Q)\right) \wedge \left((1-J) \vee (1-K) \vee (1-Q)\right). \tag{32}$$

This way we obtained two more definitions of min-max fuzzy flip-flops. It is questionable, of course, whether these new definitions play any more important role in the practice than the previous set type and reset type equations. These latter ones may be called "normalized reset type and set type" F^3-s.

Using the algebraic operations, a similar pair of normalized flip-flops may be obtained. Here we omitted the explicit formulas as it is rather easy to determine them but they are somewhat lengthy.

Although the examination of these new F^3-s may be interesting itself, in this paper we go one step further. Reset and set type behaviors are generally different and none of them possess the ″nice″ symmetrical behavior of the original J-K flip-flop. This is why a symmetrical F^3 was proposed earlier by combining the two minimal forms in the equilibrium point [10]. Only one exceptional combination of operations has been found as far where the two types completely coincided (see Section 3.5).

If we intend to combine two different extensions of the original definition we might also choose a single representative point of each interval corresponding to the *IVFS* obtained from the two normal forms. As the most obvious representative, the midpoint is proposed here. Fig. 1 depicts the graphs belonging to DNF and CNF flip-flops. The graph presents the behavior of the F^3 for the general case *J=0.45,* and *K=0.80* (in steps of 0.25 from 0 to 1).

In all earlier publications on F^3-s, however, *minterm* and *maxterm* expressions played important roles rather then the normal forms. Thus, we propose here a new idea, the interval limited by the *minterm* and the *maxterm* expressions:

$$MIVFS(\cdot) = \left[DMF(\cdot), CMF(\cdot)\right] \tag{33}$$

MIVFS gives a narrower interval according to the inequality $DNF(\cdot) \subseteq DMF(\cdot) \subseteq CMF(\cdot) \subseteq CNF(\cdot)$. Properties of *MIVFS* based F^3-s will be examined later.

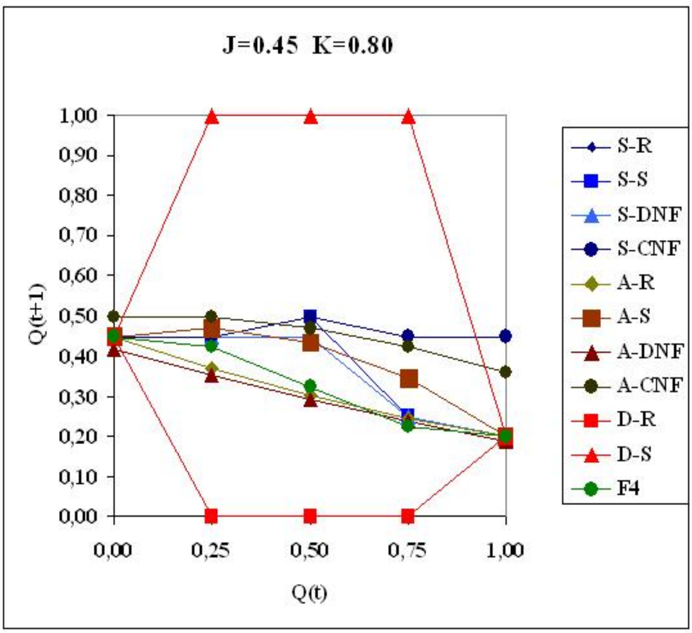


Fig. 1. The behavior of various F^3-s for a general JK value pair

5 Conclusions

After an overview of previous definitions of F^3-s, extensions for further connectives were presented. Then non-associative F^3-s were investigated and a surprising result was presented. Next Türkşen's *IVFS*-s were introduced as derived from DNF and CNF in Boolean logic. The extremal F^3-s generated by *IVFS* was discussed. Considering that the flip-flops thus defined were described by an "opening" pair of graphs, the midpoint of the *IVFS* was suggested as the definitive point for a new symmetrical F^3. Similarly to *IVFS* minterm-maxterm interval based fuzzy sets (*MIVFS*) were also proposed for defining a narrower band of F^3-s, whose midpoint offers another alternative for a symmetrical F^3.

In the future we intend to investigate the behavior of complex fuzzy sequential circuits based on *IVFS*, *MIVFS* and midpoint F^3-s. We expect the behavior of such fuzzy networks would be interesting from the viewpoint of the possible convergence/divergence behavior. We also intend to find the optimal interval or point valued F^3-s for practical applications, such as adaptive behavior and learning.

Acknowledgment. Research supported by the Nat. Sci. Res. Fund Grts. # T048832 and # T043258, further by Széchenyi Univ. and Budapest Tech. grants.

References

1. Bouchon-Meunier, B., Kreinovich, V., Nguyen, H.T.: Non-Associative Operations, Second International Conference on Intelligent Technologies, Bangkok, (2001), 39-46
2. Fodor, J.C., Kóczy, L.T.: Some remarks on fuzzy flip-flops, In: Kóczy L.T., Hirota K., (eds.), Proc. of the Joint Hungarian-Japanese Symposium on Fuzzy Systems and Applications (1991), 60-63
3. Goguen, J.A.: L-fuzzy sets, J. of Math. Analysis ands Applications, Vol. 18, 145-174
4. Hirota, K., Ozawa, K.: Concept of fuzzy flip-flop, Preprints of 2nd IFSA Congress, Tokyo, (1987), 556-559
5. Hirota, K., Ozawa, K.: Fuzzy flip-flop as a basis of fuzzy memory modules, In: Gupta M.M. et al., (eds.): Fuzzy Computing. Theory, Hardware and Applications, North Holland, Amsterdam, (1988), 173-183
6. Klir, G.J., Folger, T.A.: Fuzzy sets, uncertainty, and information, Prentice Hall, Upper Saddle River, NJ, (1988), 37-63
7. Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic: Theory and applications, Prentice Hall, Upper Saddle River, NJ, (1995), 50-88
8. Lovassy, R., Kóczy, L.T.: Comparison of Elementary Fuzzy Sequential Digital Units Based on Various Popular T-norms and Co-norms, 3rd Romanian-Hungarian Joint Symposium on Applied Computational Intelligence, Timisoara, (2006), 164-181
9. Lovassy, R., Kóczy, L.T.: Non-Associative Fuzzy Flip-Flop with Dual Set-Reset Feature, 4th Serbian-Hungarian Joint Symposium on Intelligent Systems, Subotica, (2006), 289-299
10. Ozawa, K., Hirota, K., Kóczy, L.T., Ohmori, K.: Algebraic fuzzy flip-flop circuits, Fuzzy Sets and Systems Vol. 39/2, (1991), 215-226
11. Ozawa, K., Hirota K., Kóczy, L.T.: Fuzzy flip-flop, In: Patyra M.J., Mlynek, D.M., (eds.), Fuzzy Logic. Implementation and Applications, Wiley, Chichester, (1996), 197-236

12. Türkşen, I.B.: Interval valued fuzzy sets based on normal forms, Fuzzy Sets and Systems Vol. 20, (1986), 191-210
13. Türkşen, I.B., Esper, A., Patel, K., Starks, S.A., Kreinovich, V.: Selecting a Fuzzy Logic Operation from the DNF-CNF Interval: How Practical Are the Resulting Operations?, Proceedings of the 21st International Conference of the North American Fuzzy Information Processing Society NAFIPS´2002, New Orleans, (2002), 28-33
14. Zadeh L.A.: Fuzzy Sets, Information and Control Vol. 8, (1965), 338-353.

Customized Query Response for an Improved Web Search

Vincenzo Loia and Sabrina Senatore

Dipartimento di Matematica e Informatica Università degli Studi di Salerno via Ponte Don Melillo 84084 Fisciano (SA) – Italy
{loia, ssenatore}@unisa.it

Abstract. Although search engines represent the main means to access on-line data, the increasing demand in terms of performance, precision and relevance in the information retrieval is too far from to being acceptable. The gap existing between the wanted information and the gathered information is often bound to hindrances of semantic rather than syntactic nature. The continuing growth of the Internet usage and contents makes difficult the information access, making the task of information retrieval highly critical.

The paper introduces a system for supporting the Web search activity: on the basis of the interpretation of input query, a suitable list of links to relevant web pages is presented to the user. In fact, the system builds additional queries whose content is similar to the initial one and returns a refined list, resulting from the "multiple" query submissions.

Keywords: Web search, FCM clustering, ontologies.

1 Introduction

The diffusion of the Semantic Web supplies new models to support integrated and uniform access to web resources. Nevertheless, the structure of information is still based on HTML-format for data representation, makes the conversion of the syntactic Web content into the semantic one, an expensive activity in terms of human resources and time.

Large-scale search engines often retrieve long lists of links to web pages, which actually contain only a handful of relevant pages. Usually, the returned web pages contain the terms of the input query, even though the discovered concepts do not meet the user intentions. Search engine quality is mainly based on the capabilities to return results which meet the search goal relevantly. Some search techniques are based on the classification of the discovered documents and then referenced in a search database by human experts or by artificial agents (softbots) (i.e. Yahoo!, Google) [14]. Similarly, other approaches achieve clustering of results for automatic organization (into categories) of documents (i.e. WiseNut [22] and Vivisimo [21]). Metasearch environments, instead implement strategies that apply user queries to several search engines simultaneously.

The query ambiguity and the vocabulary gap represent extant impediments that confirm the search engine technology is far from the ideal response to a certain query.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 653–662, 2007.
springerlink.com

Also, the Web search engines can not often consider the semantic relationships existing among terms.

The last trend sees the diffusion of tagging services for allowing users to add terms of their own interest and relevance to Web pages [5, 16]. Many approaches have been proposed for personalized web search systems; Web assistant systems manage information overload by proposing to the user personalized information, through the monitoring of her/his activities during the Web navigation.

Remarkable issues are reached in web usage mining through advanced fuzzy frameworks [4, 1, 12]. Some approaches exploit the agent technology for supporting users during the web browsing in adaptive Web-based environments [6, 8, 9] and for tracking the user navigational behaviour [6, 8, 9]. The opening of the Web capacity toward the online marketers and business[15], conquers large customer audience, through the online advertising spreading, customized to the evinced interests of users [23, 2].

This framework aims at assisting the user during the query formulation based on the semantic of the terms and at the same time, at producing a more accurate list of relevant results through fuzzy techniques.

The paper is organized as follows. Section 2 introduces the architecture description by sketching the working flow; then the main activities are deepened by presenting the relative formal model in Section 3. Then, the experimental results are described in Section 4 and finally the conclusions close the paper.

2 Architectural Overview

Figure 1 provides a straightforward overview of the whole framework, by presenting the activities diagram; in particular, through the tracing of the data flow, and the principal joined interactions, the main activities of the system are depicted, as follows.

- *Query Acquisition & Analysis*: the input of the query triggers this activity which starts up the system flow. The query is acquired through an aid-user interface, then it is analyzed syntactically by collecting only the relevant words and discarding all the non-informative words. The result of this activity is a sequence of ``atomic" words that represents the user query.
- *Semantical Augmentation*: once the query is reduced to the basic form, this activity aims at enriching the interpretation of the query, through the discovering of feasible ontological relationships, in order to get a possible augmentation of the input query. Conceivable result is a sequence of terms composed of query terms and (eventual) additional terms, semantically related and discovered during the analysis of ontological relationship.
- *Vector-based Profiling*: in this phase, for each query, a corresponding vector representation is produced and locally stored. Each vector is composed of a sequence of weighted values computed for some query keywords (further details are given in Section 3). All the collected vectors are arranged in data-matrix form to give as input to the *Clustering* activity. Furthermore, the obtained clustering is exploited to ``locate" the vector-based representation of the current query: the closest clustered vectors represent the queries with similar topics and they will contribute to add further expressiveness to the input query.

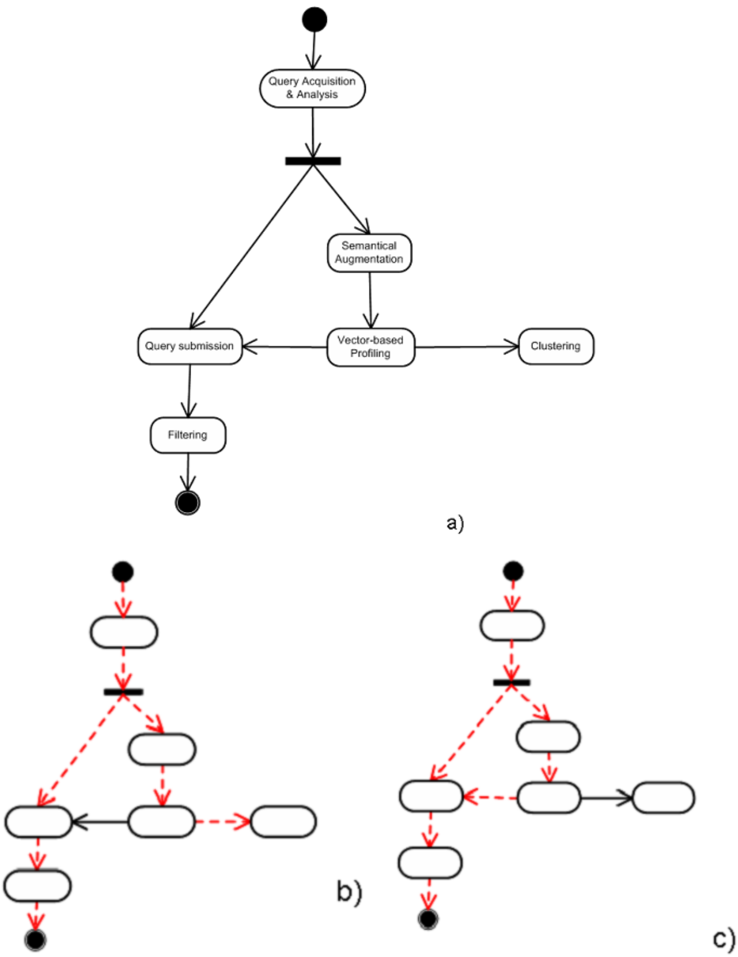


Fig. 1. Activities diagram of the system a) and the main action flow b)-c)

- *Clustering*: the queries, in form of matrix, are elaborated to generate groupings of them. The obtained clusters are associated to categories of topics related the input queries. Periodically, the clustering is re-executed, triggered by the *Vector-based Profile* activity.
- *Query Submission*: the ``atomic" input query (coming from *Query Acquisition & Analysis*) as well as some queries (in the clustered collection) that are similar the input topic (coming from *Vector-based Profiling*) are individually submitted to a search engine.
- *Filtering*: the lists of pages links, retuned by the search engine, for each one of submitted queries are then filtered and merged to get a unique ranked list of page links.

As evidenced in Figure 1 b)-c), the system consists of two main action flows:

- *start-up* : in this phase of the training, the system collects the queries, stores them and then applies the clustering to get grouping with semantic similarity. Figure 1 b) shows all the activities involved (the dotted line represent the data flow). In figure, a fork point is evidenced, producing a split of execution path. Thus, on one hand, the system replies to the input query of the user, by providing the list of links as in the typical web search approach (*Query Submission* → *Filtering*), on the other hand, collects the information (query terms, semantic relationships, etc.) to achieve a local arrangement of the acquired knowledge, through the clustering procedure (*Semantical Augmentation* → *Vector-based Profiling* → *Clustering*).
- *regime*: Figure 1 c) shows this execution flow. The system achieves the *Query Submission* activities, exploiting the clusters obtained in the previous phase (without re-execute the *Clustering* activity: *Semantical Augmentation* → *Vector-based Profiling* → *Query Submission*). More specifically, the query is compared with the clustered queries: those queries that are more similar to the given one, are exploited to be executed in order to find additional relevant pages. In fact, let us assume that although similar queries include terms semantically or syntactically close, not necessarily, their submission produce lists of page links which are very similar. Thus, merging these lists could guarantee a retrieval of more relevant web pages. At this point, an appropriate filtering of the joined lists provides one final list of links to present to the user.

3 Analysis of the Implementation Choices

This section presents a more complete description of the introduced activities. The theoretical aspects and the implementation choices will be analyzed, in order to give a more clear characterization of the system and the inherent interactions.

3.1 Query Acquisition and Processing

This activity captures the meaningful terms which compound the input query. Initially, these terms are analyzed to reduce their number to the minimum, by applying typical approaches to text parsing and analysis: after the stop-word removal, a stemming algorithm [13] reduces the variant forms of a word to a common root. Our system accepts the query sequence and then elaborates it to discriminate the main concepts.

A further semantic analysis helps to discriminate the main concepts of the query, through an RDF-based tagging of the terms and an enrichment of feasible ontological relationships. In fact, the user can arbitrary insert appropriate meta-terms, chosen in a specified ontology. The query “all the people interested in the ‘Semantic Web Services’ domain” could become:

<rdf:Class> People </rdf:Class>
<foaf:topic_interest>Semantic Web Services</foaf:topic_interest>

The prefixes rdf and foaf are associated to the namespaces of the reference ontologies where these meta-terms are defined. The prefix rdf is the "default" prefix, and introduces the basic concepts, whereas the prefix foaf refers to a specific ontology of the homonymic project [7] which describes homepages for people, groups, etc.

3.2 Semantic Augmentation

The RDF-based tagging of the terms aims at detecting and/or disambiguating the actual meaning of the words. The system maintains a prefixed set of ontologies (ad-hoc defined or already existing [17, 18]), which represents the local *natural knowledge*. Each term of query is looked for in the ontologies in order to individuate the domain of reference: if a term appears in more than one ontologies, a further analysis on the remaining query terms is applied. The reference domain is defined by the ontology which describes more query terms. If a term is not defined in any ontology, it will be not considered in the query evaluation and then, it will be put in a list of *unknown terms*. Then, the query is enriched with ontology terms that are related (through specialization/generalization) to the input query.

The augmentation of the initial query represents the *knowledge acquired* by the system for supporting the semantic interpretation of the query.

3.3 Vector-Based Profiling

During this activity, the input queries are translated into vectors which then are stored locally. In the *start-up* phase, the system predisposes the data matrix to give as input to *Clustering* activity. Specifically, each query is a sequence of terms, composed of its own terms as well as those ones elicited by the *Semantical Augmentation* activity. In order to build the data matrix, a set of common terms (among the terms of all the queries), a common set of them, called *reference set* is defined. The criterion is to select the most recurring terms among all the obtained queries terms. Once defined the *reference set*, the vectors can be built: each cell is in correspondence of a query term and has a weight value. A simple algorithm of selection provides different "weight" to the terms, according to their origin: the *initial* terms typed by the user (come from the user query) and the *derivative* term, obtained by ontological relationships.

More specifically, let T be the *reference set*, composed of *initial* and *derivative* terms. Given a term t_j, a function *weight* : $T \rightarrow [0,1]$ is defined as follows:

- $t_j \in T$ and is an *initial* term $\Rightarrow$ weight(t_j) = 1
- $t_j \in T$ and is a *derivative* term $\Rightarrow$ weight(t_j) = $1/2 + 1/(2*s + 1)$
- $t_j \notin T$ $\Rightarrow$ weight(t_j) = 0

where s is the minimum number of edges (i.e. the minimum path) included between some *initial* term t and the derivative term t_j in the reference ontology O.

We give a high relevance to the terms of the reference set that the user type in initial query, because they represent the user's intention; thus the associated weight is the maximum value. When the term is *derivative*, a relevance is evaluated according to its correlation with an initial term (computed as minimum number of edges s that separate the two terms). Thus, the weight should be greatest when the number of edges is the smallest (i.e. when $s = 1$, it means one edge is between one initial term and the derivative term t_j, i.e. a direct relationship exists, thus we associate a weight close to 1).

Hence, given a collection of queries $Q = \{q_1, q_2, \ldots, q_n\}$, each query q_i is represented by a h-dimensional vector $q_i = \{w_{i,1}, \ldots, w_{i,j}, \ldots, w_{i,h}\}$, where each $w_{i,j}$ is the weight associated to a term $t_j \in T$ with $j = 1, \ldots, h$. These vectors form a data matrix $n \times h$ which is the input of the clustering activity.

3.4 Clustering

Our clustering approach is based on the well-known Fuzzy C-Means algorithm (shortly, FCM) [3]. The FCM algorithm takes as input a collection of patterns of a universe U (in our case, the collection of queries Q) in form of matrix and produces fuzzy partitions. The output is a partition of the given patterns (queries) into (prefixed) c clusters. Each cluster is represented by its center point (prototype). This process is completely unsupervised, aimed at identifying some inherent structures in a set of data.

More formally, the FCM algorithm aims at minimizing the objective function constituted by the weighted sum of the distances $dist_{i,j}$ between data points $\underline{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,h})$ and the prototypes $v_j = (v_{j,1}, v_{j,2}, \ldots, v_{j,h})$, according to this formula:

$$\sum_{i=1}^{c} \sum_{j=1}^{n} u_{i,j}^{m} dist_{i,j}^{2} \tag{1}$$

where $c \geq 2$ is the number of clusters, $u_{i,j} \in [0,1]$ is the membership degree of $\underline{x}_j$ ($j = 1, \ldots, n$) in the i-th cluster A_i ($i = 1, \ldots, c$), and $m > 1$ is the fuzzifier, which controls the quantity of fuzziness in the classification process (common choice of fuzzifier is $m = 2$) and finally $dist_{i,j}$ just represents the euclidean distance between the data x_i and the center v_j of the j-th cluster.

The data matrix is composed of n queries, each one with h values $w_{i,k}$ ($i = 1, \ldots, n$ and $k = 1, \ldots, h$), associated to the corresponding term. In our approach the generic vector $\underline{x}$ is a query $\underline{p} \in P$ as defined above, whereas the center v of the cluster represents the sample topics of that cluster. As said, the output of clustering is a fuzzy partitions, where each cluster describes topics or subjects (relative to the queries terms). Consequently, a straightforward mapping can be defined, which associates these clusters to categories of topics. Furthermore, each query of the input set assumes a membership value for each cluster.

The clustering is a scheduled activity which is re-executed, whenever the number of queries which are not represented by no defined categories increases.

3.5 Query Submission and Filtering

The data flow coming from *Vector-based Profiling* and *Query Acquisition & Analysis* converges in the *Query Submission* activity: substantially the original query as well as the closest clustered queries (one or more) are separately "submitted" to a search engine. Before presenting to the user, the resulting lists of pages links are merged and then selected by the *Filtering* activity which combines the links, producing a unique ranked list of links. The first *k* pages links are organized as a typical web page e returned to the user. The produced final list represents a suitable balancing between the input query (often ambiguous and not enough explicative to be processed by a search engine) and the stored stereotype of queries (which can reflect the input query partially). This approach does not apply a further analysis on the possible user feedback.

4 Performance Evaluation

The system performance is evaluated, by considering the *start-up* and *regime* phases of activities. In the *start-up* phase, the performance evaluation is based on the clustering execution of the collected queries. We have built an "ad hoc" sample, by applying an appropriate selection of the query keywords. Afterwards, in the *regime* phase, the system uses the computed categories for detecting the proximity of the input query with the clustered queries. The testing set simulates the interaction of web communities of users with the system.

In our opinion, in the Web searching activity of Web communities, the argumentations are often based on a limited number of topics, bound to the common context of the groupings of people which share expertise and interests on specific subjects. This restriction meets the limited though manifold set of terms used in the initial training; on the other hand, the context of web communities is suitable to assess the performance of our system. However, the *start-up* phase, which carries out a classification of input queries, is evaluated trough the two well-known measures: the *precision* and the *recall*.

In our approach, we evaluate the recall as the ratio of the number of relevant, well-classified queries to the total collection of collected queries. A "well-classified" query is a query that, after the clustering, will appear in the expected class. In fact, according to the set of input queries, a set of categories have been predefined. The same way, the precision represents the percentage of all the well-classified queries, with respect to the expected clusters.

Let us note the recall and precision are two measures exploited for the evaluation of crisp clustering. Herein we assume that a query belongs to that cluster whose membership is the highest. The queries, whose membership is equally distributed among all the clusters, are not considered in the evaluation.

In the *regime* phase instead, the performance of the system is estimated by measuring the search effectiveness, given a set of submitted queries; thus we consider the micro-average of the individual precision-recall curves, exploiting a number λ of steps up to reach the highest recall value. According to [19], let Q be the set of queries, D all the relevant pages (i.e. all the documents that match the queries). The micro-averaging of recall and precision (at step λ) is defined as follows:

$$\mathrm{Re}\,c_{\lambda} = \sum_{R} \frac{|D_R \cap B_{\lambda,R}|}{|D|}, \qquad \mathrm{Pr}\,ec_{\lambda} = \sum_{R} \frac{|D_R \cap B_{\lambda,R}|}{|B_{\lambda}|}, \tag{2}$$

where D_R is the answer set of pages for given query request R, B_λ is the set of retrieved pages at the step λ and $B_{\lambda,R}$ is the set of all relevant pages, retrieved at the step λ.

4.1 The Experimental Results

Table 1 shows the performance of clustering in term of recall/precision (*start-up* phase). Each row represents an execution, with a given parameter setting: the number of submitted queries, the analyzed pages of the list of links, the cardinality of reference set, the number of clusters, finally the evaluation of recall and precision, expressed in percentage.

In general, the clustering performance presents interesting results: the computed values in terms of recall and precision are rather accurate for the most of experiments.

Let us note that the performance tends to improve when the ratio *query/pages* is less than 1 (i.e. when for each query, the number of associated page links is greater than one).

The *regime* phase evaluates a further hundred of queries submitted to the system: we measure the improvement of the system, by considering the "acquired knowledge", gathered in the *start-up* phase. As said, in fact, the initial query is enriched with some additional terms, extracted by those clustered queries which are more similar in topics.

Table 1. Evaluation measures of clustering performance

Exp. No.	# query	#pages	#reference terms	#clusters	Recall	Precision
1	50	50	15	3	93.6%	94.1%
2	50	100	22	3	91.9%	97%
3	80	120	21	4	93.1%	96.7%
4	100	300	35	5	86.9%	95.8%
5	120	360	35	6	95.9%	95.5%
6	150	450	44	6	94.3%	95.3%

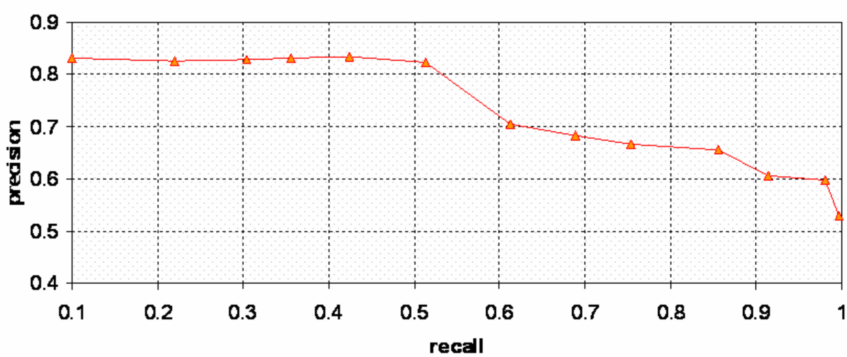


Fig. 2. Micro-averaging of recall/precision

Figure 2 shows the graph of micro-averaging of recall and precision at different steps λ. According to the characterization of these two measures, there is a reverse connection between these two measures: an higher precision usually implies a lower recall and vice versa. In fact, Figure 2 shows, on the average, the recall assumes low values (less than 0.5) when the precision assumes a high and rather constant value (greater than 0.8). Successively, the recall increases, even though the precision never assumes values lower than 0.5.

5 Conclusions

Although at a prototype stage, our approach supports the user during the web search: it improves the quality of reply during the web search activity, taking into account the collected information, the ``knowledge" of the system on a specific topic of the given query. The system provides good results in terms of the performances and the response to the input query.

This approach provides interesting results, when applied on the context of web communities: the obtained classification can be seen a feasible characterization of elicited "virtual" communities of user with common inclinations, interests and topics.

As future extensions, we would like to extend this approach by considering the feedback of the users, when a list of results is proposed. The system, in fact, could take into account the user's behaviour (the clicked links, the visited pages, etc.) when the resulting list of links is returned and then, benefit of such information to improve the query/answer quality.

References

1. Dragos Arotaritei and Sushmita Mitra. Web mining: a survey in the fuzzy framework. Fuzzy Sets and Systems, 148(1):5–19, 2004.
2. Sung Min Bae, Sang-Chan Park, and Sung Ho Ha. Fuzzy web ad selector based on web usage mining. IEEE Intelligent Systems, 18(6):62–69, 2003.
3. J.C. Bezdek. Pattern Recognition and Fuzzy Objective Function Algorithms. Plenum Press, N. York, 1981.
4. Michael Chau, Daniel Zeng, Hsinchun Chen, Michael Huang, and David Hendriawan. Design and evaluation of a multi-agent collaborative web mining system. Decis. Support Syst., 35(1):167–183, 2003.
5. Del.icio.us. http://del.icio.us/.
6. F.C. Peng, C. Ding, J.C. Patra. Personalized web search with self-organizing map. In The 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service 2005, pages 144– 147, 29 March-1 April 2005.
7. FOAF. The Friend of a Friend (FOAF) project . http://www.foaf-project.org/.
8. J. Liu and J. You Smart shopper: an agent-based web-mining approach to internet shopping. IEEE Trans Fuzzy Systems, 11(4):226–237, 2003.
9. Raymond S.T. Lee and James N.K. Liu. iJade Web-Miner: An intelligent agent framework for internet shopping. IEEE Transactions on Knowledge and Data Engineering, 16(4):461–473, 2004.

10. V. Loia , W. Pedrycz , S. Senatore and M.I. Sessa Web navigation support by means of cognitive proximity-driven assistant Agents. Journal of the American Society for Information Science and Technology (JASIST), 57(4):515–527, 2006.
11. N. Boudriga and M. S. Obaidat Intelligent agents on the web: A review. Computing in Science and Engg., 6(4):35–42, 2004.
12. Olfa Nasraoui, Hichem Frigui, Raghu Krishnapuram, and Anupam Joshi. Extracting web user profiles using relational competitive fuzzy clustering. International Journal on Artificial Intelligence Tools, 9(4):509–526, 2000.
13. M. F. Porter. An algorithm for suffix stripping. pages 313–316, 1997.
14. M. R. Henzinger Algorithmic challenges in web search engines. Internet Mathematics, 1(1):115–126, 2003.
15. R.-L. Liu Collaborative multiagent adaptation for business environmental scanning through the internet. Applied Intelligence, 20(2):119–133, 2004.
16. Shadows. http://www.shadows.com/.
17. SUMO WG. Sumo ontology site, 2003. Avaliable URL: http://ontology.teknowledge.com/#FOIS.
18. Semantic Web Technology Evaluation Ontology (SWETO). http://lsdis.cs.uga.edu/projects/semdis/sweto/.
19. C.J. Van Rijsbergen. Information Retrieval, 2nd edition. Dept. of Computer Science, University of Glasgow, 1979.
20. K. Kinshuk T. Vasilakos, V. Devedzic and W. Pedrycz Computational intelligence in web-based education: A tutorial. Journal of Interactive Learning Research, 15(4):299–318, 2004.
21. Vivissimo. http://www.vivissimo.com.
22. WiseNut. http://www.wisenut.com/.
23. Ronald R. Yager. Targeted e-commerce marketing using fuzzy intelligent agents. IEEE Intelligent Systems, 15(6):42–45, 2000.

An Effective Inductive Learning Structure to Extract Probabilistic Fuzzy Rule Base from Inconsistent Data Pattern

Hyong-Euk Lee and Z. Zenn Bien

Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Korea
helee@ctrsys.kaist.ac.kr, zbien@ee.kaist.ac.kr

Abstract. Bio-signal/behavior pattern acquisition and its use are essential in human-friendly human-robot interaction to recognize human intention. However, it is usually difficult to model and handle such interaction due to variability of the user's behavior and uncertainty of the environment in human-in-the-loop system. In this paper, we shall show the benefits of a PFR (probabilistic fuzzy rule)-based learning system to handle inconsistent data pattern in view of combining fuzzy logic, fuzzy clustering, and probabilistic reasoning in a single system as an effective engineering solution to resolve inconsistency of the complicated human behavioral characteristics. Moreover, we introduce a PFR-based inductive life-long learning structure for continual adaptation throughout incessant learning and control. The learning system gradually extracts more meaningful/reliable rule-based knowledge in incorporation of learning processes in short-term memory, interim transition memory and long-term memory. To show the effectiveness of the proposed system, we introduce a successful example as a case study in view of probabilistic fuzzy rule-based knowledge discovery to handle TV watching behavior data pattern learning.

Keywords: Probabilistic Fuzzy Rule, Inductive Learning, Life-long Learning.

1 Introduction

The robots that perform works and activities for human beings are called service robots [1]. Nowadays, increasing attention is being paid toward service robotic systems as many nations are to confront with the demographic crisis of aged population and also as the issue of welfare becomes a more keen public interest. Various service robotic systems have been developed to help human beings and to coexist with him/her in home environment as well as public place.

When one deals with a robot that is to assist people with disability or the elderly, such a machine-centered design methodology is doomed to fail because training such people to adapt to the complicated machine is simply impossible. Naturally, we may think of the other opposite approach of human-centric design. In this approach, the procedure is to study human behaviors and characteristics first and, then, design the robot that can hopefully adapt to human characteristics. This approach is often called the “human-friendly system” approach as well [2].

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 663–672, 2007.
springerlink.com

Note that HRI (Human-Robot Interaction) is frequent and intensive in service robotic environment. In this situation, bio-signal acquisition and its use are essential in human-friendly HRI to recognize human intention and to understand human's physical status and behavior. From the observations on body gestures (e.g. hand gesture) and some physiological bio-signals (e.g. EMG, EEG, ECG, etc.), we find that, as long as modeling is concerned, human shows formidably complicated characteristics such as high dimensionality, nonlinear-coupling of attributes, subjectivity, apparent inconsistency, susceptibility to environments and disturbances, and time-variance as well as situation-dependency [3]. Among the above explained complex characteristics of human behavior, more specifically, we shall discuss on the characteristics of inconsistency.

In general, an appropriate learning model and a learning algorithm are selected based on analysis about their learning target. In most pattern classification/learning problems, data pattern with well-separable classes are considered as learning target. Hence, a straightforward evaluating method of the classifiers is to use well-separable benchmark data pattern such as UCI repository of Machine Learning Database (http://www.ics.uci.edu).

However, we shall discuss, in particular, data pattern with much higher portions of inconsistent examples in it. For example, consider a problem to model TV watching patterns of an inhabitant in home environment and apply the learned knowledge to an automatic channel recommendation system. To describe this kind of human's behavioral patterns, we may need a lot of accumulated information related to the behavior. However, possible measurement data by available sensors are limited in practical situations, and thus I/O training examples can be sparse and may contain apparently inconsistent examples. That is, we may handle the inconsistency by insufficient/inappropriate features for effective classification. Note that a conventional supervised learning-based classifier, which minimizes a classification error by rejection of inconsistent data, may show low satisfactory performance due to loss of information during the learning process.

Another aspect that we have to pay attention is the inconsistency by time-varying characteristics of the target data pattern. Even though a classifier has a well-describing set of features for given classes, learning target itself may change as time goes on. In other words, currently acquired knowledge on the data pattern may be (partially) inconsistent with the previous one. In this case, we may also consider an adaptive learning methodology for effective classification. As an approach for adaptivity, a life-long learning paradigm, also termed continuous learning, emphasizes learning through the entire lifespan of a system [4]. Also, *S. Grossberg* mentions, in contrast to only adapting to a changing environment, that life-long learning suggests preserving previously learned knowledge if it does not contradict the current task [5].

In this paper, we shall show the benefits of a PFR (probabilistic fuzzy rule)-based learning system to handle inconsistent data pattern in view of combining fuzzy logic, fuzzy clustering, and probabilistic reasoning in a single system as an effective engineering solution to resolve inconsistency of the target data pattern. The learning system tries to achieve a more meaningful rule base from this data pattern with capability of distinguishing between separable clusters for classification and inseparable clusters for probabilistic interpretation. Also, we introduce a PFR-based inductive learning structure for continual adaptation and reliable knowledge extraction throughout incessant learning and control in view of life-long learning. In this context, we refine the concept of life-long learning as repeating an inductive learning process and a deductive learning process through the entire lifespan of the system in consideration of interaction between a learning agent and feedback signal (by human agent) for human-in-the-loop system [3]. In the inductive learning process, data are collected, are transformed into a certain type of knowledge and are accumulated in a database from incrementally drawn data patterns. In the deductive learning process, the learning system tries to provide control output using the acquired knowledge. And then, it strengthens/modifies the knowledge database by receiving rewards from the environment. The learning system gradually extracts more meaningful/reliable rule-based knowledge incorporation of learning processes in short-term memory, interim transition memory and long-term memory.

This paper is organized as follows. In Section 2, a PFR-based learning system is explained. Based on the PFR-based learning system, we extend it to an overall inductive learning structure in Section 3. To show the effectiveness of proposed learning structure, we illustrate a successful example as case study in view of probabilistic fuzzy rule-based knowledge discovery to handle TV watching behavior pattern as well as benchmark data sets in Section 4. Finally, a brief conclusion is stated in Section 5.

2 Probabilistic Fuzzy Rule-Based Learning System

2.1 Probabilistic Fuzzy Rule Base Representation

We briefly describe probabilistic representation of a fuzzy rule base to familiarize notational conventions. In a classical fuzzy rule-based classifier, the antecedent part of a rule defines the operating region of the rule in the M -dimensional input feature space while the consequent part of the rule describes one of the K classes, indicating a crisp class label from the label set $\{c^1, c^2, \ldots, c^K\}$. Compared with the classical fuzzy rule-based knowledge, the probabilistic fuzzy rule has the following form in Eq. (1) [6]:

$$\begin{aligned} R_i : &\textit{If } x^1 = \tilde{A}_i^1 \textit{ and } x^2 = \tilde{A}_i^2 \textit{ and}, \ldots, \textit{and } x^M = \tilde{A}_i^M \\ &\textit{then } \Pr(y = c^1 \mid x = \tilde{A}_i) = P_i^1, \Pr(y = c^2 \mid x = \tilde{A}_i) = P_i^2, \\ &\ldots, \Pr(y = c^K \mid x = \tilde{A}_i) = P_i^K, \;\; i = 1, \ldots, N. \end{aligned} \tag{1}$$

where $x = (x^1, x^2, \ldots, x^M)$, $\tilde{A}_i = (\tilde{A}_i^1, \tilde{A}_i^2, \ldots, \tilde{A}_i^M)$, and R_i denotes the i th rule. Note that the vector fuzzy set $\tilde{A}_i$ is assumed to have its membership function with shape of multivariate Gaussian function. To simplify the expression, we also assume that the covariance matrix has diagonal terms only in the fuzzy set. It is remarked that an ordinary fuzzy rule can be considered a special case of a rule expressed in Eq. (1). The totality of rules, R_i, $i = 1, \ldots, N$ is called the probabilistic fuzzy rule base (PFRB). We can use a PFRB in classification problem and assign a feature vector x to the class c^{k^*} using Eq. (2).

$$k^* = \arg\max_{k=1,2,\ldots,K} \sum_{i=1,2,\ldots,N} \mu_{\tilde{A}_i}(x)\Pr(y = c^k \mid x = \tilde{A}_i) \cdot \tag{2}$$

2.2 Probabilistic Fuzzy Rule Base Extraction

While fuzzy logic is effective to deal with linguistic uncertainty in describing the antecedent part of a rule, the probability theory can be effective to handle probabilistic uncertainty in describing the consequent part of the rule. In many fuzzy rule-based systems, a pre-defined homogeneous (grid-like) fuzzy partition is initially employed, and then, a set of fuzzy rules for the partitioned input space is acquired from numerical input-output data. The initial base can be further simplified by additionally applying a rule reduction technique [7]. Note that partitioning can affect system performance in a notable way. A well-partitioned input space, for example, can induce a reduced set of rules to describe the given data pattern with high interpretability, whereas an ill-partitioned input space may generate redundant rules some of which can be even conflicting or inconsistent. Thus, a methodology of extracting fuzzy rules with meaningful self-organized fuzzy partition is desirable.

In designing the rule base of a fuzzy classifier, an alternative approach, called "fuzzy clustering", is available, which, based on a cluster analysis and clustering algorithms with cluster validity, aims at organizing and revealing structures in data sets such as temporal waveforms, images, etc [8][9]. Note, however, that clustering itself is originally an unsupervised learning process that deals with unlabeled data patterns, and as such, a fuzzy clustering method without prior information on the number of clusters may not guarantee its performance in constructing a proper fuzzy rule base, specially a PFRB for complex data patterns. We remark that several studies about supervised clustering [10][11] are available as modified version of the FCM(Fuzzy C-Means) clustering algorithm [8], in which some prior knowledge such as the number of clusters in the data pattern is assumed given.

We think that effective combination of an unsupervised learning process with a proper supervisory scheme can be helpful to search general regularities in data patterns, and in particular, in finding more separable/analyzable groups of geometrical shapes. More specifically, we present an iterative fuzzy clustering algorithm [12] with supervising scheme in the learning process to construct a PFRB. The learning system starts with a fully unsupervised learning process with the FCM clustering algorithm and a cluster validity criterion, and then gradually constructs meaningful fuzzy partitions over the input space. The corresponding rules with probabilities are obtained through an iterative learning process of selective clustering with supervisory guidance based on cluster-pureness and class-separability.

2.3 IFCS (Iterative Fuzzy Clustering with Supervision) Learning Algorithm [12]

At the initial stage, we do not have any proper means to construct fuzzy partition in a specific shape and to extract probabilistic information without prior knowledge. This is why we start clustering in a fully unsupervised manner with the unlabeled data set.

Once the clustering process is conducted, we can evaluate validity of the obtained partition matrix U by FCM clustering algorithm in view of 1) the number of clusters, and 2) the separability of a cluster. The number of clusters is related to the unsupervised learning process and the separability of a cluster is related to the supervised learning process. To describe the IFCS algorithm in more detail, we defined cluster-pureness and class-separability.

The cluster-pureness gives information that there exists a dominant class in the cluster i. If a cluster has a dominant class, the cluster-pureness increases up to 1. The class-separability gives a maximum class-separability measure in the cluster i among all the combination of two different classes assuming that each class in the cluster i is modeled as a single normal distribution. In other words, the class-separability gives information whether there exists a pair of separable classes in each cluster. If the class-separability is near to 0, there are no separable classes in the cluster i while existence of separable classes us assured with high probability if it is near to 1.

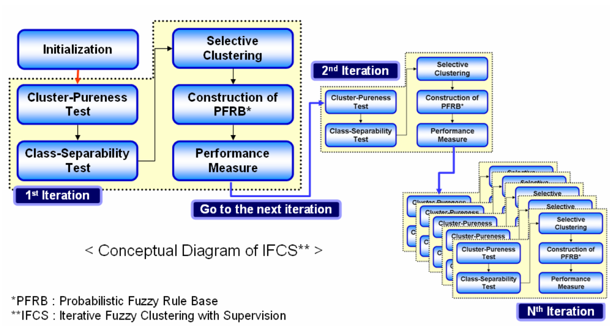


Fig. 1. Learning Procedure of IFCS (Iterative Fuzzy Clustering with Supervision) Algorithm

Given numerical data pattern, initial unsupervised clustering with a cluster-validity criterion (Xie-Beni index [13]) is conducted with selection of the first local minimum, which is the initial number of clusters. Then, the most mixed cluster with different classes is selected among the initial clusters in the cluster-pureness test. In the class-separability test, the existence of separable classes in the selected cluster is examined using the class-separability measure. If there are separable classes, re-clustering of the selected cluster is performed. If not, extraction of probabilistic information from the selected cluster is conducted.

Meaningful fuzzy partitions with a corresponding rule base are gradually obtained through an iterative process of selective fuzzy clustering with supervisory guidance based on cluster-pureness and class-separability. The iterative process enables fine tuning of the rule base with high reliability compared with the conventional validity measure and typical clustering algorithms. Overall learning procedure is shown in Fig. 1.

3 Inductive Learning Structure for PFR-Based Learning System

Fig. 2 shows a structure for an inductive learning process for extraction of probabilistic fuzzy rule base using IFCS algorithm. Note again that the concept of life-long learning includes deductive learning process, however, we shall focus on inductive learning phase and control phase only.

In the inductive learning phase, overall learning takes place in three processes called short-term memory (STM), interim transition memory (ITM) and long-term memory (LTM). During a given monitoring time, the learning system collects training examples and extracts a PFRB. After that, the learning system tries to provide control output/classification result using acquired knowledge in the LTM.

In the STM, the IFCS learning algorithm generates a PFRB from a set of training examples which are observed during each time period. PFRB in STM is considered to be a temporal rule base, while ITM is a pool of possibly reliable PFRs for control. By a fuzzy membership-similarity measure between rules, DM1(decision maker 1) adds or merges each rules from STM into ITM. If the DM1 decides to merge the incoming rule into existing rule base in the ITM, a probabilistic similarity-measure between rules is calculated. And it is reflected in updating the reliability measure value of the merged probabilistic fuzzy rule. DM2 selects probabilistic fuzzy rules from ITM to be transferred to LTM according to the reliability measure. LTM is the storage of a PFRB for control phase.

To be more specific, we introduce a reliability measure of a PFR in Eq. (2) and a similarity measure between PFRs in Eq. (3). Let an incoming PFRB to the ITM from the STM be $R_{new} = \{R_{A1}, R_{A2}, ..., R_{An_{new}}\}$ and an existing PFRB in the ITM be $R_{prev} = \{R_{B1}, R_{B2}, ..., R_{Bn_{prev}}\}$. A similarity measure of the two PFRs is defined as in Eq. (3) adopting Zadeh's sup-min compatiability. Similarly, a probability-similarity of the two PFRs can be defined as in Eq. (4).

$$Sim(R_{A_k}, R_{B_l}) = \sup_{u \in U} \min(\mu_{R_{A_k}}(u), \mu_{R_{B_l}}(u)) \tag{3}$$

where $1 \le k \le n_{new}, 1 \le l \le n_{prev}$.

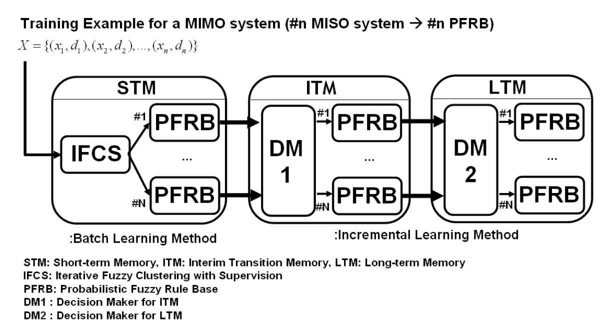
$$\text{Prob\_}Sim(R_{A_k}, R_{B_l}) = Sim(R_{A_k}, R_{B_l}) \cdot \left(\sum_{i=1}^{K} \min\left(\Pr(y = c^i \mid R_{A_k}), \Pr(y = c^i \mid R_{B_l}) \right) \right)$$

where $1 \le k \le n_{new}, 1 \le l \le n_{prev}$ (4)

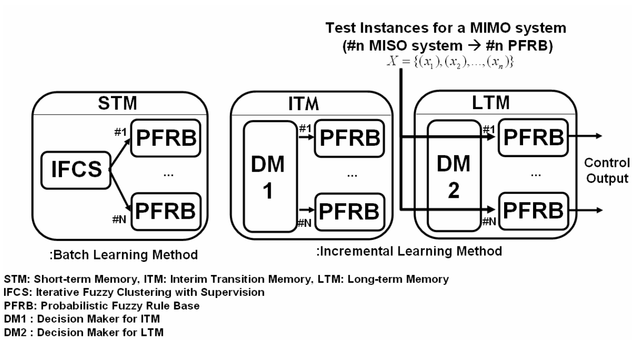
For each $R_{Ai} (i = 1, ..., n_{new})$, the learning process in the ITM finds $R_{A_{l^*}}$ such that $l^* = \arg\max_{l=1,2,...,n_{prev}} Sim(R_{A_i}, R_{B_l})$. If $Sim(R_{A_i}, R_{B_{l^*}}) < \beta (= 0.5)$, R_{A_i} is added in the ITM because the rule is regarded as a new one. If not, R_{A_i} and $R_{B_{l^*}}$ are merged in the ITM and reliability measure $r_{R_{merged}}(t)$ of the merged rule is calculated recursively as in Eq. (5).

$$r_{R_{merged}}(t) = 1 - e^{-\gamma\left(x_{new} - \frac{1}{\gamma}\ln(1 - r_{R_{merged}}(t-1))\right)}, \text{ where } x_{new} = \text{Prob\_}Sim(R_{A_k}, R_{B_{l^*}}) \quad (5)$$

If $r_{R_{merged}}(t) > \eta(=0.5)$, the merged rule is transferred to the LTM as a reliable PFR. Reliability measure of a probabilistic fuzzy rule is recursively calculated along the curve $y = 1 - e^{-\gamma x}$, where x is an accumulated value of the probabilistic similarity-measure and γ is a design parameter. More reliable PFRB are gradually stored in the LTM by means of repeated learning processes, and they are finally considered to be adequate for control phase.



(a) inductive learning phase



(b) control phase

Fig. 2. A Structure for Inductive PFR-based Learning System

4 Experimental Result

4.1 Classification Capability with Benchmark Data Pattern

Even though the proposed system is originally designed to deal with inconsistency, however, it also shows a good performance as conventional classifiers when we deal with well-separable data patterns. We consider 2 data sets which are `Wine' and `Wisconsin breast cancer' data sets from UCI data sets. The wine data contains the chemical analysis of 178 wines grown in the same region in Italy but derived from 3 different cultivars. The classification problem of the wine data set is to distinguish 3 different types based on 13 continuous attributes derived from chemical analysis. The Wisconsin breast cancer data have 683 examples with incomplete examples excluded. For the Wisconsin breast cancer data set, the task is to distinguish 2 classes between benign and malignant cancers based on the available 9 measurements: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nuclei, and mitosis. Using 10-fold cross validation method, we achieved 95.4% (3 rules) for the wine data set and 97.0% (3 rules) for the breast cancer data set by the proposed learning system.

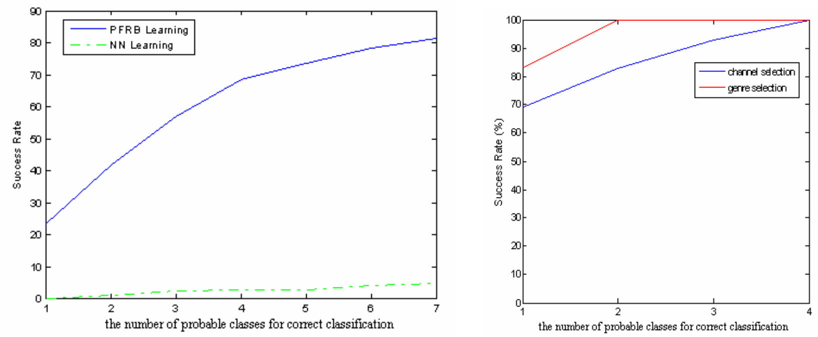
4.2 Classification Capability with TV Watching Data

We deal with TV watching data of inseparable patterns that we have obtained for a practical learning problem. The data set contains a large portion of inconsistent data which is possibly due to usage of inappropriate features. To be more specific, we consider TV watching activities in a residential space as a recognition problem of living behavior patterns. The task of the learning system is to distinguish the favorite TV channel of the inhabitant based on the channel/genre selection information. The number of classes is 200 for channel selection and 9 for genre selection. The data pattern was collected for a month based on two features (time and day of week).

Note, in general, that if the number of classes is much larger than the number of the available features, we may observe many apparently inconsistent examples. Also note that the number of usable sensors is usually scarce to characterize a behavior pattern such as TV channel selection. In this case, success rate which considers the most probable class only cannot give a reasonable (absolute) evaluation criterion of a learning system because it may not be important in the probabilistic system. Therefore, we consider, as an alternative approach to evaluate a probabilistic system, n^{th} ($1 \leq n \leq c$, c is the number of classes) probable classes for correct classification.

Let us consider an unfiltered raw TV watching data. The data pattern contains large portion of inconsistent examples because it includes not only intentional watching actions but also channel-seeking actions or unintentional actions. In this case, we can not achieve high success rate for classification as shown in Fig. 3 (a), however, we can also observe that the conventional supervised learning-based classifier such as a neural network(NN) marks much more low success rate because of the inconsistent examples (10 hidden neurons and back-propagation algorithm has been used in the NN for comparison). Otherwise, when we deal with a filtered TV watching data, where the data under 10 minutes' watching time have been excluded for training set, we achieve a higher success rate as shown in Fig. 3 (b).

Note that a merit of probabilistic system is to provide the next probable output/class in turn. It means the learning system can recommend favorite channels with 70% of success rate within 4 trials as shown in Fig. 3 (a) even with unfiltered data. Furthermore, as shown in Fig. 3 (b), the learning system can recommend the favorite TV channels with 80% of success rate and the genres with 100% of success rate within 2 trials.



(a) the result using unfiltered data set for channel selection

(b) the result using filtered data set for channel/genre selection

Fig. 3. Success Rate of the TV Watching Pattern Recognition (Channel and Genre)

5 Conclusion

Intention reading of human being becomes an important technology for human-friendly man-machine interaction in service robotic environment. And learning technique is considered to be an essential component for the intention reading by means of automatic extraction of meaningful knowledge from the collected information.

In this paper, in particular, we have proposed an effective PFR-based inductive learning structure to handle inconsistent data pattern in view of combining fuzzy logic, fuzzy clustering, and probabilistic reasoning. The learning system gradually extracts more meaningful/reliable rule-based knowledge in incorporation of learning processes in STM, ITM, and LTM from the inconsistent data pattern intermingled with separable data pattern. Note, however, that inductive learning process cannot guarantee the validity/quality of the acquired knowledge. Therefore, we remark that a more feasible validation/correction capability of the acquired knowledge (by means of effective hybridization of human knowledge/feedback) is desired in view of life-long learning especially for human-in-the-loop system.

Acknowledgments. This work was partially supported by the SRC/ERC program of the MOST/KOSEF under grant #R11-1999-008.

References

1. M. Hillman: Rehabilitation Robotics from past to present-historical perspective, Proceedings of ICORR 2003, Daejeon, Korea, April, (2003)
2. Z. Zenn Bien and Hyong-Euk Lee: Effective Learning System Techniques for Human-Robot Interaction in Service Environment", accepted to International Journal of Knowledge-based Systems (2007)
3. Z. Zenn Bien: Learning Techniques in Service Robotic Environment, Applied Artificial Intelligence: Proceedings of 7th International FLINS Conference, Genoa, Italy (2006)

4. Fred H. Hamker: Life-long learning Cell Structures-continuously learning without catastrophic interference, IEEE Transactions on Neural Networks, Vol. 14 (2001)
5. S. Grossberg: Nonlinear neural networks : principles, mechanisms and architectures", Neural Networks, Vol. 1, No. 1, (1988) 17-61
6. Nefti, S.; Oussalah, M.: Probabilistic-fuzzy clustering algorithm, IEEE International Conference on Systems, Man and Cybernetics, Vol. 5, Oct. (2004) 4786 – 4791
7. H. Ishibuchi and T. Yamamoto: Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining, International Journal of Fuzzy Sets and Systems, vol. 141, (2004) 59–88
8. J.C. Bezdek: Pattern Recognition with Fuzzy Objective Algorithms, Plenum, New York: (1981)
9. Nikhil R. Pal and J. C. Bezdek: On Cluster Validity for the Fuzzy c-Means Model, IEEE Trans. Fuzzy Systems, vol. 3, No. 3, (1995) 370-379
10. W. Pedrycz: Fuzzy clustering with a knowledge-based guidance, Pattern Recognition Letters, vol. 25, (2004) 469–480
11. J. Abonyi and F. Szeifert: Supervised fuzzy clustering for the identification of fuzzy classifiers," Pattern Recognition Letters, vol. 24, (2003) 2195–2207
12. Hyong-Euk Lee, Kwang-Hyun Park and Z. Zenn Bien: Iterative Fuzzy Clustering Algorithm with Supervision to Construct Probabilistic Fuzzy Rule Base from Numerical Data, Submitted to IEEE Transactions on Fuzzy Systems (2006)
13. X.L. Xie and G. Beni, "A Validity Measure for Fuzzy Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 8 (1991) 841-847

Robust Stability Analysis of a Fuzzy Vehicle Lateral Control System Using Describing Function Method

Jau-Woei Perng[1], Bing-Fei Wu[1], Tien-Yu Liao[1], and Tsu-Tian Lee[2]

[1] Department of Electrical and Control Engineering, National Chiao Tung University 1001, Ta-Hsueh Road, 300 Hsinchu, Taiwan
jwperng@cn.nctu.edu.tw, bwu@cc.nctu.edu.tw, tienyu@cssp.cn.nctu.edu.tw

[2] Department of Electrical Engineering, National Taipei University of Technology 1, Sec. 3, Chung-Hsiao E. Rd. 106 Taipei, Taiwan
ttlee@ntut.edu.tw

Abstract. In this paper, the robust stability analysis of a fuzzy vehicle lateral system with perturbed parameters is presented. Firstly, the fuzzy controller can be linearized by utilizing the describing function method with experiments. After the describing function is obtained, the stability analysis of the vehicle lateral control system with the variations of velocity and friction is then carried out by the use of parameter plane method. Afterward some limit cycle loci caused by the fuzzy controller can be easily pointed out in the parameter plane. Computer simulation shows the efficiency of this approach.

Keywords: Describing function, vehicle lateral system, fuzzy control, parameter plane, limit cycle.

1 Introduction

It is well known that the describing function is an useful frequency domain method for analyzing the stability of a nonlinear control system especially when the system has hard nonlinear elements, such as relay, deadzone, saturation, backlash, hysteresis and so on. The fundamental description of describing functions can be referred in [1-3]. Some academic and industry researches have been applied the describing function to meet the design specifications. For PI controller design, an iterative procedure for achieving gain and phase margin specifications has been presented in [4] based on two relay tests and describing function analysis. In [5], the describing function utilized for the stability analysis and limit cycle prediction of nonlinear control systems has been developed. The hysteresis describing function was applied to the class AD audio amplifier for modeling the inverter [6]. Ackermann and Bunte [7] employed the describing function to predict the limit cycles in the parameter plane of velocity and road tire friction coefficient. The algorithm computes the limit cycles for a wide class of uncertain nonlinear systems was considered in [8]. As for the intelligent control, some researchers have developed the experimental and analytic describing functions of fuzzy controller in order to analyze the stability of fuzzy control systems [9-11].

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 673–682, 2007.
springerlink.com

Besides, the describing function was also applied to find the bounds for the neural network parameters to have a stable system response and generate limit cycles [12]. In fact, the uncertainties are often existed in the practical control systems. It is well known that the frequency domain algorithms of parameter plane and parameter space [13, 14] have been applied to fulfill the robust stability of an interval polynomial.

In this paper, we apply the describing function of fuzzy controller mentioned in [11] and parameter plane method [14] to consider the robust stability of a vehicle lateral control system with perturbed parameters. A systematic procedure is proposed to solve this problem. The characteristics of limit cycles can be found out in the parameter plane. Computer simulation results verify the design procedure and show the efficiency of this approach.

2 Description of Vehicle Model and Design Approach

In this section, the classical linearized single track vehicle model is given first. The describing function of fuzzy controller is also introduced. In order to analyze the stability of perturbed parameters, a systematic procedure is proposed to solve this problem by the use of parameter plane method. In addition, the control factors are also addressed.

2.1 Vehicle Model [7]

Figure 1 shows the single track vehicle model and the related symbols are listed in Table 1. The equations of motion are

$$\begin{bmatrix} mv(\dot{\beta}+r) \\ ml_f l_r \dot{r} \end{bmatrix} = \begin{bmatrix} F_f + F_r \\ F_f l_f - F_r l_r \end{bmatrix} \tag{1}$$

The tire force can be expressed as

$$F_f(\alpha_f) = \mu c_{f0}\alpha_f, \quad F_r(\alpha_r) = \mu c_{r0}\alpha_r \tag{2}$$

with the tire cornering stiffnesses c_{f0}, c_{r0}, the road adhesion factor μ and the tire side slip angles

$$\alpha_f = \delta_f - (\beta + \frac{l_f}{v} r), \alpha_r = -(\beta - \frac{l_r}{v} r) \tag{3}$$

The state equation of vehicle dynamics with β and r can be represented as

$$\begin{bmatrix} \dot{\beta} \\ \dot{r} \end{bmatrix} = \begin{bmatrix} -\frac{\mu(c_{f0}+c_{r0})}{mv} & -1+\frac{\mu(c_{r0}l_r - c_{f0}l_f)}{mv^2} \\ \frac{\mu(c_{r0}l_r - c_{f0}l_f)}{ml_f l_r} & -\mu\frac{(c_{f0}l_f^2 + c_{r0}l_r^2)}{ml_f l_r v} \end{bmatrix} \begin{bmatrix} \beta \\ r \end{bmatrix} + \begin{bmatrix} \frac{\mu c_{f0}}{mv} \\ \frac{\mu c_{f0}}{ml_r} \end{bmatrix} \delta_f \tag{4}$$

Hence, the transfer function from δ_f to r is

$$G_{r/\delta_f} = \frac{c_{f0} m l_f \mu v^2 s + c_{f0} c_{r0} l \mu^2 v}{l_f l_r m^2 v^2 s^2 + l(c_{r0} l_r + c_{f0} l_f) m \mu v s + c_{f0} c_{r0} l^2 \mu^2 + (c_{r0} l_r - c_{f0} l_f) m \mu v^2} \tag{5}$$

The numerical data in this paper are listed in Table 2.

According to the above analysis of a single track vehicle model, the transfer function from the input of front deflection angle δ_f to the output of yaw rate r can be obtained as

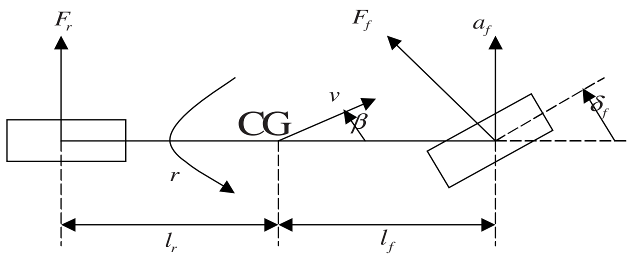


Fig. 1. Single track vehicle model

$$G_{r/\delta_f}(s,\mu,v) = \frac{(1.382\times10^8 \mu v^2 s + 1.415\times10^{10} \mu^2 v)}{6.675\times10^6 v^2 s^2 + 1.08\times10^9 \mu v s + (1.034\times10^7 \mu v^2 + 4\times10^{10} \mu^2)} \tag{6}$$

The operating area Q of the uncertain parameters μ and v is depicted in Fig. 7.

In addition, the steering actuator is modeled as

$$G_A(s) = \frac{\omega_n^2}{s^2 + \sqrt{2}\omega_n s + \omega_n^2} \tag{7}$$

where $\omega_n = 4\pi$.

The open loop transfer function $G_O(s)$ is defined as

$$G_O(s,\mu,v) = G_A(s) G_{r/\delta_f}(s,\mu,v) \tag{8}$$

2.2 Describing Function of Fuzzy Controller

In this subsection, the fuzzy controller given in [11] is adopted here. Figure 2 shows the block diagram for determining the describing function of the fuzzy controller from experimental evaluations. The membership functions of the fuzzy controller are shown in Fig. 3 (a)-(c) and the rules of the fuzzy controller are listed in Table 3.

Figure 4 shows the control surface. According to the analysis in [11], the describing function $N(A)$ with input signal ($x(t) = A\sin\omega t$) and scaling factors ($k_p = 6,\ k_d = 0.001$) can be obtained in Fig. 5.

Table 1. Vehicle system quantities

F_f, F_r	lateral wheel force at front and rear wheel
r	yaw rate
β	side slip angle at center of gravity (CG)
v	velocity
a_f	lateral acceleration
l_f, l_r	distance from front and rear axis to CG
$l = l_f + l_r$	wheelbase
δ_f	front wheel steering angle
m	mass

Table 2. Vehicle parameters

c_{f0}	50000 N/rad
c_{r0}	100000 N/rad
m	1830 kg
l_f	1.51 m
l_r	1.32 m

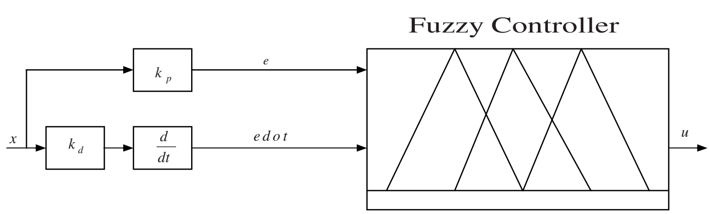
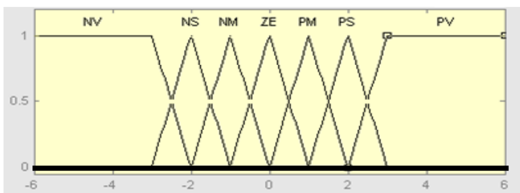


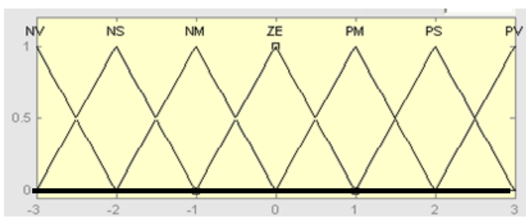
Fig. 2. Fuzzy Controller

Table 3. Rules of the fuzzy controller

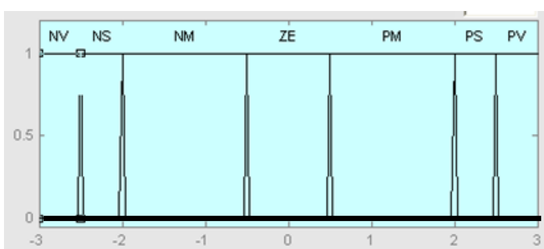
edot \ e	NV	NS	NM	ZE	PM	PS	PV
NV	NV	NV	NV	NV	NS	NM	ZE
NS	NV	NV	NV	NS	NM	ZE	PM
NM	NV	NV	NS	NM	ZE	PM	PS
ZE	NV	NS	NM	ZE	PM	PS	PV
PM	NS	NM	ZE	PM	PS	PV	PV
PS	NM	ZE	PM	PS	PV	PV	PV
PV	ZE	PM	PS	PV	PV	PV	PV



(a) Input of e



b) Input of edot



(c) Output of u

Fig. 3. Fuzzy membership functions

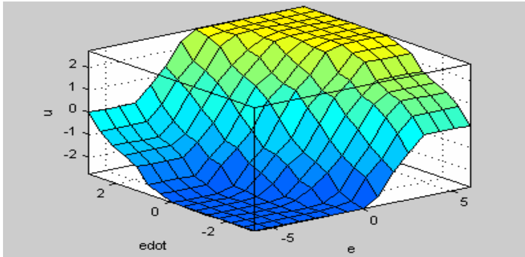


Fig. 4. Control Surface

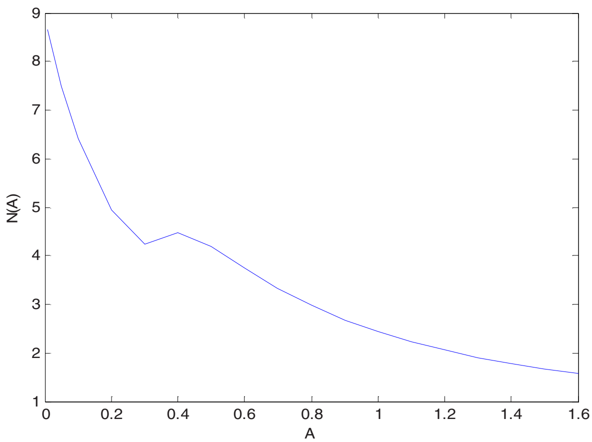


Fig. 5. Describing function of the fuzzy controller

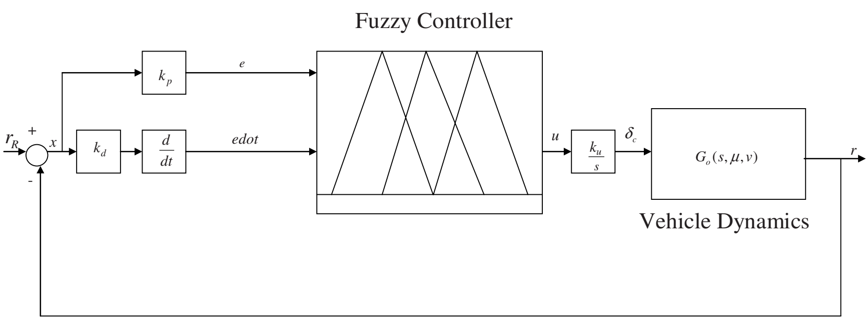


Fig. 6. Block diagram of a fuzzy vehicle lateral control system

2.3 Parameter Plane Method

Figure 6 shows the overall block diagram of a fuzzy vehicle lateral control system. In this study, $k_u = 0.1$ is selected.

After some simple manipulations by using (5) to (8), the characteristic equation in Fig. 6 can be obtained as

$$\begin{aligned} & f(s,\mu,v) \\ & = C_4\mu^2 + C_3v^2 + C_2\mu v + C_1\mu^2 v + C_0\mu v^2 \\ & = 0 \end{aligned} \tag{9}$$

where

$$C_4 = 4.0045\times10^{10}s(s^2+17.7715s+157.9137)\,,$$
$$C_3 = 6.675\times10^{6}s^3(s^2+17.7715s+157.9137)\,,$$
$$C_2 = 1.0746\times10^{9}s^2(s^2+17.7715s+157.9137)\,,$$
$$C_1 = 2.2345\times10^{11}N_1\,,$$
$$C_0 = 1.03395\times10^{8}s(s^2+17.7715s+157.9137)+2.1818\times10^{9}N_1 s\,.$$

Let $s = j\omega$, (9) is divided into two stability equations with real part X and imaginary part Y of characteristic equation, one has

$$f(j\omega,\mu,v) = X + jY = 0 \tag{10}$$

where

$$X = -7.1165\times10^{11}\omega^2\mu^2 + 1.0075\times10^{12}\omega^4v^2 + (1.0746\times10^{9}\omega^4$$
$$-1.6970\times10^{11}\omega^2)\mu v + 2.2345\times10^{11}N_1\mu^2 v - 1.8375\times10^{9}N_1\omega^2\mu v^2,$$
$$Y = (6.3236\times10^{12}\omega - 4.0045\times10^{10}\omega^3)\mu^2 + (6.6750\times10^{6}\omega^5$$
$$-1.0541\times10^{9}\omega^3)v^2 - 1.9098\times10^{10}\omega^3\mu v + (1.6327\times10^{10}\omega - 1.03395\times10^{8}\omega^3$$
$$+2.1818\times10^{9}N_1\omega)\mu v^2.$$

In order to obtain the solution of μ and v, the following equation is solved

$$\begin{cases} X = 0 \\ Y = 0 \end{cases} \tag{11}$$

when k_p, k_d, k_u, N_1 are fixed and ω is changed from 0 to ∞. As the amplitude A is changed, the solutions of μ and v called limit cycle loci can be displayed in the parameter plane.

3 Simulation Results

In this work, the scaling factors of $k_p = 6$, $k_d = 0.001$, and $k_u = 0.1$ in Fig. 6 are selected. The membership functions and rules of fuzzy controller are the same with the above section. So the describing function of the fuzzy controller shown in Fig. 5 can be applied to analyze the stability of the closed loop system. After doing so, the solution of (11) can be solved when A is fixed and ω is changed from 0 to ∞. Figure 7

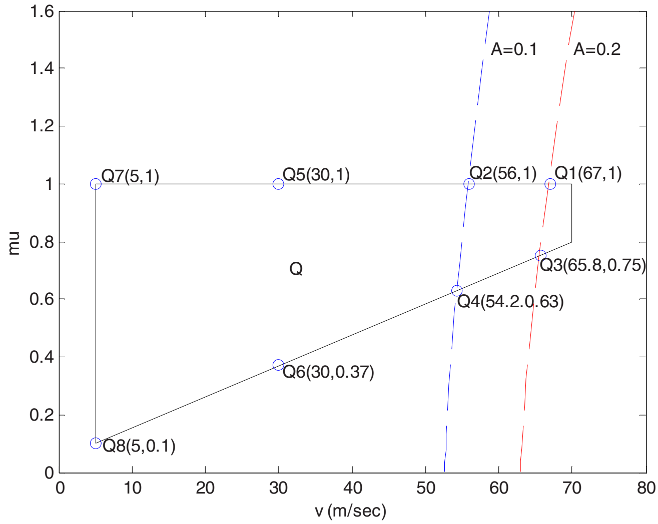


Fig. 7. Operating area and limit cycle loci

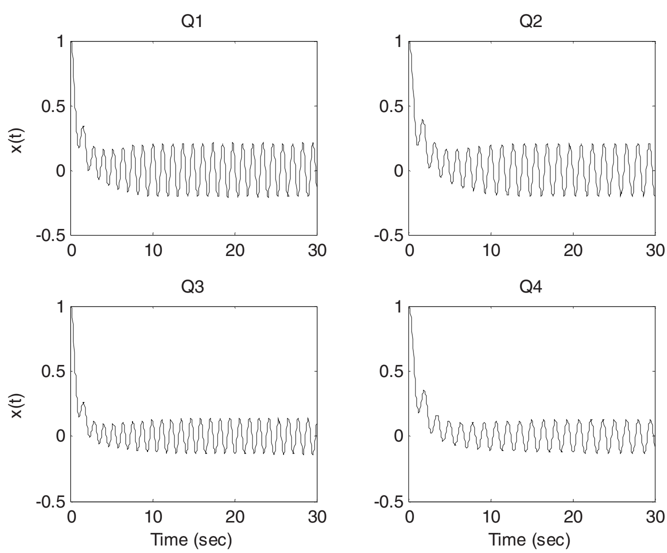


Fig. 8. Input signal of Q1 to Q4

shows the operating area of the vehicle system. Two limit cycle loci in the μ - v parameter plane as $A = 0.1$ and $A = 0.2$ are plotted in Fig. 7, respectively. In order to test the predicted result in Fig. 7, eight operating points with Q1 ($\mu = 1$, $v = 67$), Q2($\mu = 1$, $v = 56$), Q3($\mu = 0.75$, $v = 65.8$), Q4($\mu = 0.63$, $v = 54.2$), Q5($\mu = 1$, $v = 30$), Q6($\mu = 0.37$, $v = 30$), Q7($\mu = 1$, $v = 5$), Q8($\mu = 0.1$, $v = 5$) are

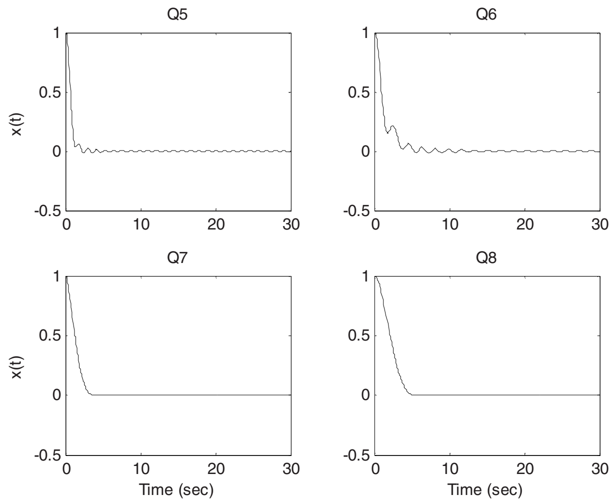


Fig. 9. Input signal of Q5 to Q8

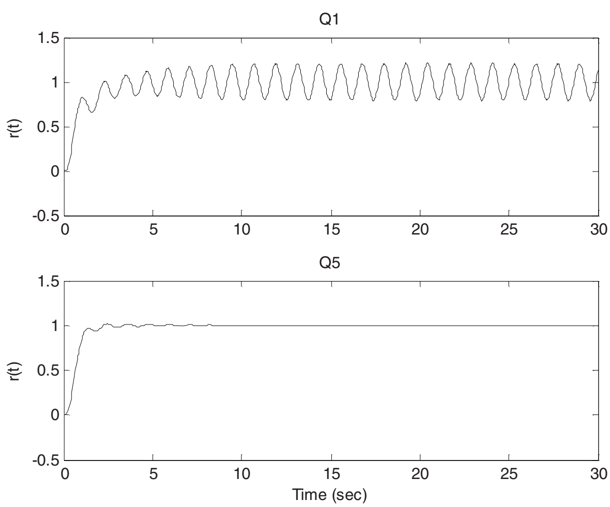


Fig. 10. Output signal of Q1 and Q5

chosen for the time simulations. Figure 8 shows the time responses of the input signal $x(t)$ at Q1 to Q4. The limit cycles are existed when operating at these four points and it consists with the predicted results (the amplitude of limit cycles) in Fig. 7.

4 Conclusions

Based on the approaches of describing function and parameter plane, the robust stability of a fuzzy vehicle lateral system is considered in this paper. A systematic procedure is presented to deal with this problem. Simulation results show that more information about the characteristic of limit cycle can be obtained by this approach.

On the other hands, the time responses of the input signal $x(t)$ at the other four points are displayed in Fig. 9. No limit cycle occurs in these four cases and this result is also matched with Fig. 7. Finally, the unit step responses of the output signal $r(t)$ at two operating points Q1 and Q5 are given in Fig. 10.

Acknowledgments. This paper is supported by the National Science Council in Taiwan through Grant NSC95-2752-E009-010.

References

1. Gelb A, Velde W. E. V.: Multiple Input Describing Functions and Nonlinear System Design. McGraw-Hill, New York (1968)
2. Siljak, D. D.: Nonlinear Systems - The Parameter Analysis and Design. New York: Wiley (1969).
3. Atherton, D. P.: Nonlinear Control Engineering. Van Nostrand Reinhold Company, London (1975)
4. Arruda, G. H. M., Barros, P. R.: Relay-Based Gain and Phase Margins PI Controller Design. IEEE Trans. Instru, and Measure. Vol 52 5 (2003) 1548-1553
5. Krenz, G. S., Miller, R. K.: Qualitative Analysis of Oscillations in Nonlinear Control Systems: A Describing Function Approach. IEEE Trans. Circuits and Systems. Vol. CAS-33 5 (1986) 562-566
6. Ginart, A. E., Bass, R. M., Leach, W. M. J., Habetler, T. G.: Analysis of the Class AD Audio Amplifier Including Hysteresis Effects. IEEE Trans. Power Electronics. Vol. 18 2 (2003) 679-685
7. Ackermann, J., Bunte, T.: Actuator Rate Limits in Robust Car Steering Control. Prof. IEEE Conf. Decision. and Control. (1997) 4726-4731
8. Nataraj, P. S. V., Barve, J. J.: Limit Cycle Computation for Describing Function Approximable Nonlinear Systems with Box-Constrained Parametric Uncertainties. Int. J. of Robust and Nonlinear Cont. Vol. 15 (2005) 437-457
9. Gordillo, F., Aracil, J., Alamo, T.: Determining Limit Cycles in Fuzzy Control Systems. Proc. IEEE Int. Conf. Fuzzy Syst. Vol. 1 (1997) 193-198
10. Kim, E., Lee, H., Park, M.: Limit-Cycle Prediction of a Fuzzy Control System Based on Describing Function Method. IEEE Trans. Fuzzy Syst. Vol. 8 1 (2000) 11-21
11. Sijak, T., Kuljaca, O., Tesnjak, S.: Stability Analysis of Fuzzy Control System Using Describing Function Method. MED2001. Dubrovnik, Croatia (2001)
12. Delgado, A.: Stability Analysis of Neurocontrol Systems Using a Describing Function. Proc. Int. Join. Conf. Neural Network. Vol. 3 (1998) 2126-2130
13. Siljak, D. D.: Parameter Space Methods for Robust Control Design: A Guide Tour. IEEE Trans. Automat. Contr. Vol. 34 7 (1989) 674-688
14. Han, K. W.: Nonlinear Control Systems - Some Practical Methods. Academic Cultural Company, California (1977)

An Adaptive Location Service on the Basis of Fuzzy Logic for MANETs

Ihn-Han Bae and Yoon-Jeong Kim

School of Computer and Information Communication Eng.,
Catholic University of Daegu, GyeongSan 712-702, Korea
ihbae@cu.ac.kr, kimyj2845@naver.com

Abstract. Location services are used in mobile ad hoc and hybrid networks either to locate the geographic position of a given node in the network or for locating a data item. One of the main usages of position location services is in location based routing algorithms. In particular, geographic routing protocols can route messages more efficiently to their destinations based on the destination node's geographic position, which is provided by a location service. In this paper, we propose an adaptive location service on the basis of fuzzy logic called FHLS (*Fuzzy Hierarchical Location Service*) for mobile ad hoc networks. The FHLS uses the adaptive location update scheme using the fuzzy logic on the basis of the mobility and the call preference of mobile nodes. The performance of the FHLS is to be evaluated by using a simulation, and compared with that of existing HLS scheme.

Keywords: Mobile Ad-hoc Network, Location Service, Fuzzy Logic.

1 Introduction

Mobile ad hoc networks (MANETs) enable wireless communication between mobile hosts without relying on a fixed infrastructure. In these networks the mobile hosts themselves forward data from a sender to a receiver, acting both as router and end-system at the same time. MANETs have a wide range of applications, e,g., range extension of WLAN access points, data transmission in disaster areas and inter-vehicular communication. Due to scare bandwidth, varying network connectivity and frequent topology changes caused by node mobility and transient availability, routing algorithms tailored for wired networks will not operate well if directly transposed to MANETs.

Since no fixed infrastructure of servers is assumed in MANETs, it is useful to devise a scheme through which various services offered within the network may be located. With the availability of such location services, it is tempting to adapt and exploit them for storing routing information. By storing the geographic location of mobile nodes in designed location servers in the network, it is possible to introduce a new family of routing algorithms that may potentially perform better than the traditional approach of discovering and maintaining end-to-end routes.

In this paper, we present an adaptive location service on the basis of fuzzy logic called FHLS (*Fuzzy Hierarchical Location Service*) for MANETs. FHLS divides the

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 683–690, 2007.
springerlink.com

area covered by the network into a hierarchical of regions. The top-level region covers the complete network. A region is subdivided into several regions of the next lower level until the lowest level is reached. We call a lowest level region a cell. Using the hierarchy as a basis, the FHLS uses the adaptive location update scheme using the fuzzy logic on the basis of the mobility and the call preference of mobile nodes. The performance of the FHLS is to be evaluated by a simulation, and compared with that of existing HLS scheme.

The remainder of this paper is organized as follows. The next section provides an overview of related work. In Section 3, we present the FHLS algorithm, which serves as a basis for a routing algorithm. In Section 4, we undertake a simulation study for the FHLS. We finally conclude and describe future work in Section 5.

2 Related Work

In routing protocol of MANETs, the location service uses the location information of a node for packet routing. So, many researches for location management in MANETs were performed recently. A location service that uses flooding to spread position information is DREAM (*Distance Routing Effect Algorithm for Mobility*) [1]. With DREAM, each node floods its position information in the network with varying flooding range and frequency. The frequency of the flooding is decreased with increasing range. Thus, each node knows the location of each other node, whereas the accuracy of this information depends on the distance to the node.

The GLS (*Grid Location Service*) [2] divides the area containing the ad hoc network into a hierarchy of square forming a quad-tree. Each node selects one node in each element of the quad-tree as a location server. Therefore the density of location servers for a node is high in areas close to the node and becomes exponentially less dense as the distance to the node increases. The update and request mechanisms of GLS require that a chain of nodes based on node IDs is found and traversed to reach an actual location server for a given node. The chain leads from the updating or requesting node via some arbitrary and some dedicated nodes to a location server.

DLM (*Distributed Location Management Scheme*) [3] partitions the mobile node deployment region into a hierarchical grid with square of increasing size. The location service is offered by location servers assigned across the grid, storing node location information. Nodes that happen to be located in these regions offer the location service. The selection mechanism for the predetermined regions is carried out through a hash function, which maps node identifiers to region addresses. DLM distinguishes between a full length address policy and a partial length address policy. Under the full length address policy, location servers store the precise position of nodes. When nodes change regions due to mobility, it is necessary to update all location servers. Under the partial length address policy, the accuracy of node location information stored at the location servers increases along with the proximity of the location servers to the nodes. To the contrary of the full length address policy, several queries are necessary to locate a node. Nevertheless, the partial addressing scheme offers overall increasing performance, since it reduces the scope and frequency of location server updates due to node mobility.

HLS (*Hierarchical Location Service*) [4] divides the area covered by the network into a hierarchy of regions. The lowest level regions are called cells. Regions of one level are aggregated to form a region on the next higher level of the hierarchy. Regions on the same level of the hierarchy do not overlap. For any given node A, one cell in each level of the hierarchy is selected by means of a hash function. These cells are called the responsible cells for node A. As a node moves through the area covered by the network, it updates its responsible cells with information about its current position. When another node B needs to determine those cells that may potentially be responsible for A, it then queries those cells in the order of the hierarchy, starting with the lowest level region. There are two different methods for HLS to update location servers, the direct location scheme and the indirect location scheme. To update its location servers according to the direct location scheme, a node computes its responsible cells. Position updates are then sent to all responsible cells. The location information in the responsible cells is represented as a pointer to the position of the node. The network load can be reduced with the indirect location scheme where the location servers on higher hierarchy levels only know the region of the next lower level a node is located in. More precise location information is not necessary on higher levels.

3 Fuzzy Hierarchical Location Service

We present an adaptive location service on the basis of fuzzy logic called FHLS (*Fuzzy Hierarchical Location Service*) to minimize the sum of location update cost and paging cost. FHLS is adapted to the location update rate and call arrival rate in an environment where the characteristic of nodes movement changes all the time. The FHLS uses the fuzzy control logic for location update. The input parameters of the fuzzy logic are the linguistic variables of location update rate and call arrival rate for a mobile node, and the output is a direct location update scheme or an indirect location update scheme on the basis of a different level in hierarchical region.

3.1 Area Partitioning

FHLS partitions the area containing the ad-hoc network in cells. This partitioning must be known to all participating nodes. The shape and size of the cells can be chosen arbitrary according to the properties of the network. The only prerequisite is that a node in a given lowest-level cell must be able to send packets to all other nodes in the same cell.

The cells are grouped hierarchically into regions of different levels. A number of cells form a region of level one, a number of level-one regions forms a level-two region and so on. Regions of the same level must not intersect, i.e., each region of level n is a member of exactly one region of level n+1. An example for the area partitioning is shown in Fig. 1.

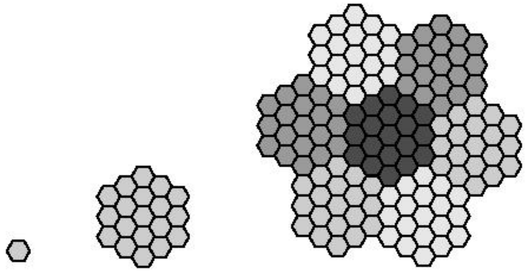


Fig. 1. Grouping cells to form regions: cell, region on level one and region on level two

3.2 Location Service Cells

FHLS places location information for a node T in a set of cells. We call these cells the LSC (location service cells) of T. A node T selects one LRC for each level in the hierarchy. For a given level n, the LRC is selected according to the following algorithm:

1. Compute the set *S*(*T, n*) of cells belonging to the region of level n which contains T.
2. Select one of these cells with a hash function based on the characteristics of S and the node ID of T.

A possible hash function is the simple, modulo-based function: $H(T,n) = ID(T) \bmod \|S(T,n)\|$. With the number of calculated with this hash function, one of the cells in *S*(*T, n*) is chosen. As a result of the above selection, T has exactly one location service cell on each level n and it is guaranteed that the LRC for T of level n and node T share the same level-n region. An example for the selection of LSCs for a three-level hierarchy is shown in Fig. 2. All location service cells are candidates for location service. These candidate cells, tree-like structure are called candidate tree.

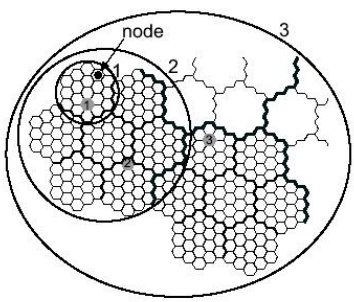


Fig. 2. Example for location service cells of a node

3.3 Location Updates

We propose the adaptive location update scheme using the fuzzy logic on the basis of the mobility and the call preference for mobile nodes. The input parameters of the fuzzy logic are the linguistic variables of location update rate and call arrival for a mobile node, and the output is a direct location update scheme or an indirect location update scheme on the basis of a different level in the hierarchy.

The mobility of a mobile node is measured by location update rate per minute. We map the update rate to the linguistic variable for the node mobility using the membership function as shown in Fig. 3.

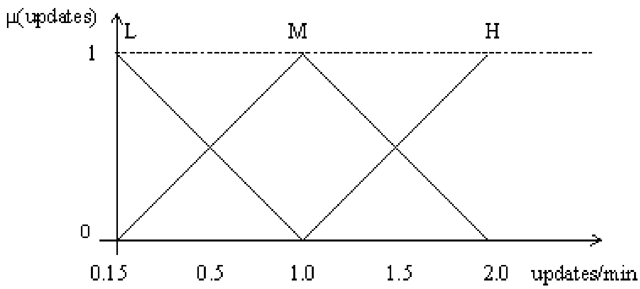


Fig. 3. The membership function and the linguistic variables for node mobility

The call preference of a mobile node is measured by call arrival rate per minute. We map the call arrival rate to the linguistic variable for the node preference using the membership function as shown in Fig. 4.

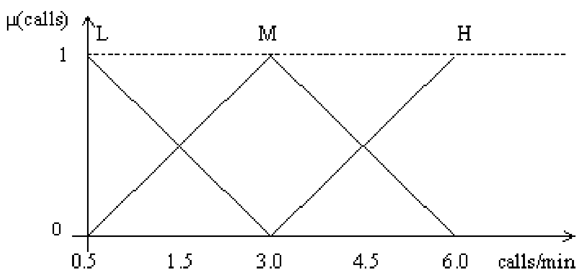


Fig. 4. The membership function and the linguistic variables for node preference

Fuzzy logic uses linguistic variable to map the input fuzzy variable to the output variable. This is accomplished by using IF-THEN rules [5, 6]. We use 9 fuzzy rules described in Table 1. According to the mobility and the preference for a mobile node, a direct location update scheme or an indirect location update scheme on the basis of a different level in the hierarchy is used for updating the location information of the mobile node. For a mobile node, if the linguistic variable for node mobility is M and the linguistic variable for node preference is M, FHLS uses the level-i indirect location update scheme for updating the location information of the mobile node, where the input parameter for FHLS, the level-i represents the level of the first location service cell in the hierarchy.

Table 1. Fuzzy control rules for location update

Mobility / Preference	L	M	H
L	level-(i+1) direct location update scheme	level-(i+1) indirect location update scheme	level-(i+1) indirect location update scheme
M	level-i direct location update scheme	level-i indirect location update scheme	level-i indirect location update scheme
H	level-(i-1) direct location update scheme	level-(i-1) direct location update scheme	level-(i-1) indirect location update scheme
(Input variables) L – Low, M – Medium, H – High (Output variables) a direct location update scheme or an indirect location update scheme on the basis of a different level in the hierarchy			

3.4 Location Requests

To successfully query the current location of a target node T, the request of a source node S needs to be routed to a location server of T. When querying the position of T, S knows the ID of T and therefore the structure of the candidate tree defined via the hash function and T's ID. It furthermore knows that T has selected a LSC for each region it resides in. Thus, the request only needs to visit each candidate cell of the regions containing S.

S computes the cell that T would choose as a location service cell. If the location service cell were in the same level-one region, the S sends its request to this cell. When the request packet arrives at the first node A within the boundaries of the candidate cell, it is processed as follows:

1. Node A broadcasts the request to all nodes within the candidate cell. This is called cellcast request.
2. Any node that receives this cellcast request and has location information in its location database sends an answer to A.
3. If A receives an answer for its cellcast request, the request is forwarded to the target node T.
4. Otherwise A forwards the request to the corresponding candidate cell the next level.

With this mechanism, the request is forwarded from candidate cell to candidate cell until a location server for T is founded or the highest level candidate cell has been reached.

4 Performance Evaluation

We evaluate the performance of the FHLS in terms of the total cost that consists of location update and paging costs. Table 2 shows the parameter values for the simulation.

Table 2. Simulation parameters

Parameters	Values
The number of cells	64
Cell size (kilometer×kilometer)	2×2
Update rate per minute	random (0.15~2.0)
Call rate per minute	random (0.5~6.0)
Simulation time (minute)	100

The simulation result of total cost over time is illustrated in Fig. 5. As shown in the Fig. 5, the performance of the proposed FHLS is better than that of existing HLS. This is because the FHLS uses an adaptive location update scheme according to the characteristic of mobile node that is location update rate and call arrival rate. Also, the performance of the FHLS (level=4) that places the first location service cell on level 4 in the hierarchy is superior to that of the FHLS (level=3) that places the first location service cell on level 3 in the hierarchy. This is because as the level of the first location service cell increasing, the number of location service cells is decreased and the paging cost is reduced by the spatial locality in location query.

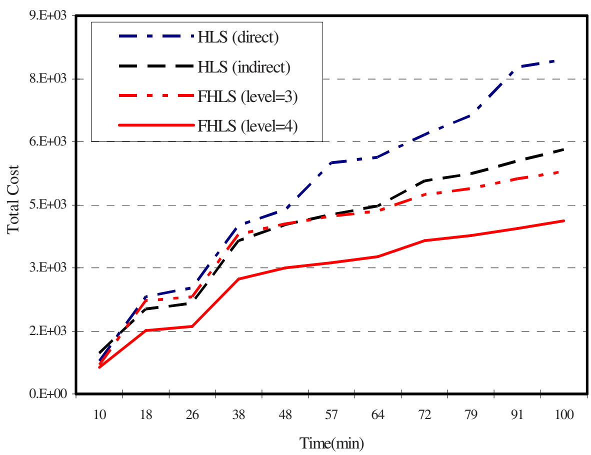


Fig. 5. Total cost for location service

Additionally, we know that the performance of the HLS (indirect) that uses indirect location update scheme is better than that of HLS (direct) that uses direct location update scheme.

5 Conclusions

In this paper, we have presented the Fuzzy Hierarchical Location Service (FHLS) for MANETs. FHLS uses the adaptive location update scheme using the fuzzy logic on the basis of the mobility and the call preference for mobile nodes. The input parameters of the fuzzy logic are the linguistic variables of location update rate and call arrival rate for a mobile node, and the output is a direct location update scheme or an indirect location update scheme on the basis of a different level in the hierarchy.

The performance of the FHLS has been evaluated by using a simulation. Because the FHLS uses an adaptive location update scheme according to the characteristic of mobile node that is location update rate and call arrival rate, the performance of the proposed FHLS is better than that of existing HLS. Also, we know that the performance of the FHLS (level=4) that places the first location service cell on level 4 in the hierarchy is superior to that of the FHLS (level=3) that places the first location service cell on level 3 in the hierarchy.

As part of our future work, we plan to evaluate our proposed FHLS in terms of various factors i.e., the number of cells, cell sizes, etc.

References

1. Basagni, S., Chlamtac, I., Syrotiuk, V. R., and Woodward, B. A.: A Distance Routing Effect Algorithm for Mobility (DREAM). Proceeding of the fourth annual ACM/IEEE International Conference on Mobile computing and networking (1998) 76-84
2. Li, J., Jannotti, J., DeCouti, D. S. J., Karger D. R. and Morris, R.: A Scalable Location Service for Geographic Ad Hoc Routing. Proceeding of the sixth annual ACM/IEEE International Conference on Mobile computing and networking (2000) 120-130
3. Xue, Y., Li. B. and Nahrstedi. K.: A Scalable Location Management Scheme in Mobile Adhoc Networks. Proceeding of the IEEE Conference on Local Computer Networks (2002) 102-111
4. Kieb W., Fubler H, Widmer and Mauve M.: Hierarchical Location Services for Mobile Ad-Hoc Networks. Mobile Computing and Communication Review 8(4) (2004) 47-58
5. Baldwin, J. F.: Fuzzy logic and fuzzy reasoning. In Fuzzy Reasoning and It's Applications, E. H. Mamdani and B. R. Gaines (eds), Academic Press, London (1981)
6. Bae, I-H, Oh, S-J and Olariu, S.: A Fuzzy Logic-Based Location Management Method for Mobile Networks. software Engineering Research and Application, Springer-Verlag Berlin Heidelberg (2004) 304-319
7. Karp, B. N. and Kung, H. T.: GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. Proceeding of the sixth annual ACM/IEEE International Conference on Mobile computing and networking (2000) 243-254

Self-tunable Fuzzy Inference System: A Comparative Study for a Drone

Hichem Maaref, Kadda Meguenni Zemalache, and Lotfi Beji

IBISC Laboratory, CNRS-FRE 2873,Université d'Evry, 40 rue du Pelvoux 91020
Evry Cedex, France
hichem.maaref,kadda.zemalache,lotfi.beji@ibisc.univ-evry.fr

Abstract. The work describes an automatically on-line Self-Tunable Fuzzy Inference System (STFIS) of a mini-flying called XSF drone. A Fuzzy controller based on an on-line optimization of a zero order Takagi-Sugeno fuzzy inference system (FIS) by a back propagation-like algorithm is successfully applied. It is used to minimize a cost function that is made up of a quadratic error term and a weight decay term that prevents an excessive growth of parameters. Simulation results and a comparison with a Static Feedback Linearization controller (SFL) are presented and discussed. A path-like flying road, described as straight-lines with rounded corners permits to prove the effectiveness of the proposed control law.

Keywords: Fuzzy Inference System, Optimization, Static Feedback Linearization controller, Tracking control, Dynamic systems, Drone.

1 Introduction

Modelling and controlling aerial vehicles (blimps, mini rotorcraft, drone) are one of the principal preoccupation of the IBISC-group. Within this optic, the attracted contest of the DGA-ONERA is the XSF project which consists of a drone with revolving aerofoils [2], (Fig. 1 (top)). It is equipped with four rotors where two are directionals, what we call in the following X4 Bidirectional Rotors or X4 Stationary Flyer (XSF). The XSF is an engine of $68cm \times 68cm$ of total size and not exceed $2Kg$ in mass. It is designed in a cross form and made of carbon fibber. Each tip of the cross has a rotor including an electric brushless motor, a speed controller and a two-blade ducted propeller. The operating principle of the XSF can be presented thus: two rotors turn clockwise, and the two other rotors turn counter clockwise to maintain the total equilibrium in yaw motion. A characteristic of the XSF compared to the existing quad-rotors, is the swivelling of the supports of the motors 1 and 3 around the pitching axis thanks to two small servomotors controlling the swivelling angles ξ_1 and ξ_3. This swivelling ensures either the horizontal rectilinear motion or the rotational movement around the yaw axis or a combination of these two movements which gives the turn (Fig. 2 (bottom)). This permits a more stabilized horizontal flight and a suitable cornering.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 691–700, 2007.
springerlink.com

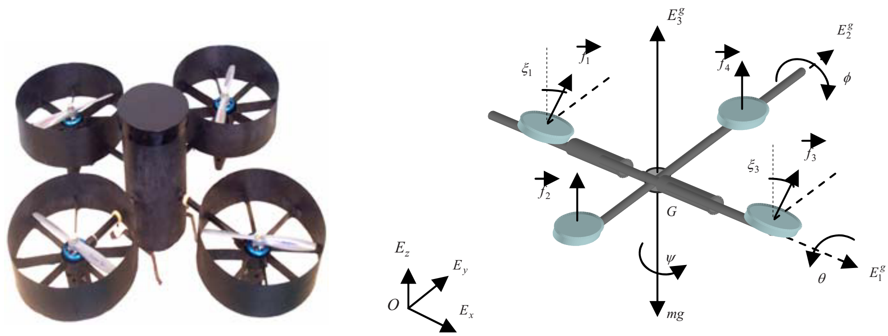


Fig. 1. Conceptual form of the XSF drone (top). Frames attached to the XSF (bottom).

Several recent works were completed for the design and control in pilotless aerial vehicles domain such that Quadrotor [1, 8], X4-flyer [3, 5]. A model for the dynamic and configuration stabilization of quasi-stationary flight conditions of a four rotors VTOL was studied by Hamel et al [5] where the dynamic motor effects are incorporating and a bound of perturbing errors is obtained for the coupled system.

All the reviewed techniques require the well knowledge of the system model and parameters. In this paper, a STFIS control strategy is presented based on the systems output measures. This technique early used for autonomous wheeled robot, is adapted for the use with the XSF [10, 11]. In autonomous wheeled robot, many developed learning techniques have arisen in order to generate or to tune fuzzy rules. Most of them are based on the so-called "Neuro-Fuzzy learning algorithms" as proposed by [6, 7].

In this paper, the Self-Tunable Fuzzy Inference System (STFIS) controller is presented and compared with a Static Feedback Linearization controller (SFL) to stabilize the XSF by using the point to point steering stabilization. The dynamic model of the XSF drone is briefly given in the second section. The developed ideas of control for the XSF by the STFIS and the SFL controllers are described in the third section. Motion planning and simulation results are introduced in the fourth section. The robustness of the proposed controller is then evaluated in the fifth section. Finally, conclusion and future works are given in the last section.

2 Motion Dynamic

We consider the translation motion of $\Re_G$ with respect to $\Re_O$. The position of the mass centre wrt $\Re_O$ is defined by $\overline{OG} = (x, y, z)^T$, its time derivative gives the velocity wrt to $\Re_O$ such that $d\overline{OG}/dt = (\dot{x}, \dot{y}, \dot{z})^T$, while the second time derivative permits to get the acceleration $d^2\overline{OG}/dt^2 = (\ddot{x}, \ddot{y}, \ddot{z})^T$ (Fig. 1 (bottom)).

The model is a simplified one's. The constraints as external perturbation and the gyroscopic torques are neglected. The aim is to control the engine vertically along z

axis and horizontally according to x and y axis. The vehicle dynamics, represented on Figure 1 (bottom), is modelled by the system of equations $Eq.\,1$, [2, 3, 9].

$$\begin{aligned}
m\ddot{x} &= S_{\psi}C_{\theta}u_2 - S_{\theta}u_3 \\
m\ddot{y} &= \left(S_{\theta}S_{\psi}S_{\phi}\right)u_2 + \left(C_{\theta}S\phi\right)u_3 \\
m\ddot{z} &= \left(S_{\theta}S_{\psi}C_{\phi}\right)u_2 + \left(C_{\theta}C\phi\right)u_3 \\
\ddot{\theta} &= \tilde{\tau}_{\theta}; \quad \ddot{\phi} = \tilde{\tau}_{\phi}; \quad \ddot{\psi} = \tilde{\tau}_{\psi}
\end{aligned} \tag{1}$$

Where C_{θ} and S_{θ} represent $\cos\theta$ and $\sin\theta$, respectively. m is the total mass of the XSF. The vector u_2 and u_3 combines the principal non conservative forces applied to the engine airframe including forces generated by the motors and drag terms. Drag forces and gyroscopic due to motors effects are not considered in this work.

3 Advanced Strategies of Control

The aim is to make comparison between model based approaches and experts analysis involving fuzzy systems. Classical model based techniques such as the static feedback linearization and backstepping techniques have been investigated and used for stabilization with motion planning [3, 9].

3.1 Static Feedback Linearization Controller

Control input for z-y motions: We propose to control the y/z motion through the input u_2/u_3. So, we have the following proposition.

Proposition 1. Consider $(\psi,\theta) \in \left]-\pi/2, \pi/2\right[$, with the static feedback laws.

$$\begin{aligned}
u_2 &= m\nu_y C_{\phi}C_{\psi}^{-1} - m(\nu_z + g)S_{\phi}C_{\psi}^{-1} \\
u_3 &= m\nu_y\left(S_{\phi}C_{\theta}^{-1} - C_{\phi}tg_{\psi}tg_{\theta}\right) + m(\nu_z + g)\left(C_{\phi}C_{\theta}^{-1} - S_{\phi}tg_{\psi}tg_{\theta}\right)
\end{aligned} \tag{2}$$

The dynamic of y and z are linearly decoupled and exponentially-asymptotically stable with the appropriate choice of the gain controller parameters. ν_y and ν_z are detailed in the following.

We can regroup the two dynamics as:

$$\begin{pmatrix} \ddot{y} \\ \ddot{z} \end{pmatrix} = \frac{1}{m}\begin{pmatrix} S_{\theta}S_{\psi}S_{\phi} + C_{\psi}C_{\theta} & C_{\theta}S_{\phi} \\ S_{\theta}S_{\psi}C_{\phi} - C_{\psi}S_{\theta} & C_{\theta}C_{\phi} \end{pmatrix}\begin{pmatrix} u_2 \\ u_3 \end{pmatrix} - \begin{pmatrix} 0 \\ g \end{pmatrix} \tag{3}$$

For the given conditions in ψ and θ the (2×2) matrix in $Eq.\,3$ is invertible. Then a nonlinear decoupling feedback permits to write the following decoupled linear dynamics.

$$\begin{aligned}
\ddot{y} &= \nu_y \\
\ddot{z} &= \nu_z
\end{aligned} \tag{4}$$

Then we can deduce from $Eq.\ 4$ the linear controller:

$$\begin{aligned} \nu_y &= \ddot{y}_r - k_y^1(\dot{y}-\dot{y}_r) - k_y^2(y-y_r) \\ \nu_z &= \ddot{z}_r - k_z^1(\dot{z}-\dot{z}_r) - k_z^2(z-z_r) \end{aligned} \quad (5)$$

with the k_y^i, k_z^i are the coefficients of a Hurwitz polynomial.

Our second interest is the (x, z) dynamics which can be also decoupled by a static feedback law.

Input u_2 for the motion along x: The aim here is to stabilize the dynamics of x with the control input u_2. While we keep the input u_3 for the altitude z stabilization.

Proposition 2. with the new inputs

$$\begin{aligned} u_2 &= (S_\psi C_\phi - S_\theta C_\psi S_\phi)^{-1}(m\nu_x C_\phi C_\theta + m(\nu_z + g)S_\theta) \\ u_3 &= (S_\psi C_\phi - S_\theta C_\psi S_\phi)^{-1}(-m\nu_x(S_\psi S_\theta C_\phi - C_\psi C_\phi) + m(\nu_z + g)S_\psi C_\theta) \end{aligned} \quad (6)$$

The dynamic of x and z are decoupled and exponentially stable. ν_x and ν_z can be deduced as in $Eq.\ 7$ [3].

$$\begin{aligned} \ddot{x} &= \nu_x \\ \ddot{z} &= \nu_z \end{aligned} \quad (7)$$

Remark. To switch between the two controllers $Eq.\ 2$ and $Eq.\ 6$ continuity is recommended allowing the avoidance of the peak phenomenon. This can be asserted if we impose some constraints to the reference trajectories. In order to ensure this, we take $u_2(\psi=0)=u_2(\psi=\pi/2)=0$ with $\phi=\theta=0$. For $\psi=0$, one uses (2) and for $\psi=\pi/2$ expression $(Eq.\ 6)$ is considered. As soon as we get $u_3(\psi=0)=u_3(\psi=\pi/2)=mg$ taking $\phi=\theta=0$.

3.2 Self-tunable Fuzzy Inference System

A Fuzzy controller based on an on-line optimization of a zero order Takagi-Sugeno fuzzy inference system is successfully applied. It is used to minimize a cost function that is made up of a quadratic error term and a weight decay term that prevents an excessive growth of parameters of the consequent part. The main idea is to generate the conclusion parts w_i (so-called weight) of the rules automatically thanks to an optimization technique. The used method is based on a back-propagation algorithm where the parameters values are free to vary during the optimization process. The shape of the used membership functions is triangular and fixed in order to extract and represent the knowledge from the final results easily. To deduce the truth value, we use the MIN operator for the composition of the input variables. For the control of the XSF, we use the architecture known as “mini-JEAN” as illustrated in the Figure 2 (top) [7].

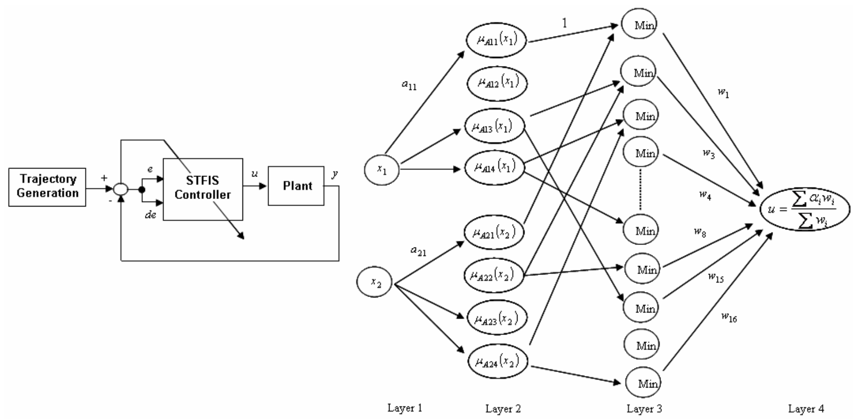


Fig. 2. Control architecture mini-JEAN (top). Architecture of STFIS controller (bottom).

Self-tunable fuzzy inference system presentation: A Sugeno type fuzzy system is determined in three stages:

1. Given an input x a membership degree μ is obtained from the antecedent part of rules.
2. A truth value degree α_i is obtained, associated to the premise of each rule R_i : IF x_1 is X_1 AND IF x_2 is X_2 THEN u IS w_i.
3. An aggregation stage to take into account all rules by $y = \sum_{i=1}^{r} \alpha_i w_i \Big/ \sum_{i=1}^{r} w_i$ permits to obtain a crisp value y.

4 Motion Planning and Simulation Results

The XSF is tested in simulation in order to validate some motion planning algorithms considering the proposed STFIS control. The objectives are to test the capability of the engine to fly with rounded intersections and crossroads. The two internal degree of freedom lead to a longitudinal/lateral forces which permit to steer the system with rounded corner flying road. The proposed control inputs permit to perform the tracking objectives; flying road with straight and round corners like connection.

Starting with a preinitialized rules table, when XSF begins to fly, it performs the acquisition of the distances (observations), calculates the cost function to back-propagation, updates the triggered rules in real time, begins to move and so on. The weights w_i are then adjusted locally and progressively.

The universes of discourse are normalized and shared in five fuzzy subsets for all displacements. The linguistic labels are defined as NB : Negative Big, NS : Negative Small, Z : Approximately Zero, PZ : Positive Small and PB : Positive Big. The results of the simulation are reported in the Table 1 for the z displacement. The optimization phase tends to stable weights (Figure 3 (bottom)). In these circumstances the outputs linguistic labels could be interpreteted as follows (Figure 3 (top)) VW: {1, 5} Very Weak, W: {7, 11} Weak, M: {19, 23} Medium, B: {28, 32} Big and VB: {34, 42} Very Big.

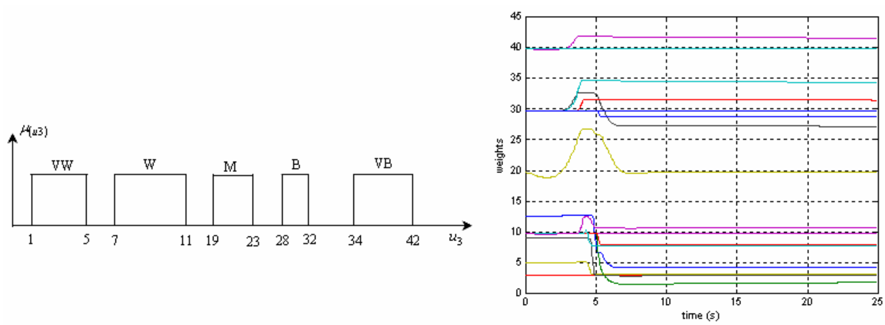


Fig. 3. Linguistic translation representation of the controller u_3 (top). The weights evolutions (bottom).

Table 1. Weight for z displacement

de/e	*NB*	*NS*	*Z*	*PS*	*PB*
NB	29,62	29,71	7,59	2,55	2,99
NS	29,52	31,34	10,63	3,02	1,69
Z	34,22	29,73	19,54	4,84	2,79
PS	39,72	41,37	22,32	1,74	9,61
PB	39,92	39,82	28,27	7,96	9,62

Table 2. Linguistic table for z displacement by learning (top) and by expertise (bottom)

de/e	*NB*	*NS*	*Z*	*PS*	*PB*
NB	B	B	W	VW	VW
NS	B	B	W	VW	VW
Z	VB	B	M	W	VW
PS	VB	VB	M*	VW*	W
PB	VB	VB	B	W	W

de/e	*NB*	*NS*	*Z*	*PS*	*PB*
NB	B	B	W	VW	VW
NS	B	B	W	VW	VW
Z	VB	B	M	W	VW
PS	VB	VB	M*	VW*	W
PB	VB	VB	B	W	W

The Table 2 (top), illustrates the linguistic translation of the table obtained by on-line optimization for the z displacement (Table 1). By comparing the table proposed by human expertise, Table 2 (bottom) and Table 2 (top), we can observe that the two sets of linguistic rules are quite close. Two cases noted with (*) are different and they differ from only one linguistic concept (M instead B and VW instead W). So, we can claim that the extracted rules are quite logical and coherent.

On the other hand, the main advantage of the described technique is the optimization of the controller with respect to the actual characteristics of the engine.

The use of a cost function gathering a quadratic error and a term of regression of the weights enabled to achieve our goal. For this behavior, the building of the navigation controller is done entirely on-line by the optimization of a zero order Takagi-Sugeno fuzzy inference system by a back-propagation-like algorithm.

The Figure 4 (b) and (d) illustrate the controlled positions xyz using STFIS controller where u_3 and u_2 denote the command signals for z and for x and y directions respectively. Note that the input $u_3 = mg$ at the equilibrium state is always verified. The input u_2 tend to zero after having carried out the desired orientation of the vehicle. The 3D displacement is depicted with straight and round corners like connection on the Figure 4 (a) and (c). These figures show the effectiveness of the used controller.

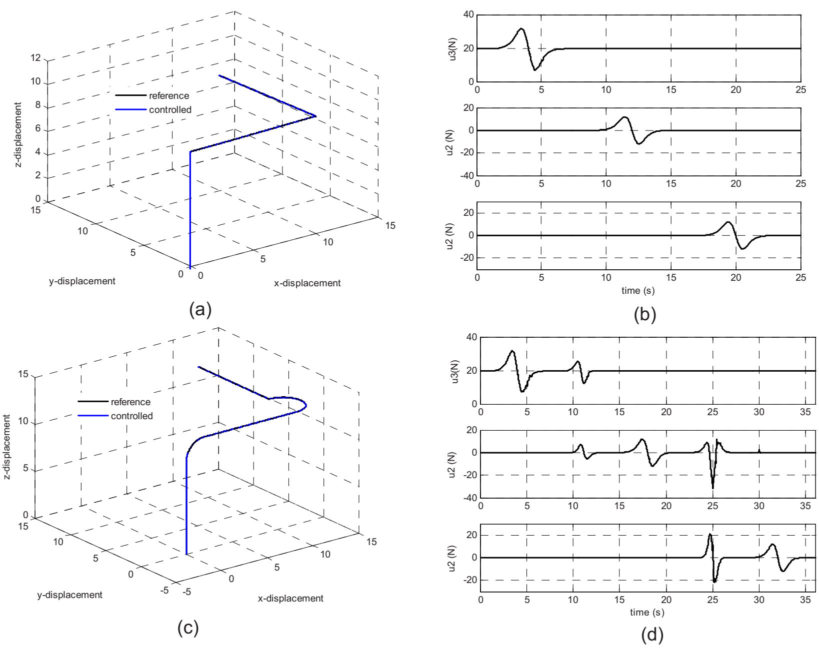


Fig. 4. Inputs u_3 and u_2 (b) for the realization of a straight corners (a). Inputs u_3 and u_2 (d) for the realization of a round corners (c).

5 Controllers Robustness

The robustness of the developed controllers are evaluated regarding external disturbances and performance degradations in the case of actuator/sensor failures and wind influence. In the case of the XSF, a resistance or a drag force is opposed to its movement in flight. The work produced by this force involves an additional energy consumption at the actuators levels which limits its maneuvering capacities in flight. This force can be expressed as follow:

$$F_i = \frac{1}{2} C_x \rho A V_i^2 \tag{8}$$

where $F_i[N]$ is the drag force following the i axis, $V_i[m/s]$ is the drone velocity, $A[m^2]$ is the coss-sectional area perpendicular to the force flow and $\rho[Kg/m^3]$ is the body density. The equation (8) induced a drag coefficient C_x which is a dimensionless quantity that describes a characteristic amount of aerodynamic drag depending on the XSF structure and which is experimentally determined by windtunnel tests. This coefficient is equal to 0.5 for the x and y directions and 0.08 for the z displacement. The surface characteristic A of the XSF drone is equal to $A = 0.031 m^2$ and its density is considered equal to $\rho = 1.22\, Kg/m^3$.

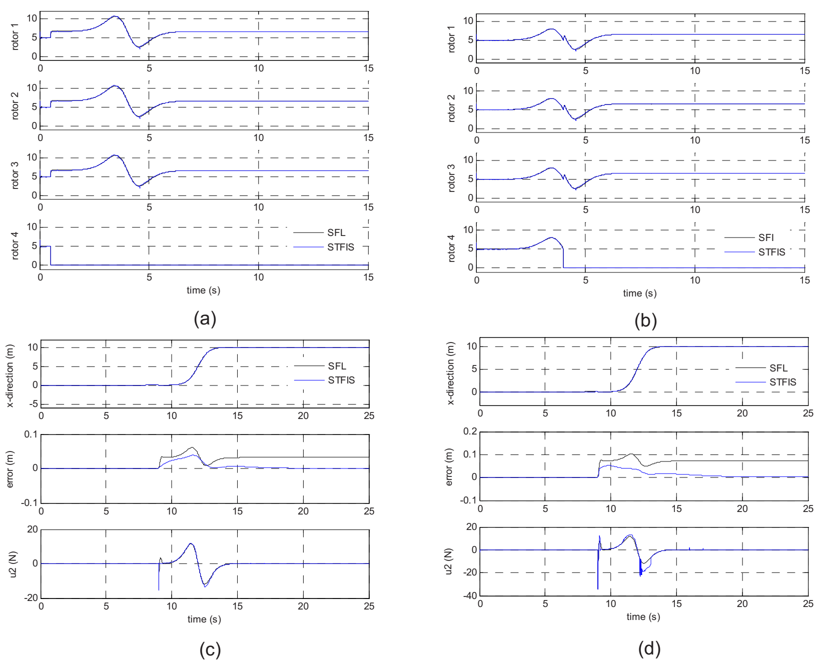


Fig. 5. XSF Forces in the case of motor 4 failures (a) at $t = 0.5\,\text{sec}$, (b) at $t = 4\,\text{sec}$. (c) Wind influence with a drag force of $1.4N$ and (d) $2.1N$ for the x direction.

The Figure 5 (a) and (b) illustrate the simulation results in the case of actuator 4 failure after take-off at the instant $t_1 = 0.5\,\text{sec}$ and $t_2 = 4\,\text{sec}$ in the z direction. To maintain its equilibrium, the three remain actuators share the drone load compensation where the practically results are in an equitable distribution of the developed forces ($F_1 + F_2 + F_3 = mg$ at steady state). The STFIS and the SFL controllers behave in the same way.

The Figure 5 (c) and (d) present the simulation results in the case of a drag force of $F_{dg} = 1.4N$ and of $F_{dg} = 2.1N$ according to the x displacement. The STFIS controller exhibits chattering signal problems in the transition phase while the SFL controller presents static errors that varies proportionally with the drag force amplitude F_{dg} . The same observations are found according to the two directions y and z .

6 Conclusion

In this paper, we studied a new configuration of flyer engine called XSF. We have considered in this work, the stabilizing/tracking control problem for the three decoupled displacements. The objectives are to test the capability of the engine to fly with straight and rounded intersections. We have presented and implemented an optimization technique allowing an on-line adjustment of the fuzzy controller parameters. The descent gradient algorithm, with its capacities to adapt to unknown situations by the means of its faculties of optimization, and the fuzzy logic, with its capacities of empirical knowledge modelling, are combined to control a new bidirectional drone. Indeed, we have obtained an on-line optimized Takagi-Sugeno type SIF of zero order. A comparison between the STFIS set rules and that deduced by human expertise, shows the validity of the proposed technique. An analysis of the STFIS and the SFL controllers and their robustness regarding disturbances, shows the advantages and the disadvantages of these two techniques.

Future works will essentially investigate the real time implementation of the STFIS and the based-model control techniques. Obstacles avoidance and flying multi-drones are also envisaged thanks to the SIF faculties and its optimization capabilities.

References

1. Altug E (2003) Vision based control of unmanned aerial vehicles with applications to an autonomous four rotor helicopter, quadrotor. Ph.D. thesis, Faculties of the University of Pennsylvania
2. Azouz N, Bestaoui Y (2005) Modelling and simulation of a mini quad-rotor helicopter. Proceeding of DTM2005, ASME Design Theory and Methodology Coferences, Long Beach, California, USA
3. Beji L, Abichou A, Zemalache K M (2005) Smooth control of an X4 bidirectional rotors flying robots. Fifth International Workshop on Robot Motion and Control, Dymaczewo, Poland

4. Castillo P, Dzul A, Lozano R (2004) Real-time stabilization and tracking of a four rotor mini-rotorcraft. IEEE Transactions on Control Systems Technology, Vol. 12, N 4, pp 510-516
5. Hamel T, Mahony R, Lozano R, Ostrowski J P (2002) Dynamic modelling and configuration stabilization for an X4-flyer. IFAC 15th World Congress on Automatic Control, Barcelona, Spain
6. Jang J S R (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions System, Man and Cybernet.23, pp 665-685
7. Maaref H, Barret C (2001) Progressive optimization of a fuzzy inference system. IFSA-NAFIPS'2001, Vancouver, pp 665-679
8. Tayebi A, McGilvray S (2004) Attitude stabilization of a four-rotor aerial robot. 43rd IEEE Conference on Decision and Control, Atlantis, Bahamas
9. Zemalache K M, Beji L, Maaref H (2005) Control of an under-actuated system: application to a four rotors rotorcraft. 5th IEEE International Conference on Robotics and Biomimetics, Hong Kong, China, Vol. 1, pp 160-230
10. Zemalache K M, Beji L, Maaref H (2006) Neuro-Fuzzy system controlling a X4-Flyer. 5th International Symposium on Robotics and Automation 2006, San Miguel Regla Hidalgo, México, pp 460-465
11. Zemalache K M, Beji L, Maaref H (2006) Motion planning and stabilisation of a X4 Bidirectional rotors using a Neuro-Fuzzy controller. 9th International Symposium on Climbing and Walking Robots and Associated Technologies, Royal Military Academy, Brussels, Belgium, pp 678-683

A Kind of Embedded Temperature Controller Based on Self-turning PID for Texturing Machine

Tan Dapeng, Li Peiyu, and Pan Xiaohong

College of Mechanical and Energy Engineering, Zhejiang University,
Hangzhou 310027, China
{tandapeng,lipeiyu,pan_xh}@zju.edu.cn

Abstract. Temperature control system (TCS) of texturing machine (TM) has many control channels, requires high control precision, and it is prone to be influenced by peripheral environmental factors. Since traditional temperature control method hardly realize ideal control effect, a kind of TCS realization method based on self-turning PID technology for constant temperature box (CTB) of TM was put forward based on the wide analysis of characteristics and methods of TCS. According to structure characteristic and working environment of TM-CTB, its discrete mathematics model was established by thermodynamics principle. The self-turning PID minimum phase control system for TM temperature controller was designed using pole points placement method, and its application instance was established by embedded system technology. System parameters were identified by recursion least square method, and forgetting factor was introduced to improve the model tracing ability of TCS. Experiments results showed that the method can satisfy the multi-channel requirement of texturing machine TCS, and its temperature control error can be limited under one percent.

1 Introduction

TM is a tensile strain machine used for producing long polyester fiber filament, it can executes the treatment of tensile strain and stretcher strain for the fiber. By the process of TM, general fiber can turned into special fiber with high resilience. CTB is the crucial controlled target of TM, and the precision of the TCS will influent the quality index such as thrown silk curl degree, coloration, shrinkage ratio in boiling water. CTS of TM-CTB requires to control many channels at the same time, and has a high precision desire of target temperature. What's more, the decalescence of the fiber material, speed, ambient temperature, power of the heating pipe can impact the TCS control effect [1]. Traditional temperature control methods hardly realize ideal control effect, so it is necessary to develop a TM-TCS which can resolve the problems mentioned above.

2 Discussion and Analysis of TCS

2.1 Conventional TCS Model

Temperature control can be presented by the mathematic model of single-order inertia and lag annulus.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 701–709, 2007.
springerlink.com

$$G(s)=\frac{K}{(1+T_d s)(1+\tau s)}=\frac{K}{T_d\tau s^2+(T_d+\tau)s+1} \tag{2.1}$$

In expression (2.1), K is amplification coefficient; $T_d = C/A$ is time constant, C is specific heat and A is radiating coefficient. Consequently, it has a great variation according to the different controlled targets and ambient conditions. The pure lag annulus can arouse system overshoot and sustained oscillation, and the parameter of temperature control target will also have a large scale change [2]. Therefore, it's difficult to achieve high control accuracy by this TCS method.

2.2 TCS Based on PID

Proportion-integral-differential (PID) controller is a linearity controller, its control rule is presented as followed formulation.

$$u(t)=K_P\left[e(t)+\frac{1}{T_I}\int_0^t e(t)dt+\frac{T_D de(t)}{dt}\right] \tag{2.2}$$

K_p is proportional constant, T_I is integral time constant, T_D is differential time constant [3].The practice effect and theory analysis prove that this control rule can get better effectiveness in large number of industrial control process, that mainly express two facets as followed: firstly, its principle and structure are very simple, can fill the requests of mass industrial processes; secondly, it can appropriate for many different controlled targets and its arithmetic has a stronger robustness [4].The current PID control method can be divided into two modes: digital control and analogue control. The digital control method is prone to be realized by computer, and it has many good advantages such as good flexibility, low cost and reliable performance. It can make TCS reach the level with high accuracy and intelligence [5].

The self-turning control is an effective method which can resolve the non-linearity and time-variation process. The self-turning PID control method integrates the merits of self- turning and PID control, it only require few parameters to set and can modify the controller parameter on-line according to the variation of system parameters. Moreover, it can improve the dynamic characteristics of speed regulating system by the adequate pole assignment. Because it has smaller calculating quantity than the control method of generalized minimum variance with pole assignment, easy to control real-time system and has successful instances in the process of industry control, so this paper adopts it to design the TCS of TM [6]-[7].

3 System Design of TM Temperature Controller

3.1 Mathematic Model of Controlled Target

The working process of TM is shown in Fig.1-a, in order to be analyzed and processed by computer conveniently, system discrete mathematics model was established by thermodynamics principle and shown in Fig.1-b.

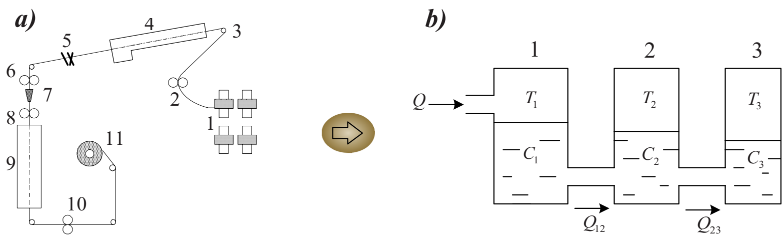


Fig. 1. *a)* TM process flow. 1.protofilament 2.first roller 3.stop twister 4.first heat-box 5.false twister 6.second roller 7.network snoot 8.ZBIS roller 9.second heat-box 10.third roller 11.package *b)* Discrete mathematics model of TM-CTB. 1,2,3 respectively represent air cell, insulating layer and outside air; T_1, T_2, T_3 are temperatures of air cell, insulating layer and outside air; C_1, C_2, C_3 are specific heat correspondingly; Q is input thermal power; Q_{12}, Q_{23} are thermal transmissions from air cell to insulating layer and insulating layer to outside air.

According to the thermodynamics principle, some expressions can be obtained as followed.

$$Q - Q_{12} = C_1 m_1 \frac{dT_1}{dt} \tag{3.1}$$

$$Q_{12} = K_{12} A_1 (T_1 - T_2) \tag{3.2}$$

$$Q_{12} - Q_{23} = C_2 m_2 \frac{dT_2}{dt} \tag{3.3}$$

$$Q_{23} = K_{23} A_2 (T_2 - T_3) \tag{3.4}$$

In above expressions, K_{12}, K_{23} are the heat transfer coefficient; m_1, m_2 are the mass of air cell and insulating mediator; A_1, A_2 are the heat transfer areas. Because $c_1 m_1$, $c_2 m_2$, $K_{12} A_1$, $K_{23} A_2$ are all constant, so they are recorded as c_1, c_2, K_{12}, K_{23} respectively. T_1 is output quantity and T_3 is constant [8]. The expression of system transfer function is

$$G(s) = \frac{G(T_1)}{G(Q)} = \frac{K(T_4 s + 1)}{T_5^2 s^2 + 2\xi T_5 s + 1} \tag{3.5}$$

$$K = \frac{K_{12} + K_{23}}{K_{12} K_{23}},\ T_5^2 = \frac{c_1 c_2}{K_{12} K_{23}},\ T_4 = \frac{c_2}{K_{12} K_{23}},\ \xi = \frac{1}{2} \frac{c_1 K_{12} + c_1 K_{23} + c_2 K_{12}}{\sqrt{K_{12} K_{23} c_1 c_2}}.$$

The expression of difference model is

$$y(k) + a_1 y(k-1) + a_2 y(k-2) = b_0 u(k-d) + b_1 u(k-d-1) + e(t) \tag{3.6}$$

In above expression, d represents the number of sampling period for system lag. Obviously, TCS of CTB is a SISO system, its control block diagram is shown in Fig.2 [9]-[10].

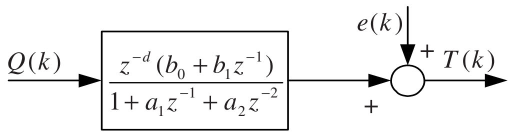


Fig. 2. Control block diagram of target

3.2 Parameter Recognition of System Model

After establishment of system model, this paper required to recognize the relative parameter referred. To reduce the calculation load, the algorithm of recursive least-square (RLS) is adopted. According to the control characteristic, this paper can obtain the loop-locked system control block diagram via recursive recognition, that's shown in Fig.3 [11].

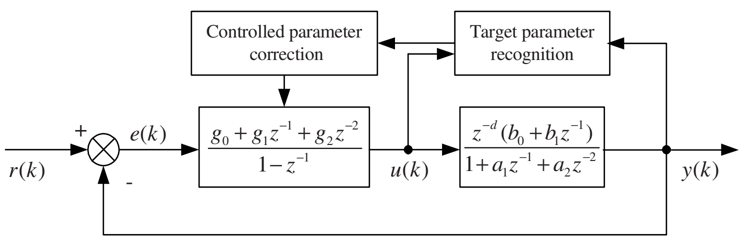


Fig. 3. Loop-locked system recognition control block diagram

The Z transfer function of this digital PID controller is

$$G(z) = \frac{1}{1-z^{-1}}\left[K_P(1-z^{-1}) + K_I + K_D(1-z^{-1})^2\right] \tag{3.7}$$

Above expression can be represented as followed:

$$G(z) = \frac{g_0 + g_1 z^{-1} + g_2 z^{-2}}{1 - z^{-1}} \tag{3.8}$$

$$g_0 = K_P + \frac{K_P T}{T_I} + \frac{K_P T_D}{T}, \quad g_1 = -\left(K_P + \frac{2K_P T_D}{T} \right), \quad g_2 = \frac{K_P T_D}{T}$$

From above description, it is clearly that PID control is a second-order control, its number of order equals to control targets', so it can satisfy the condition of system lock-loop recognition. The self-turning algorithm of TM-TCS adopts the designation method of pole assignment. Pole assignment self-turning controller is based on normal procedure control strategy, its designation steps can be described as followed: *1)* Determining the desired pole position of lock-loop system; *2)* Estimating and recognizing system parameter on-line; *3)* Calculating the parameters of controller; *4)* Calculating the efficiency of controller [12]. Because the order-number of system is two, so its lock-loop transfer function can be expressed by

$$\frac{y(k)}{r(k)} = \frac{\dfrac{g_0(1+g_1' z^{-1} + g_2' z^{-2})}{(1-z^{-1})(1+f_1' z^{-1})} \bullet \dfrac{z^{-d} \cdot b_0 \left(1+\dfrac{b_1}{b_0} z^{-1}\right)}{1+a_1 z^{-1} + a_2 z^{-2}}}{1+\dfrac{g_0(1+g_1' z^{-1} + g_2' z^{-2})}{(1-z^{-1})(1+f_1' z^{-1})} \bullet \dfrac{z^{-d} \cdot b_0 \left(1+\dfrac{b_1}{b_0} z^{-1}\right)}{1+a_1 z^{-1} + a_2 z^{-2}}} \tag{3.9}$$

Form the expression (3.5), the zeros of control targets transfer function are located left-half plane of root plane($T_4 > 0$), so it is a minimum phase system [13]. For this kind of system, this paper supposes

$$1+g_1' z^{-1} + g_2' z^{-2} = 1 + a_1 z^{-1} + a_2 z^{-2}, 1 + \frac{b_1}{b_0} z^{-1} = 1 + f_1' z^{-1}, g_2' = a_2, g_1' = a_1, f_1' = \frac{b_1}{b_0}.$$

Therefore, the relational expression of target parameters and controller parameters is established, so this paper can obtain that

$$G(z) = \frac{g_0(1+g_1' z^{-1} + g_2' z^{-2})}{(1-z^{-1})(1+f_1' z^{-1})} = \frac{g_0(1+a_1 z^{-1} + a_2 z^{-2})}{(1-z^{-1})\left(1+\dfrac{b_1}{b_0} z^{-1}\right)} \tag{3.10}$$

$$u(k) = u(k-1) + g_0 e(k) + g_0 \hat{a}_1 e(k-1) + g_0 \hat{a}_2 e(k-2) + \frac{\hat{b}_1}{\hat{b}_0}\left[u(k-2) - u(k-1)\right] \tag{3.11}$$

In expression (3.10), $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_0$, $\hat{b}_1$ are system parameters which obtained by recognition on-line. So the PID control parameters will vary with the variation of controlled targets, it is called self-turning PID control.

4 Realization of TCS Application Instance

4.1 Embedded Hardware Control System Based on ARM

Compare to the common industry temperature controller, TM-TCS has so many controlled points, which requires controller have the ability of multi-channel selection and multi-channel digit output. In circuit design part, 120 unit select-channel of system is composed of multi-channel analog switch chip. On the sides of digit output, this paper adopts five *8255* chips to extend to 120 unit output [14].

The core of hardware system is an embedded computer which has *Intel PXA-255* CPU and corresponding LCD human computer interface, it takes charge of selection and switchover of channel, calculation and display of temperature, process of depressed key interrupt and digit output, and it sends control information to TM-CTB by the process of driver circuit and power amplification circuit. System hardware structure is shown in Fig.4.

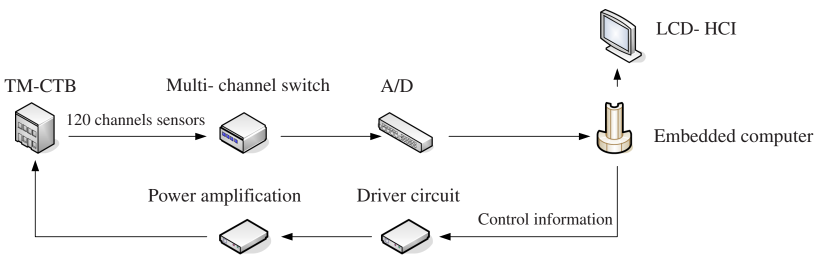


Fig. 4. Hardware structure of TM-TCS

4.2 Anti-interference Design and System Software Flow

According to analysis and investigation of TM work environment, this paper finds there are interference sources as followed: power, signal line and ground wire. In allusion to above interference sources, two methods were adopted to resolve it. In the facet of hardware, some anti-interference measure such as optoelectronic isolating must be adopted. In the other hand, the system software needs to be optimized and make it have some fault tolerant ability.

So the designation of software self-check program is realized according to its character, it adopt watch-dog method that one interrupt program and the status of main program are monitored by other interrupt program. This kind of anti-interference method improved the stability of system in large extent, the system software process is shown in Fig.5 [15].

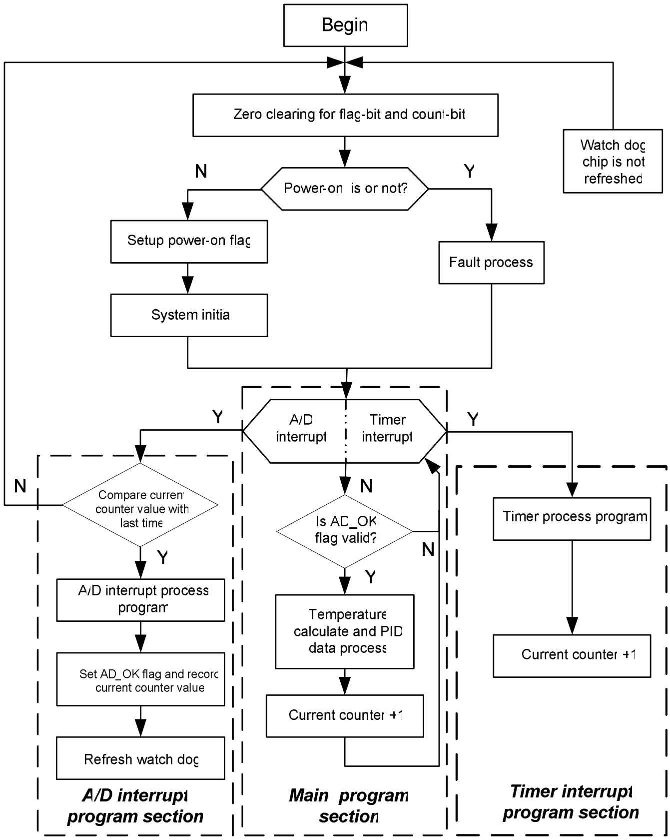


Fig. 5. System monitoring software process flow

5 Parameters Recognition and Experiments Outcome Analysis

5.1 Parameters Recognition

In order to overcome the problem of data-saturation, this paper fetches the forgetting factor into the process of parameters recognition, that can quickly trace the variation of model parameters and reduce the transition process, and obtains the emulation calculation outcome (as shown in Tab.1, k is the data length) by using RLS method [16]. Data outcome of the table can prove this method's validity clearly.

Table 1. Parameters recognition outcome

Pra.	a_1		a_2		b_0		b_1	
	k<150	k>150	k<150	k>150	k<150	k>150	k<150	k>150
Exp.	-1.49	-1.05	0.69	0.49	0.09	0.069	0.0056	0.0038
Cal.	-1.50	-1.05	0.70	0.49	0.10	0.070	0.0050	0.0035

5.2 Verification Experiments

In order to verify the validation and superiority of self-turning PID control further, this paper make comparison experiment of common PID and self-turning PID. The heat device is a 300W electric cup which holds water with room temperature, temperature sensor submerges in the water. In the incipient stage of experiment, this paper adopts common PID control algorithm, at the time point of t, this paper change the control algorithm instead of self-turning PID to compare the effect of two algorithms. Fig.6 is the compare figure of control effect, the horizontal ordinate is time and the vertical ordinate is temperature. It is clearly see that the accuracy of TCS is improved apparently after adopting the self-turning algorithm.

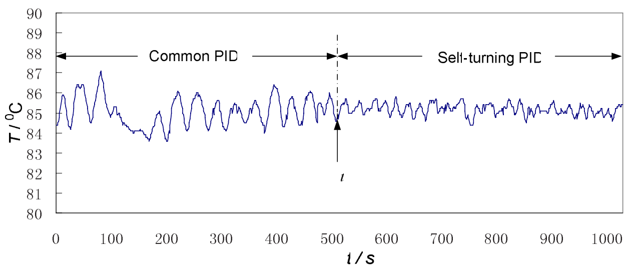


Fig. 6. Outcome data wave of verification experiments

6 Conclusion

1) A kind of TCS realization method for TM-TCB is put forward, it can resolve the problems of TM temperature control such as too many control points, environment factors sensitivity etc; *2)* According to characteristic analysis of TM-CTB, its discrete mathematics model was established by thermodynamics principle; *3)* By pole points placement method, self-turning PID minimum phase control system for TM-TCS was finished, and its application instance was established using embedded system technology; *4)* Because of non-determinacy of industry environment, so the future research work will be carried out around by stability design and system optimization.

References

1. L. Li, K. Y. Li, "The reasons for the HDS-EX2 type textured machine take-up package unevenness and their settlements," Polyester Industry, vol. 17, no. 1, pp. 43-45, Feb. 2004.
2. D. Zhang, M. H. Du, "Ringing Artifacts Reduction Using Adaptive Filtering," Computer Engineering and Applications, vol. 41, no. 12, pp. 135-139, Dec. 2005.
3. J. Q. Xiong, Computer Control Technology. Bei-jing: Publishing House of Electronics Industry, 2003, pp. 230-242.
4. J. L. Ye, "Application of Self-tuning PID Method in the Governing System of the Hydroelectric Generating Unit," Transactions of China Electro-technical Society, vol. 17, no. 4, pp. 53-56, Apr. 2002.

5. S. G. Shi, X. Shi, “Auto-control System of High-precision Const-Temperature Based on Digital PID and Digital Potentiometer,” Journal of Wu-han Poly-technic University, vol. 24, no. 1, pp. 75-78, Feb. 2005.
6. Q. C. Chen, L. Peng, Y. Kang, “PID Controller Parameters' Optimization for DC/DC Converters Based on Pole Deploying,” Telecom Power Technologies, vol. 22, no. 4, pp. 78-81, Apr. 2005.
7. M. Wang, L. Shi, “Influence analysis of covariance analysis model,” Journal of Yun-nan University(Natural Sciences Edition),vol. 25, no. 5, pp. 391-394, May. 2003.
8. H. C. Wu, J. Deng, J. Zhou, “Application of Thermodynamics Model on Aqueous Two-phase System,” Journal of Si-chuang University of Science & Engineering(Natural Science Edition), vol. 18, no. 1, pp. 74-78, Jan. 2005.
9. W. K. Liu, Y. D. Lou, X. H. Zhang, “Implementation of Single-Frequency Pseudo-ranges Double-Differenced Model Applied to Mobile-Mobile Target Monitoring. Journal of Geomatics, vol. 29, no. 2, pp. 36-38, Apr. 2004.
10. J. Hu, X. W. Li, “An improved analyzing way of the stability condition about a liner time-invariant SISO system,” Journal of Shan-dong University of Technology(Science and Technology), vol. 19, no. 2, pp. 35-39, Feb. 2005.
11. Y. Sun, B. Jing, J. Zhang, “Multi-sensor Information Fusion Based on the Combination of Optimal Weight and Recursive Least Square Algorithm,” Chinese Journal of Sensors and Actuators, vol. 17, no. 4, pp. 157-160, Apr. 2004.
12. Z. Yang, “The Method of Designing and Key Points of the Industrial Auto-turning PID Controllers,” Control and Instruments in Chemical Industry, vol. 18, no. 1, pp. 74-78, Jan. 2005.
13. J. Zhang, G. X. Wang, L. He, “Robust control of systems with unstable zero-dynamics,” Control Theory & Applications, vol. 22, no. 1, pp. 99-103, Jan. 2005.
14. Z. C. Yan, Z. M. Li, D. Q. Song, “Application of Embedded SCM System in Subminiature pH Meter,” Journal of Northeast China Institute of Electric Power Engineering, vol. 23, no. 1, pp. 55-57, Feb. 2003.
15. S. W. Jia, K. Gu, “Study of the CCM of the Embedded SCM System in Electronic Equipment,” Avionics Technology, vol. 35, no. 3, pp. 48-51, Sep. 2004.
16. Z. L. Feng, J. Liu, K. Z. Liu, “Dead Reckoning Method for Autonomous Navigation of Autonomous Underwater Vehicles,” Robot, vol. 27, no. 2, pp. 168-172, Apr. 2005.

Trajectory Tracking Using Fuzzy-Lyapunov Approach: Application to a Servo Trainer

Jose A. Ruz-Hernandez[1], Jose L. Rullan-Lara[1], Ramon Garcia-Hernandez[1], Eduardo Reyes-Pacheco[2], and Edgar Sanchez[3]

[1] Universidad Autonoma del Carmen, Calle 56 # 4 por Avenida Concordia, CP 24180, Colonia Aviación, Cd. del Carmen, Campeche, Mexico
{jruz,jrullan,rghernandez}@pampano.unacar.mx
[2] Universidad Tecnologica de Campeche, San Antonio, Cardenas, Mexico, CP 24381
ereyes@utcam.edu.mx
[3] CINVESTAV, Unidad Guadalajara, Apartado Postal 31-430, Plaza La Luna, C.P. 45091, Guadalajara, Jalisco, MEXICO, on sabbatical leave at CUCEI, Universidad de Guadalajara
sanchez@gdl.cinvestav.mx

Abstract. This paper presents a Fuzzy-Lyapunov approach to design trajectory tracking controllers. This methodology uses a Lyapunov function candidate to obtain the rules of the Mamdani-type fuzzy controllers which are implemented to track a desired trajectory. Two fuzzy controllers are implemented to control the position and velocity of a servo trainer and real time results are presented to evaluate the performance of designed controllers against the performance of classical controller.

1 Introduction

The most difficult aspect in the design of fuzzy controllers is the construction of rule base. The process of extracting the knowledge of human operator, in the form of fuzzy control rules, is by no means trivial, nor is the process of deriving the rules based on heuristics and a good understanding of the plant and control theory [1], [2].

We present an application of *fuzzy Lyapunov synthesis method* to an educational servo trainer equipment [3]. We design two controllers, the first one is used to solve the trajectory tracking problem when a desired reference is used to determine the behavior of the angular position and the second one has the same goal when a desired reference is used to determine the behavior of the angular velocity. Real time results using the designed fuzzy controllers are illustrated and compared to the obtained results via conventional controllers, which have been suggested by the servotrainer designers [3].

Additionally to this section, the paper is organized as follows. In section 2 we present the theory related to fuzzy Lyapunov synthesis method as described in [1], [4], [5], [6], [7], [8], [9] and [10]. Section 3 describes the mathematical model of servo trainer, which is used as the plant to apply the designed controllers. In section 4 the design of both position and velocity fuzzy controller is presented. Real time results are presented in section 5. Finally, the conclusions are established in section 6.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 710–718, 2007.
springerlink.com

2 Fuzzy Lyapunov Synthesis Method

In this section we describe the proposed fuzzy Lyapunov synthesis method to design controllers. We start with the conventional case when we have an exact mathematical description of the plant and then describe its extension to the fuzzy case.

Consider the single-input, single output system

$$\dot{x} = F(x,u), \quad y = h(x). \tag{1}$$

where $\mathrm{F}(\cdot) = (F_1(\cdot), F_2(\cdot), \ldots, F_n(\cdot))^T$ with $F_i(\cdot)'s$ being continuous functions, $u \in R$ and $y \in R$ are the input and output of the system, respectively, and $\mathrm{x} = (x_1, x_2, \ldots, x_n)^T \in R^n$ is the state vector of the system. The control objective is to design a feedback control $u(x)$ so that $\mathbf{0}$ will be stable equilibrium point of (1).

One way of achieving this goal is to choose a Lyapunov function candidate $V(x)$ and then determine the conditions on u necessary to make it Lyapunov function. The Lyapunov function candidate follows the next requirements

1. $V(\mathbf{0}) = 0$
2. $V(x) > 0, \quad x \in N \setminus \{\mathbf{0}\}$
3. $\dot{V} = \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} \dot{x}_i < 0, \quad x \in N \setminus \{\mathbf{0}\}$

where $N \setminus \{\mathbf{0}\} \in R^n$ is some neighborhood of $\mathbf{0}$ excluding the origin $\mathbf{0}$ itself, and $\dot{x}_i \ (i = 1, 2, \ldots, n)$ is given by (1). If $\mathbf{0}$ is an equilibrium point of (1) and such V exist, then $\mathbf{0}$ is locally asymptotically stable.

We determine a Lyapunov function candidate V, derive an expression for its derivative, and then obtain conditions so that V would indeed be a Lyapunov function. However, in this case, the conditions obtained can be extended to a fuzzy controller design [1], [2], [8], [10].

We can formulate the resulting conditions in the form of rules in one of two possible representations. The first form is:

IF x_1 is $\langle lv \rangle$ and/or x_2 is $\langle lv \rangle$ $\cdots$ and/or x_n is $\langle lv \rangle$ THEN u must be $\langle lv \rangle$

where the $\langle lv \rangle$ are linguistic values. The second possible form is

IF x_1 is $\langle lv \rangle$ and/or x_2 is $\langle lv \rangle$ $\cdots$ and/or x_n is $\langle lv \rangle$ THEN u must be $f(x_1, x_2, \cdots, x_n)$

where $f(\ )$ is a linear function.

These rules constitute the rule base for Mandani-type or Takagi-Sugeno-Kang-type (TSK) fuzzy controller, respectively. We refer to the method of deriving the rules as the *fuzzy Lyapunov synthesis method*.

3 Servo Trainer Apparatus

The equipment used as plant to control in this paper is the CE110 Servo Trainer from Tecquipment [3]. This apparatus is used to help in teaching linear control theory and to implement some applications in real time like the proportional-integral-derivative controller. The apparatus have a variable load which is set using a current direct generator, by changes of different inertial load and using the engage a gearbox or by set all of them together. Besides, the apparatus have three modules to introduce some nonlinearity which can modify the behavior of servotrainer.

The mathematical model of the servotrainer is expressed by the equations

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{1}{\tau}x_2 + \frac{G_1 G_2}{\tau}u \end{aligned} \tag{2}$$

where $x_1 = \theta$ and $x_2 = \omega$ are the angular position and the angular velocity, respectively. The gains G_1 and G_2 are defined by $G_1 = k_i k\omega$ and $G_2 = k_\theta / 30k_\omega$ where k_i is the motor constant, k_ω is the velocity sensor constant and k_θ is the angle sensor constant; τ is the time constant of the system. Table 1 shows all values of these parameters.

Table 1. Parameters of the servo trainer

Parameter	Value	Units
k_i	3.229	rev / sec-Volts
k_ω	0.3	Volts / (rev/sec)
k_θ	20	Volts / rev
G_1	1	Undimentional
G_2	2.22	Seconds
τ	1.5 small load 1 medium load 0.5 full load	Seconds

4 Design of Mamdani-Type Controllers

In this section, we present the design of two controllers using fuzzy Lyapunov methodology. The design is done as follows.

4.1 Position Mamdani-Type Controller

The objective is to design $u(x_1, x_2)$ such that the position behavior of servo trainer x_1 follows a reference signal y_θ .The reference y_θ and its derivatives $\dot{y}_\theta$ and $\ddot{y}_\theta$ are bounded and available to the controller. We choose the Lyapunov function

candidate $V = \frac{1}{2}\left(e^2 + \dot{e}^2\right)$ where $e = x_1 - y_\theta$. This function hold conditions (1) and (2) and we only need to consider condition (3). Differentiating V yields $\dot{V} = e\dot{e} + \dot{e}\ddot{e} = e\dot{e} + \dot{e}\left(\dot{x}_2 - \ddot{y}_\theta\right)$
and denoting

$$w = \dot{x}_2 - \ddot{y}_\theta \, , \tag{3}$$

we require that

$$\dot{V} = e\dot{e} + \dot{e}w < 0 \, . \tag{4}$$

There are three possible combinations that hold the expressions (4) depending of the sign of error, e.

1. If e and $\dot{e}$ are both positive we must have $w < -e$
2. If e and $\dot{e}$ are both positive we must have $w > -e$
3. If e and $\dot{e}$ have opposite signs we must have $w = 0$

These combinations can be used to generate a rules base for w as follows:

IF e is positive and $\dot{e}$ is positive THEN w is negative big
IF e is negative and $\dot{e}$ is negative THEN w is positive big
IF e is *positive* and $\dot{e}$ is *negative* THEN w is *zero*
IF e is *negative* and $\dot{e}$ is *positive* THEN w is *zero*

Triangular membership functions, center of gravity defuzzifier and the product inference engine are used to obtain the following w

$$w = f_1\left(e, \dot{e}\right) + f_2\left(e, \dot{e}\right) . \tag{5}$$

By means of the equations (2) y (3) we obtain the following control law

$$u = \frac{\tau}{G_1 G_2}\left(w + \ddot{y}_\theta\right) + \frac{1}{G_1 G_2} x_2 \, . \tag{6}$$

4.2 Velocity Mamdani-Type Controller

For this case, the objective is to design $u\left(x_1, x_2\right)$ such that the velocity behavior x_2 of servo trainer follows a reference signal y_ω. As well done in the design of position controller the conditions of the signal reference y_θ are held. The Lyapunov function candidate is still the same but the error signal is $e = x_2 - y_\omega$. Differentiating V yields $\dot{V} = e\dot{e} + \dot{e}\ddot{e} = e\dot{e} + \dot{e}\left(\dot{x}_2 - \ddot{y}_\omega\right)$ and denoting

$$w = \ddot{x}_2 - \ddot{y}_\omega \, , \tag{7}$$

and taking into account the same conditions as established in equation (4), we obtain

$$\dot{V} = e\dot{e} + \dot{e}w < 0\ . \tag{8}$$

Finally, using equations (2) and (7) we obtain the following control law

$$u = \frac{\tau}{G_1 G_2} \int \left(w + \ddot{y}_{\omega} + \frac{1}{\tau} \dot{x}_2 \right) dt\ . \tag{9}$$

5 Real-Time Results

In this section we show real time results using both position and velocity controller and its performances are compared with respect to conventional controller's performances which are taken from user manual of educational equipment [3]. All results are done considering that the servo trainer is operating only with a small load, which corresponds to a time constant equal to 1.5 seconds. Figure 1 shows the block diagraman of the application in real time. We used the Data Adquisition Target *PCI6071* manufactured by National Instruments and the *Xpc-Target Toolbox* of Matlab[1].

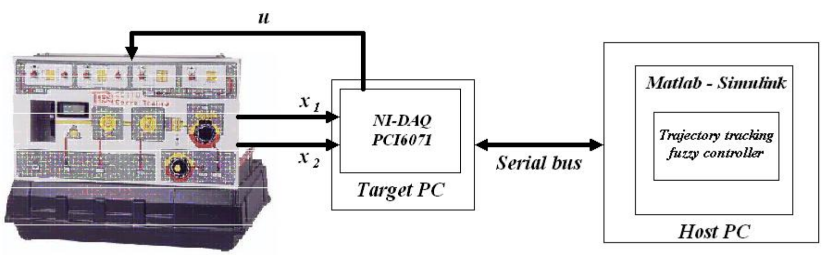


Fig. 1. Real time application using NI-DAQ PCI6071 and Xpctarget of Matlab 6.5

5.1 Application of the Position Controller

Figures 2–4 shows the real time results using the initial conditions (4π/9,0) and the signal reference $y_\theta = 5sin(0.5t)$ where its amplitude corresponds to 80 degree of angular position. The initial conditions were adjustment manually by using an independent supply applied to the actuator of servotrainer until. The proportional controller has a velocity feedback loop; the value of proportional gain is 10 and the gain of velocity feedback loop is 0.01 which are proposed to implement a conventional controller as indicated in [2].

As illustrated in figure 2, after 5 seconds the angular position of both fuzzy and proportional controllers track the signal reference, however a lag phase with the proportional controller is holded. This lag phase is reflected in the error plot of figure 4 where the steady state error is different to zero when the conventional controller is used. The error signal converges to zero after 5 seconds when our fuzzy controller is

[1] Matlab is a registered trademark of Mathworks Inc.

implemented. The control signal computed by the fuzzy controller presents some suddenly changes but is smaller than the control signal computed by proportional controller which is satured during two seconds at beginning.

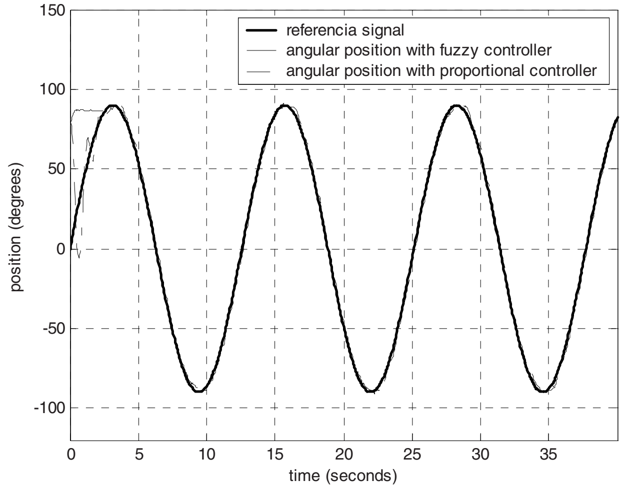


Fig. 2. The position angle of servotrainer x1 with fuzzy controller (dashed line), with a conventional controller (dashed-dot line) and the reference signal yθ = 5sin(0.5t) (solid line) when initial condition (4π/9,0) is used

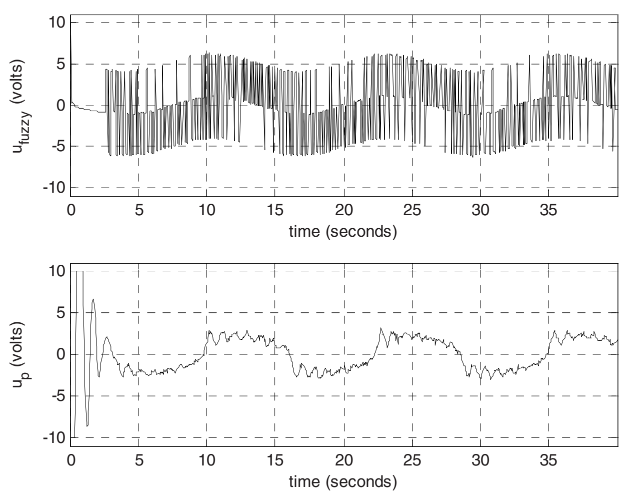


Fig. 3. The control signal u for angular position tracking when initial condition (4π/9,0) is used. The first plot (above) corresponds to fuzzy control and the second plot (below) corresponds to conventional control.

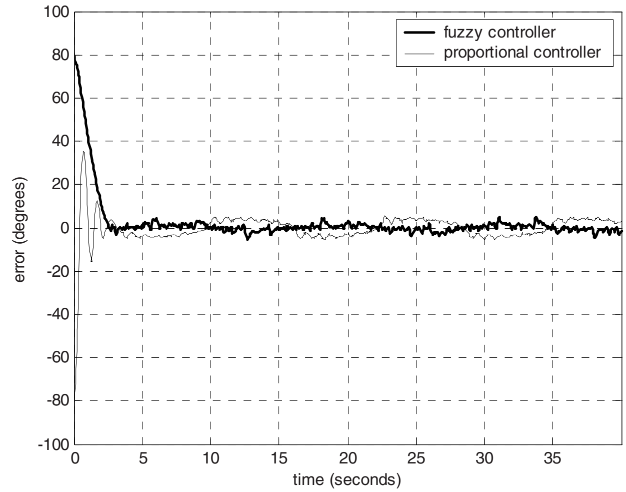


Fig. 4. The error signal for angular position tracking when initial condition (4π/9,0) is used. The error signal for fuzzy controller is defined by $e = x_1 - y_\theta$ and the error signal for proportional controller is defined by $e = y_\theta - x_1$.

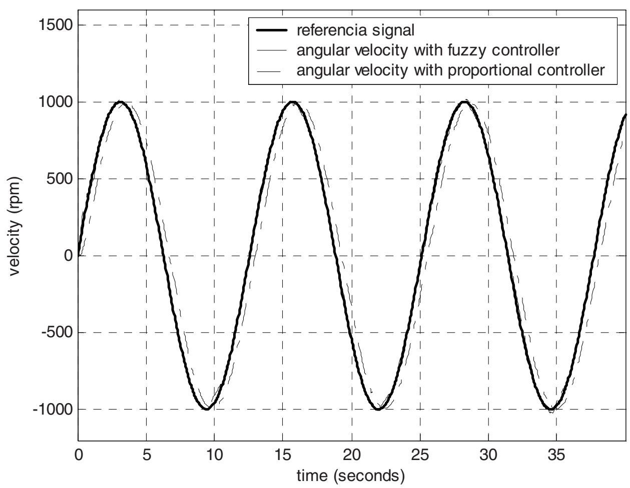


Fig. 5. The angular velocity of servotrainer x_2 with fuzzy controller (*dashed line*), with PI controller (*dashed-dot line*) and the reference signal $y_\omega = 5sin(0.5t)$ (solid line) when initial condition (0,0) is used

5.2 Application of the Velocity Controller

Figure 5 shows the simulation results when the initial conditions (0,0) and the reference signal $y_\omega = 5sin(0.5t)$ are used. The amplitude of this reference corresponds to 1000 revolutions per minute for angular velocity. In this case a proportional-integral

(PI) controller is used to compare the performance of fuzzy controller. The proportional gain is 0.9 and the integral gain is 2.4.

After 6 seconds both controller converge to reference signal. However, when the PI controller is used a lag phase is holded. The control signals of both controllers are showed in figure 6. The PI control signal and the fuzzy control signal are bounded. However, PI controller has a steady state error different to zero against the performance of fuzzy controller which converge to zero after 2 seconds.

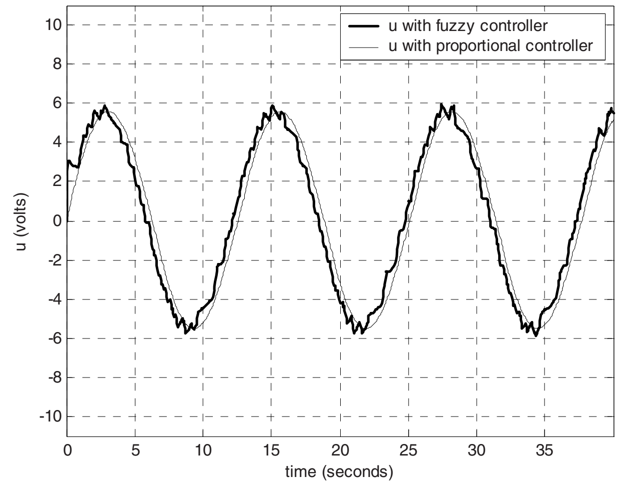


Fig. 6. The control signal u for angular velocity tracking when initial condition (0,0) is used

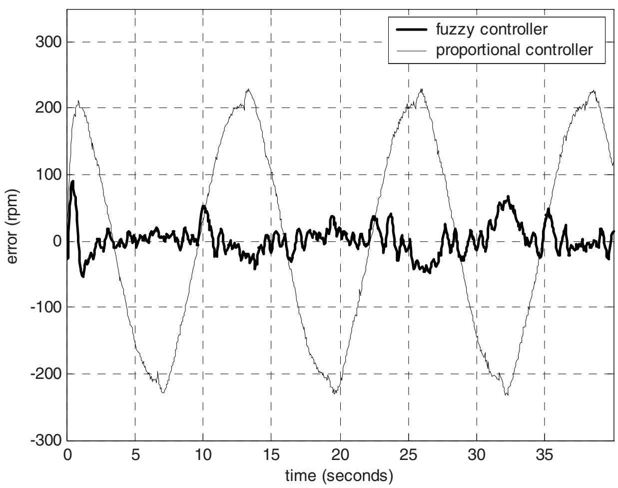


Fig. 7. The error signal for angular velocity tracking when initial condition (0,0) is used. The error signal for fuzzy controller is defined by $e = x_2 - y_\omega$ and the error signal for proportional controller is defined by $e = y_\omega - x_2$.

6 Conclusions

In this paper, controllers using *fuzzy Lyapunov synthesis approach* are designed and implemented to solve the trajectory tracking problem when signal references are used to determine the behavior of the angular position and velocity of a servo trainer. Real time results illustrate that the performance of the fuzzy controller is better than the performance using conventional controller suggested by the servo trainer designers. Fuzzy Lyapunov approach is an efficient methodology to sistematically derive the rule bases of fuzzy controllers, which are stable and improve the behavior of classical proportional and proportional-integral controllers. We are encouraged to extend this methodology considering different loads and perturbations to achieve new results in short time.

Acknowledgement. The authors thank support of UNACAR on project PR/61/2006.

References

1. Margaliot M., Langholz G.: Fuzzy Lyapunov-based approach to design of fuzzy controllers. Fuzzy Sets and Systems, Elsevier, No. 106, (1999) 49-59.
2. Margaliot M., Langholz G.: Design and Analysis of Fuzzy Schedulers using Fuzzy Lyapunov Synthesis. Third International Conference on Knowledge-Based Intelligent Inforation Engineering Systems, Adelaide, Australia (1999).
3. TecQuipment LTD: CE110 Servo Trainer, User's Manual, England (1993).
4. Margaliot M., Langholz G.: Adaptive Fuzzy Controller Design via Fuzzy Lyapunov Synthesis. Proceedings of IEEE World Congres on Computational Intellegence, Vol. 1, (1998), 354-359.
5. Margaliot M., Langholz G.: New approaches to fuzzy modeling and control: design and analysis, World Scientific, Singapore (2000).
6. Zhou C.: Fuzzy-Arithmetic-Based Lyapunov Synthesis in the Design of Stable Fuzzy Controllers: A Computing-With-Words Approach. Intenational Journal of Applied Mathematics and Computational Scients, Vol. 12, No. 3 (2002), 411-421.
7. Margaliot M., Langholz G.: Fuzzy control of benchmark problem: a computing with words approach, IEEE Transaction on Fuzzy Systems, Volume 12, Issue 2, (2004) 230-235
8. Cazarez N. R., Castillo O, Aguilar L. and Cardenas S.: Lyapunov Stability on Type-2 Fuzzy Logic Control, Proceedings of International Seminar on Computational Intelligence, Mexico D. F (2005) 32-41.
9. Mannani A., Talebi H. A.,: A Fuzzy Lyapunov Síntesis-Based Controller for a Flexible Manipulator: Experimental Results. Proceedings of the IEEE Conference on Control Applications, Toronto, Canada, (2005), 606-611.
10. Castillo O., Cazarez N., Melin P.: Design of Stable Type-2 Fuzzy Logic Controller based on a Fuzzy Lyapunov Approach. IEEE International Conference on Fuzzy Systems (2006).

Part XV

Neural Networks

A New Method for Intelligent Knowledge Discovery

Keith Douglas Stuart[1] and Maciej Majewski[2]

[1] Polytechnic University of Valencia, Department of Applied Linguistics
Camino de Vera, s/n, 46022 Valencia, Spain
KStuart@idm.upv.es
[2] Koszalin University of Technology, Faculty of Mechanical Engineering
Raclawicka 15-17, 75-620 Koszalin, Poland
Maciej.Majewski@tu.koszalin.pl

Abstract. The paper describes the application of an artificial neural network in natural language text reasoning. The task of knowledge discovery in text from a database, represented with a database file consisting of sentences with similar meanings but different lexico-grammatical patterns, was solved with the application of neural networks which recognize the meaning of the text using designed training files. We propose a new method for natural language text reasoning that utilizes three-layer neural networks. The paper deals with recognition algorithms of text meaning from a selected source using an artificial neural network. In this paper we present that new method for natural language text reasoning and also describe our research and tests performed on the neural network.

Keywords: Knowledge Discovery, Artificial Neural Networks, Natural Language Processing, Artificial Intelligence.

1 Introduction

For linguistic research, there is a need for consciously created and organized collections of data and information that can be used to carry out knowledge discovery in texts and to evaluate the performance and effectiveness of the tools for these tasks. Knowledge discovery in text is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in unstructured textual data [14,15]. These patterns are unknown, hidden or implicit in semi-structured and unstructured collections of text. Below are some of the kinds of knowledge discovery tasks that many subject disciplines are interested in:

- Identification and retrieval of relevant documents from one or more large collections of documents.
- Identification of relevant sections in large documents (passage retrieval).
- Co-reference resolution, i.e., the identification of expressions in texts that refer to the same entity, process or activity.
- Extraction of entities or relationships from text collections.
- Automated characterization of entities and processes in texts.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 721–729, 2007.
springerlink.com

- Automated construction of ontologies for different domains (e.g., characterization of medical terms).
- Construction of controlled vocabularies from fixed sets of documents for particular domains.

The need to construct controlled vocabularies for subject domains has meant that terminological extraction from corpora has become an important process in tasks related to knowledge discovery in text [14].

The proposed system for knowledge discovery in text uses neural networks for natural language understanding in Fig. 1.

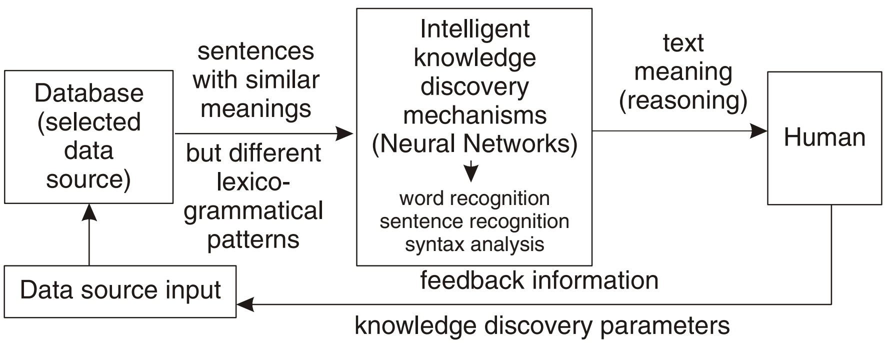


Fig. 1. Steps involved in proposed knowledge discovery in text

The system consists of a selected data source, 3-layer artificial neural networks, network training sets, letter chain recognition algorithms, syntax analysis algorithms, as well as coding algorithms for words and sentences.

2 The State of the Art

Knowledge discovery is a growing field: There are many knowledge discovery methodologies in use and under development. Some of these techniques are generic, while others are domain-specific.

Learning algorithms are an integral part of knowledge discovery. Learning techniques may be supervised or unsupervised. In general, supervised learning techniques enjoy a better success rate as defined in terms of usefulness of discovered knowledge. According to [1,2], learning algorithms are complex and generally considered the hardest part of any knowledge discovery technique. Machine discovery is one of the earliest fields that has contributed to knowledge discovery [4]. While machine discovery relies solely on an autonomous approach to information discovery, knowledge discovery typically combines automated approaches with human interaction to assure accurate, useful, and understandable results.

There are many different approaches that are classified as knowledge discovery techniques [16]. There are quantitative approaches, such as the probabilistic and statistical approaches. There are approaches that utilize visualization techniques. There are classification approaches such as Bayesian classification, inductive logic, data cleaning/pattern discovery, and decision tree analysis [2,4]. Other approaches include deviation and trend analysis, genetic algorithms, neural networks, and hybrid approaches that combine two or more techniques.

The probabilistic approach family of knowledge discovery techniques utilizes graphical representation models to compare different knowledge representations [7]. These models are based on probabilities and data independencies. The statistical approach uses rule discovery and is based on data relationships. An inductive learning algorithm can automatically select useful join paths and attributes to construct rules from a database with many relations [3]. This type of induction is used to generalize patterns in the data and to construct rules from the noted patterns.

Classification is probably the oldest and most widely-used of all the knowledge discovery approaches [3,7,16]. This approach groups data according to similarities or classes. There are many types of classification techniques e.g. the Bayesian approach, pattern discovery and data cleaning, and the decision tree approach.

Pattern detection by filtering important trends is the basis for the deviation and trend analysis approach. Deviation and trend analysis techniques are normally applied to temporal databases [4,6].

Neural networks may be used as a method of knowledge discovery. Neural networks are particularly useful for pattern recognition, and are sometimes grouped with the classification approaches. A hybrid approach to knowledge discovery combines more than one approach and is also called a multi-paradigmatic approach. Although implementation may be more difficult, hybrid tools are able to combine the strengths of various approaches. Some of the commonly used methods combine visualization techniques, induction, neural networks, and rule-based systems to achieve the desired knowledge discovery. Deductive databases and genetic algorithms have also been used in hybrid approaches.

2 Method Description

In the proposed knowledge discovery system shown in Fig. 2, sentences are extracted from the database. The separated words of the text are the input signals of the neural network for recognizing words [5]. The network has a training file containing word patterns. The network recognizes words as the sentence components, which are represented by its neurons in Fig. 3. The recognized words are sent to the algorithm for coding words [12]. Then, the coded words are transferred to the sentence syntax analysis module. It is equipped with the algorithm for analyzing and indexing words. The module indexes words properly and then they are sent to the algorithm for coding sentences [13]. The commands are coded as vectors and they are input signals of the sentence recognition module using neural networks. The module uses the 3-layer Hamming neural network in Fig. 4, either to recognize the sentence in order to find out its meaning or just does not recognize the sentence. The neural network is equipped with a training file containing patterns of possible sentences whose meanings are understood.

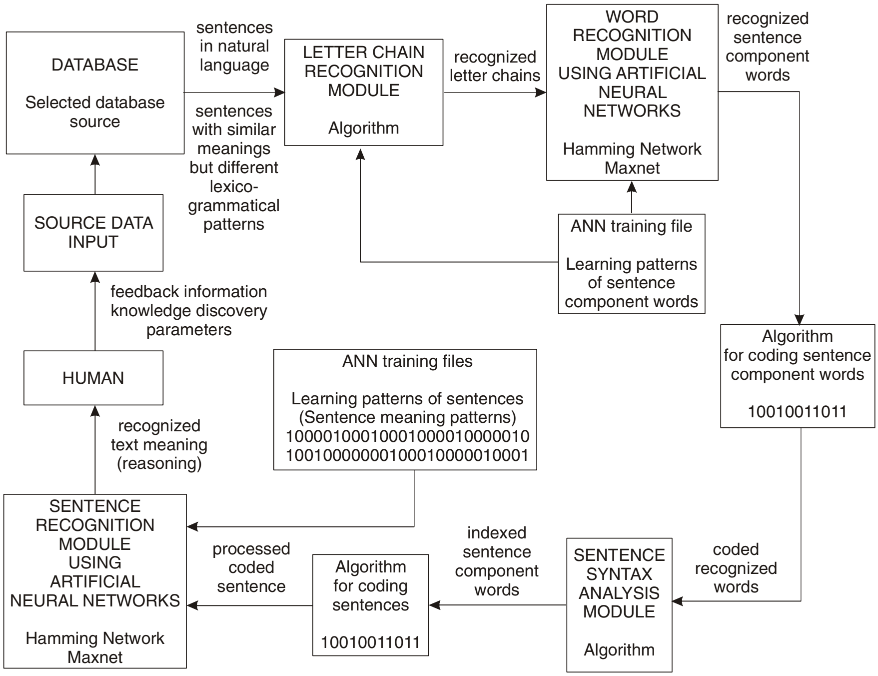


Fig. 2. Scheme of the proposed system for knowledge discovery in text

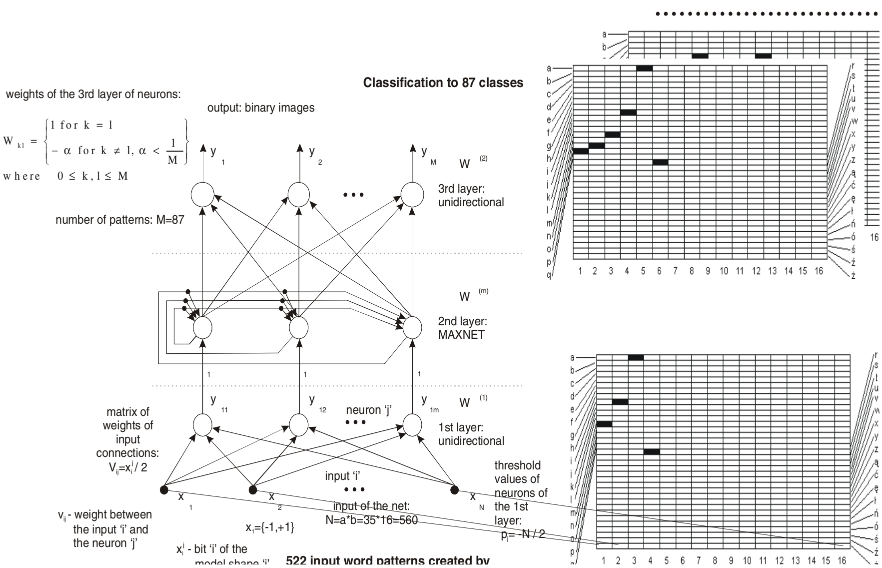


Fig. 3. Scheme of the 3-layer neural network for word recognition

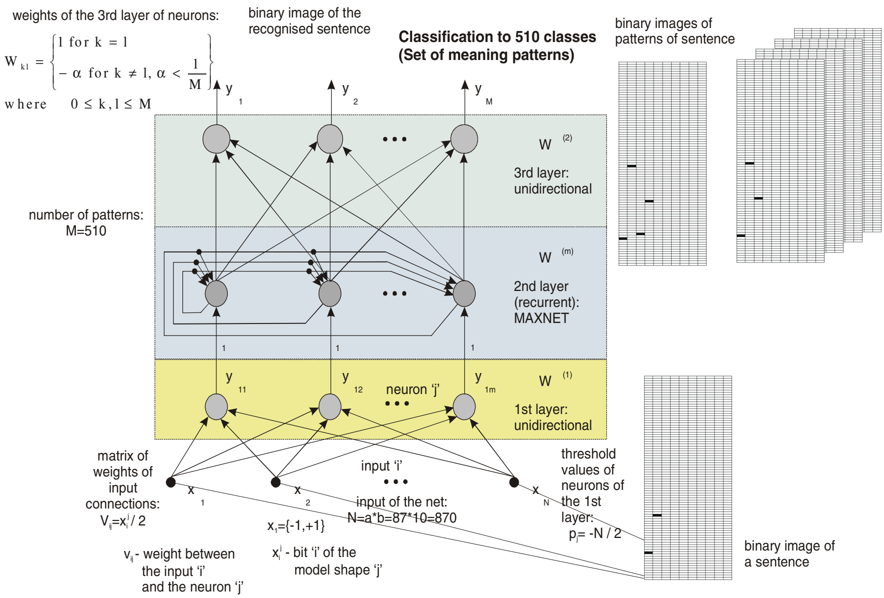


Fig. 4. Scheme of the 3-layer neural network for sentence recognition

Because of the binary input signals, the Hamming neural network is chosen (fig. **5**) which directly realizes the one-nearest-neighbour classification rule [9, 10, 11]. Each training data vector is assigned a single class and during the recognition phase only a single nearest vector to the input pattern x is found and its class C_i is returned. There are two main phases of the operation of the expert-network: training (initialization) and classification. Training of the binary neural network consists of copying reference patterns into the weights of the matrix W_{pn}, as follows (**1**):

$$w_i = x_i \ , \quad 1 \le i \le p \tag{1}$$

where p is the number of input patterns-vectors x, each of the same length n, w_i is the i-th row of the matrix W of dimensions p rows and n columns. For given n the computation time is linear with the number of input patterns p.

The goal of the recursive layer N_2. is selection of the winning neuron. The characteristic feature of this group of neurons is a self connection of a neuron to itself with a weight $m_{ii}=1$ for all $1 \le i \le p$, whereas all other weights are kept negative. Initialization of the N_2. layer consists in assigning negative values to the square matrix M_{pp} except the main diagonal. Originally Lippmann proposed initialization [8] (**2**):

$$m_{kl} = -(p-1)^{-1} + \xi_{kl} \quad for\ k \ne l, \quad 1\ for\ k = l$$
$$where\ 1 \le k, l \le p,\ p > 1 \tag{2}$$

where ξ is a random value for which $|\xi| \ll (p-1)^{-1}$. However, it appears that the most efficient and still convergent solution is to set equal weights for all neurons $N_{2.}$ which are then modified at each step during the classification phase, as follows (**3**):

$$m_{kl} = \varepsilon_k(t) = -(p-t)^{-1} \quad for\ k \neq l, \quad 1\ for\ k = l \qquad (3)$$
$$where\ 1 \leq k, l \leq p,\ p > 1$$

where t is a classification time step. In this case the convergence is achieved in p-1-r steps, where $r>1$ stands for the number of nearest stored vectors in W.

In the classification phase, the group $N_{1.}$ is responsible for computation of the binary distance between the input pattern z and the training patterns already stored in the weights W. Usually this is the Hamming distance (**4**):

$$b_i(z,W) = 1 - n^{-1} D_H(z, w_i), \quad 1 \leq i \leq p \qquad (4)$$

where $b_i \in [0,1]$ is a value of an i-th neuron in the $N_{1.}$ layer, $D_H(z, w_i) \in \{0,1,...,n\}$ is a Hamming distance of the input pattern z and the i-th stored pattern w_i (i-th row of W).

In the classification stage, the $N_{2.}$ layer operates recursively to select one winning neuron. This process is governed by the following equation (**5**):

$$a_i[t+1] = \varphi\left(\sum_{j=1}^{n} m_{ij} a_j[t]\right) = \varphi\left(a_i[t] + \sum_{j=1, i \neq j}^{n} m_{ij} a_j[t]\right) \qquad (5)$$

where $a_i[t]$ is an output of the i-th neuron of the $N_{2.}$ layer at the iteration step t, φ is a threshold function given as follows (**6**):

$$\varphi(x) = x \quad for \quad x > 0, \quad 0 \ \ otherwise \qquad (6)$$

Depending on the chosen scheme (2)-(3) of the m_{ij} weights in (5), we obtain different dynamics of the classification stage. The iterative process (5) proceeds up to a point where only one neuron has value different than 0 – this neuron is a winner.

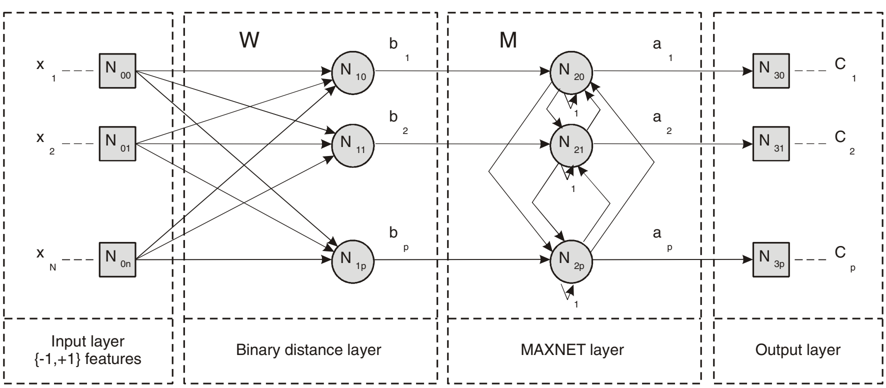


Fig. 5. Structure of the Hamming neural network as a classifier-expert module

3 Research Results

The dataset for the tests carried out contained a database of 1500 sentences, files consisting of 522 letter chains, 87 word training patterns and 510 sentence meaning training patterns.

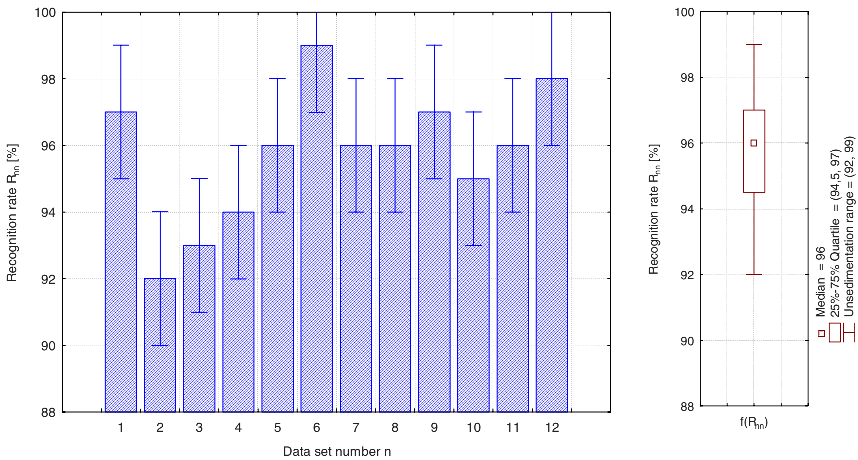


Fig. 6. Sentence meaning recognition rate as a set of words recognized earlier

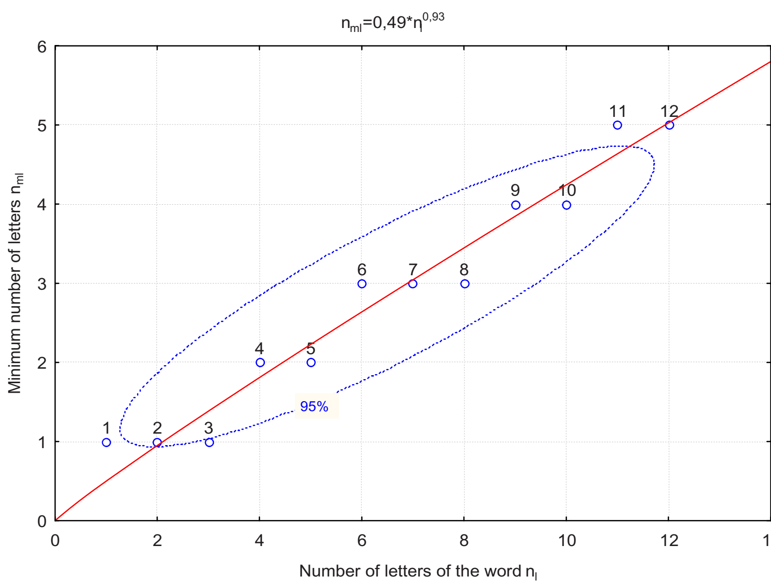


Fig. 7. Sensitivity of word recognition: minimum number of letters of the word being recognized to number of word component letters

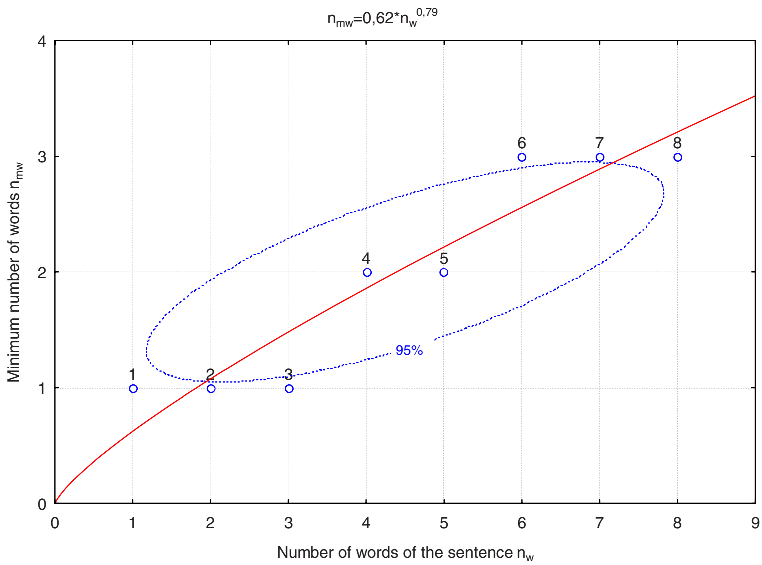


Fig. 8. Sensitivity of sentence meaning recognition: minimum number of words of the sentence being recognized to number of sentence component words

The first test measured the performance of the sentence meaning recognition with the sentence recognition module using artificial neural networks as a set of words recognized earlier in Fig. 6.

As shown in Fig. 7, the ability of the implemented neural network for word recognition to recognize the word depends on the number of letters. The neural network requires a minimum number of letters of the word being recognized as its input signals. As shown in Fig. 8, the ability of the neural network for sentence meaning recognition to recognize the sentence depends on the number of sentence component words. Depending on the number of component words of the sentence, the neural network requires a minimum number of words of the given sentence as its input signals.

4 Conclusions and Perspectives

Knowledge discovery is a rapidly expanding field with promise for great applicability. Knowledge discovery purports to be the new database technology for the coming years. The need for automated discovery tools had caused an explosion in research.

The motivation behind using the binary neural networks in knowledge discovery comes from the possible simple binarization of words and sentences, as well as very fast training and run-time response of this type of neural networks. Application of binary neural networks allows for recognition of sentences in natural language with similar meanings but different lexico-grammatical patterns, which can be encountered in documents, texts, vocabularies and databases. The presented methods can be easily extended.

It is anticipated that commercial database systems of the future will include knowledge discovery capabilities in the form of intelligent database interfaces. Some types of information retrieval may benefit from the use of knowledge discovery techniques. Due to the potential applicability of knowledge discovery in so many diverse areas there are growing research opportunities in this field.

References

1. Berry, M., Linoff, G.: Mastering Data Mining. John Wiley & Sons (2000)
2. Dunham, M. H.: Data Mining Introductory and Advanced Topics. Prentice Hall (2003)
3. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann (2006)
4. Hand, D. J., Mannila, H., Smyth, P.: Principles of Data Mining. MIT Press (2000)
5. Kacalak, W., Douglas Stuart, K., Majewski, M.: Intelligent Natural Language Processing. International Conference on Natural Computation ICNC2006. Xi'an, China. Lectures Notes in Artificial Intelligence 4221, Springer-Verlag, Berlin Heidelberg New York (2006) 584-587
6. Kantardzic, M.: Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-IEEE Press (2002)
7. Larose, D. T.: Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Inc. (2004)
8. Lippman, R.: An Introduction to Computing with Neural Nets. IEEE Transactions on Acoustic, Speech, and Signal Processing. ASSP-4, 3(1987). IEEE Signal Processing Society. Piscataway (1987) 4-22
9. Majewski, M., Kacalak, W.: Intelligent Layer of Two-Way Voice Communication of the Technological Device with the Operator. KES 2005 Melbourne, Australia. Lectures Notes in Artificial Intelligence 3683, Springer-Verlag, New York (2005) 930-936
10. Majewski, M., Kacalak, W.: Intelligent human-machine voice communication system. Engineering Mechanics International Journal for theoretical and applied mechanics 12(3) (2005) 193-200
11. Majewski, M., Kacalak, W.: Automatic recognition and evaluation of natural language commands. International Symposium on Neural Networks ISNN 2006. Chengdu, China. Lecture Notes in Computer Science 3973: Advances in Neural Networks, Springer-Verlag, Berlin Heidelberg New York (2006) 1155-1160
12. Majewski, M., Kacalak, W.: Natural language human-machine interface using artificial neural networks. International Symposium on Neural Networks ISNN 2006. Chengdu, China. Lecture Notes in Computer Science 3973: Advances in Neural Networks. Springer-Verlag, Berlin Heidelberg New York (2006) 1161-1166
13. Majewski, M., Kacalak, W., Douglas Stuart, K.: Selected Problems of Intelligent Natural Language Processing. International Journal of Research on Computing Science 20(2006) (ISSN 1665-9899)
14. Stuart, K.: Corpus Linguistics: theory, method and applications. XXIV International AESLA Conference, Language Learning, Language Use and Cognitive Modelling: Applied Perspectives across Disciplines. Madrid, Spain (2006)
15. Stuart, K.: A software application for text analysis. International Conference on Engineering Education, Valencia, Spain (2003)
16. Tan, P. N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson Addison Wesley (2005)

New Method of Learning and Knowledge Management in Type-I Fuzzy Neural Networks

Tahseen A. Jilani and Syed Muhammad Aqil Burney

Department of Computer Science, University of Karachi,
Karachi-74270, Pakistan
tahseenjilani@ieee.org, burney@computer.org

Abstract. A new method for modeling and knowledge extraction at each neuron of a neural network using type-I fuzzy sets is presented. This approach of neuron modeling provides a new technique to adjust the fuzzy neural network (FNN) structure for feasible number of hidden neurons and efficient reduction in computation complexity. Through repeated simulations of a crisp neural network, we propose the idea that for each neuron in the network, we can obtain reduced model with high efficiency using wavelet based multiresolution analysis (MRA) to form wavelet based fuzzy weight sets (WBFWS). Triangular and Gaussian membership functions (MFs) are imposed on wavelet based crisp weight sets to form Wavelet Based Quasi Fuzzy Weight Sets (WBQFWS) and Wavelet Based Gaussian Fuzzy Weight Sets (WBGFWS). Such type of WBFWS provides good initial solution for training in type-I FNNs. Thus the possibility space for each synoptic connection is reduced significantly, resulting in fast and confident learning of FNNs. It is shown that propsed modeling approach hold low computational complexity as compared to existing type-I fuzzy neural network models.

Keywords: extraction of fuzzy rules, fuzzy neural networks, neuro-fuzzy modeling Wavelet based multiresolution analysis.

1 Introduction

Fission of artificial neural networks [10] and fuzzy sets have attracted the growing interest of researchers in various scientific and engineering areas due to the growing need of adaptive intelligent systems to solve the real world problems. A crisp or fuzzified neural network can be viewed as a mathematical model for brain-like systems. The learning process increases the sum of knowledge of the neural network by improving the configuration of weight factors. An overview of different FNN architectures is discussed by [5] and [9]. It is much more difficult to develop the learning algorithms for the FNN than for the crisp neural networks. This is because the inputs, connection weights and bias terms related to a FNN are fuzzy sets; see [19] and [22].

The new technique in mathematical sciences called wavelets can be introduced to reduce the problem complexity as well as the dimensions so that a FNN may provide a fast track for optimization. Wavelet based MRA provides better analysis of complex signals than Fourier based MRA, see [7].

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 730–740, 2007.
springerlink.com

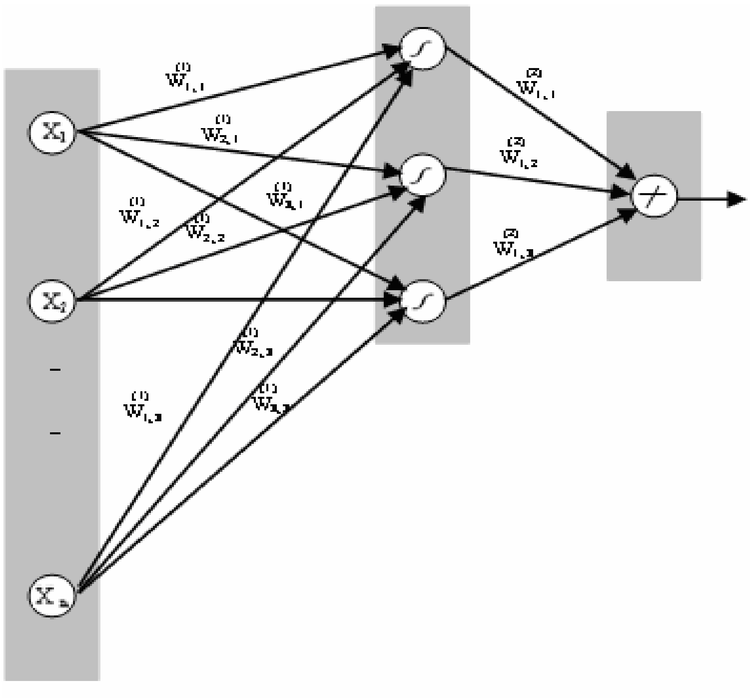


Fig. 1. Structure of a crisp neural network

Based on this approach of knowledge discovery for each synaptic connection, we can convert the probability concept of network connections into possibility perception. These sets provide the initial design for type-I neuro-fuzzy networks as discussed in [1], [3] and [11]. When jumbled with [20], this new approach assures lower computational complexity due to improved selection of seed values of the network. To our knowledge, the concept of obtaining WBFWS through crisp neural networks has not been investigated in the literature. The work is organized as follows. In section 2 we have briefly discussed wavelet based multiresolution analysis technique. In section 3, we have given new approach for fuzzy modeling of network connections. In section 4, simulation experiments are presented. To determine accuracy of WBFWS, a comparison is also made between proposed WBFWS and Gaussian confidence intervals for each hidden synaptic connection of the neural network. Finally, discussions and concussions are given in section 5.

2 Wavelet Based Multiresolution Analysis

In recent years, researchers have developed powerful wavelet techniques for the multi scale representation and analysis of signals see [6], [7] and [17]. These new methods differ from the traditional Fourier technique. Wavelets localize the information in the time-frequency space which makes them especially suitable for the analysis of non-stationary signals [15]. One important area of applications where wavelets have been found to be relevant is fuzzy neural systems is discussed in [9] and [14]. This whole area of research is still relatively new but is evolving very rapidly. We examine the very important property of wavelet transformation i.e. maximization of signal energy using data compression for FNNs. There are essentially two types of wavelet decompositions, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT), see [21]. Continuous wavelets are usually preferred for signal analysis, feature extraction and detection tasks where as the second type is obviously more adequate whenever it is desirable to perform some kind of data reduction or when the orthogonality of the representation is an important factor [7]. However, the choice between them is optional depending upon the computational considerations. We will use the decomposition in terms of DWT using Mallat's pyramid algorithm which is faster than a CWT and obtained very satisfactory results. Let $f(t)$ be a signal defined in $L^2(R)$ space, which denotes a vector space for finite energy signals, where R is a real continuous number system. The wavelet transformation of $f(t)$ in terms of continuous wavelets is then defined as

$$CWT_\psi f(a,b) = \langle f, \psi_{a,b} \rangle = \int_{\infty}^{\infty} f(t)\psi_{a,b}(t)\,dt$$
$$\text{where} \quad \psi_{a,b}(t) = |a|^{\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \tag{1}$$

$\psi(t)$ is the base function or the mother wavelet with $a, b \in R$ are the scale and translation parameters respectively. Instead of continuous dilation and translation, the mother wavelet may be dilated and translated discretely by selecting $a = a_0^{\ m}$ and $b = nb_0 a_0^{\ m}$, where a_0 and b_0 are fixed values with $a_0 > 1$, $b_0 > 0$, $m, n \in Z$ and Z is the set of positive integers. Then the discretized mother wavelet becomes

$$\psi_{m,n}(t) = a^{-\frac{m}{2}} \psi\left(\frac{t - nb_0 a_0^{\ m}}{a_0^{\ m}}\right) \tag{2}$$

and the corresponding discrete wavelet transform is given by

$$DWT_\psi f(m,n) = \langle f, \psi_{m,n} \rangle = \int_{-\infty}^{\infty} f(t)\psi_{m,n}(t)\,dt \tag{3}$$

DWT provides a decomposition and reconstruction structure of a signal using MRA through filter bank. The role of mother scaling and mother wavelet functions $\phi(t)$ and $\psi(t)$ are represented through a low pass filter L and a high pass filter H. Consequently, it is possible to obtain a signal f through analysis and synthesis by using wavelet based MRA [21].

$$f(t) = \sum_{n \in Z} c_{p,n} \phi_{p,n}(t) + \sum_{0 \le m \ge p} \sum_{n \in Z} d_{m,n} \psi_{m,n}(t) \tag{4}$$

where the sum with coefficients $c_{p,n}$ represents scaling or approximation coefficients and sums with coefficients $d_{m,n}$ represent wavelet or detail coefficients on all the scales between 0 and p. Data compression and energy storage in wavelets can be achieved by simply discarding certain coefficients that are insignificant. We combine this property of wavelets with neural networks and found a special class of mother wavelets db4, the most appropriate based on our data. We studied the effect of crisp weights on different neurons by reducing them using wavelets according to their energy preservation.

3 New Technique for Fuzzy Neural Network Synaptic Weights

Methods of fuzzy logic are commonly used to model a complex system by a set of rules provided by the experts [1]. To form WBQFWS from crisp neural network, we can obtain an appropriate unique MF by applying fuzzy aggregation operations. Such aggregation operations allow modeling each free parameter by combining multiple possible alternatives, see [16] and [23]. As in our proposed method, each synaptic fuzzy weight is obtained by ordered selection from crisp simulations, thus we can also impose ordered weighted aggregation operation (OWA) on re-simulated crisp neural networks. The quasi fuzzy weight sets follows all axioms of fuzzy aggregation functions like boundary condition, monotonicity, continuity, idempotency and strict buoyancy.

Let $h:[0,1]^n \to [0,1]$ denote the fuzzy aggregation from n dimensional space to a unique value given as $A(x) = h(A_1(x), A_2(x), ..., A_n(x))$ for each $x \in X$, where $A_i(x)$ are fuzzy sets defined over X. Thus a quasi-fuzzy set **A** is defined as

$$\min(a_1, a_2, ..., a_n) \le h(a_1, a_2, ..., a_n) \le \max(a_1, a_2, ..., a_n) \tag{5}$$

for all n-tuples $(a_1, a_2, ..., a_n) \in [0,1]^n$. We call $a_l = \min(a_1, a_2, ..., a_n)$ and $a_r = \max(a_1, a_2, ..., a_n)$. For middle parameter, we define simple averaging of quasi fuzzy sets as

$$h_\alpha(a_l, a_r) = \left(\frac{a_l^\alpha + a_r^\alpha}{2} \right)^{\frac{1}{\alpha}} \tag{6}$$

For simplicity, we have assumed $\alpha = 1$. For FNN with WBQFWS, we can also introduce non-symmetric fuzzy MFs by varying the parameter α.

4 Experiment

A crisp neural network with three hidden and one output neuron is trained and repeated the simulations for 700 times with an average rate of 7 simulations before a successful simulation. Through wavelet decomposition, we reduced dimensions by preserving 95% of the energy of original signal. The decomposed signal at level 5 using db4 wavelets for one of the input weight vectors is shown in Fig. 2(a), and its compressed version along with original signal in Fig. 2(b). A threshold of 5% is used for the compression of signals, thus reducing the data dimensionality up to 50%. So that in place of high data requirements of QFWS in [4], we obtained better performance using WBFWS, see [2]. From Fig. 1, the parameters of triangular MF based WBQFWS are given in Eq. (1) and Eq. (2). Although other MF can also be used depending upon the problem addressed.

As most of the actuarial problems are full of fuzziness [24], thus skewed MFs like Gamma and Beta can also be imposed. In Fig. 5 and Fig. 6, we have shown Gaussian MF based modeling for synaptic connections of a neural network.

In Table 1, results of WBQFWS provide superior mapping of weight space obtained using repeated simulations of crisp neural network than 95% Gaussian confidence interval. Interesting to note that the mean and standard errors are similar up to five decimal places in WBQFWS showing consistency of proposed interval sets but not in the case of Gaussian bounds. When each of the 100 simulated values of weights are validated for significance then we observe considerable differences in the prediction capacity of two types of interval sets. Similarly, the 95% WBGFWS also perform better then 95% Gaussian intervals. In validation process, the proposed WBFWS provide more accurate bound as compared to 95% Gaussian confidence intervals (less than 93% close sets) as show in Table 3 for WBQFWS and in Table 6 for WBGFWS.

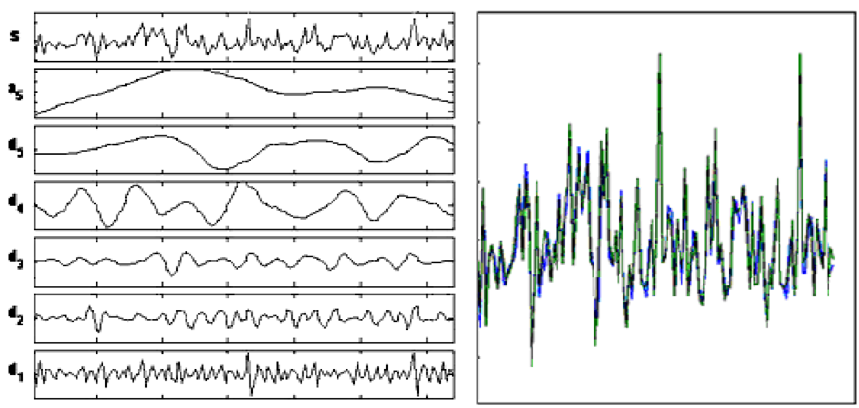


Fig. 2. (a) Signal decomposition (b) original and compressed signals

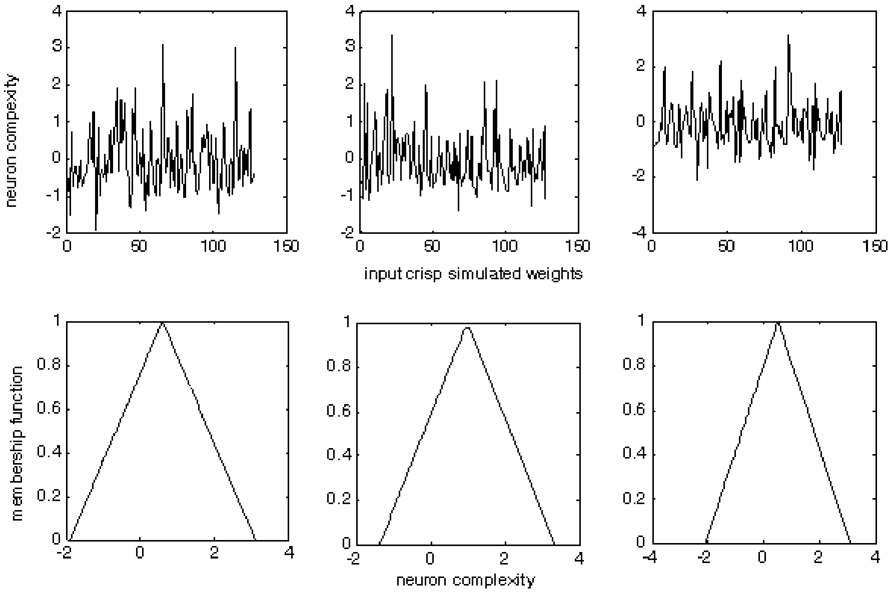


Fig. 3. Wavelet based weights and corresponding triangular MF

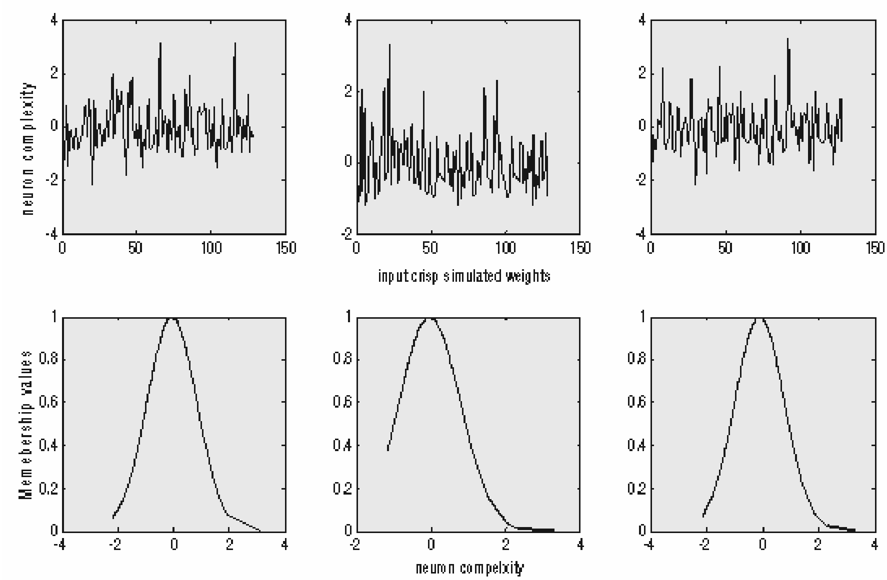


Fig. 4. (a) Input weight matrix for first input vector, (b) Corresponding Gaussian MF

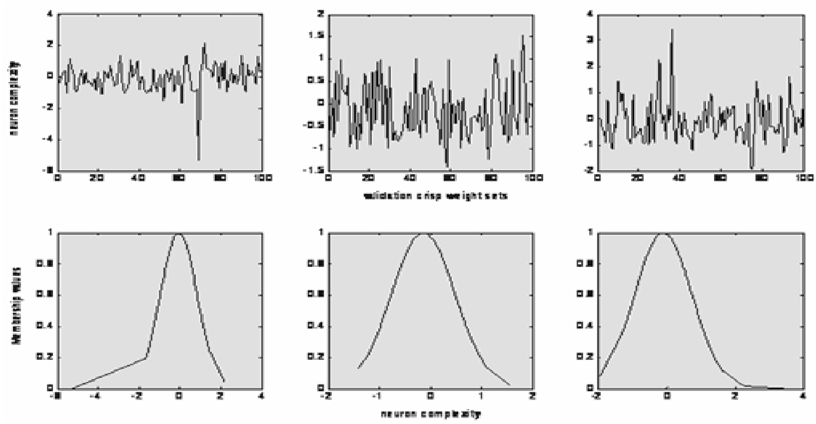


Fig. 5. (a) Validation weight matrix for first input vector, (b) Corresponding Gaussian MF

Table 1. A Comparison of WBQFWS and 95% Gaussian Confidence Intervals

Weight position (from Fig. 1)	(a) WBQFWS Min	Max	Mean	S. E.	(b) 95% Gaussian C. I. Confidence Bound		Mean	S. E.
$W_1(1,1)$	-1.92	3.09	-0.06	0.86	-1.52	1.40	-0.06	0.89
$W_1(2,1)$	-1.37	3.36	-0.04	0.80	-1.39	1.30	-0.04	0.82
$W_1(3,1)$	-2.09	3.13	-0.11	0.85	-1.56	1.33	-0.11	0.88
$W_2(1,1)$	-2.90	3.22	0.26	1.20	-1.82	2.33	0.25	1.26
$W_2(1,2)$	-1.98	3.13	0.19	1.09	-1.65	2.04	0.19	1.12
$W_2(1,3)$	-2.26	2.92	0.31	1.03	-1.45	2.07	0.31	1.07

Table 2. Validation of WBQFWS and 95% Gaussian Confidence Intervals

Weight position (from Fig. 1)	(a) WBQFWS Min	Max	Mean	S. E.	(b) 95% Gaussian C. I. Confidence Bound		Mean	S. E.
$W_1(1,1)$	-1.92	3.09	-0.06	0.89	-1.52	1.40	-0.06	0.89
$W_1(2,1)$	-1.37	3.36	-0.13	0.62	-1.16	0.88	-0.13	0.62
$W_1(3,1)$	-2.09	3.13	-0.09	0.83	-1.46	1.27	-0.09	0.83
$W_2(1,1)$	-2.90	3.22	-0.01	0.92	-1.52	1.52	-0.01	0.92
$W_2(1,2)$	-1.98	3.13	0.09	1.13	-1.77	1.97	0.09	1.13
$W_2(1,3)$	-2.26	2.92	0.39	1.12	-1.45	2.23	0.39	1.12

Table 3. Number of Insignificant WBQFWS and Gaussian Confidence Interval

Weight position (from Fig. 1)	WBQFWS	95% Gaussian C.I
$W_1(1,1)$	1	5
$W_1(2,1)$	1	2
$W_1(3,1)$	1	7
$W_2(1,1)$	0	7
$W_2(1,2)$	0	13
$W_2(1,3)$	0	13
Deficiency	0.43%	7.80%
Accuracy through validation	99.57%	92.20%

Table 4. A Comparison of WBGFWS with 95% Gaussian Confidence Intervals

Weight position (from Fig. 1)	(a) WBGFWS				(b) 95% Gaussian C. I.			
	Min	Max	Mean	S. E.	Confidence Bound		Mean	S. E.
$W_1(1,1)$	-1.92	3.09	-0.06	0.86	-1.52	1.40	-0.06	0.89
$W_1(2,1)$	-1.37	3.36	-0.04	0.80	-1.39	1.31	-0.04	0.82
$W_1(3,1)$	-2.09	3.13	-0.11	0.85	-1.56	1.3	-0.11	0.88
$W_2(1,1)$	-2.90	3.22	0.26	1.21	-1.82	2.36	0.25	1.26
$W_2(1,2)$	-1.98	3.13	0.19	1.09	-1.65	2.04	0.19	1.12
$W_2(1,3)$	-2.26	2.92	0.31	1.04	-1.45	2.07	0.31	1.07

Table 5. Validation of 95% WBGFWS and 95% Gaussian Confidence Intervals

Weight position (from Fig. 1)	(a) 95% WBGFWS				(b) 95% Gaussian C. I.			
	Confidence Bound		Mean	S. E.	Confidence Bound		Mean	S. E.
$W_1(1,1)$	-1.81	1.68	-0.06	0.89	-1.53	1.40	-0.06	0.89
$W_1(2,1)$	-1.36	1.08	-0.14	0.62	-1.16	0.89	-0.15	0.63
$W_1(3,1)$	-1.73	1.54	-0.09	0.83	-1.47	1.28	-0.09	0.84
$W_2(1,1)$	-1.81	1.82	-0.01	0.92	-1.53	1.52	-0.00	0.93
$W_2(1,2)$	-2.13	2.33	0.09	1.14	-1.77	1.97	0.09	1.14
$W_2(1,3)$	-1.80	2.59	0.39	1.12	-1.45	2.24	0.39	1.12

Table 6. Number of Insignificant WBGFWS and Gaussian Confidence Interval

Weight position (from Fig. 1)	95 % WBGFWS	95% Gaussian C.I
$W_1(1,1)$	2	5
W1(2,1)	1	2
$W_1(3,1)$	2	7
$W_2(1,1)$	1	7
$W_2(1,2)$	9	13
$W_2(1,3)$	8	13
Deficiency	3.80%	7.80%
Accuracy through validation	96.20%	92.20%

5 Conclusions

Learning with compressed wavelet neural networks using fuzzy weights is efficient and demonstrates much higher level of generalization and shorter computing time as compared to FNNs. We described the architecture of WBFWS based FNN that provide better initial search for synaptic weights. For WBQFWS, results showed that less than 1% chance of bound independent values is possible, thus providing above 99% accurate mapping, in comparison with Gaussian bounds that is below 93%.

In the following, we have presented some aspects of WBFWS regarding the computational complexity of learning in FNN,

1. In FNNs, one of the methods of learning is based on level sets where each fuzzy synaptic connection is divided into υ intervals that satisfy fuzzy arithmetic operations. When assuming υ level sets, we get 2υ parameters of the form

$$\left\{\left(w_L^{h_1}, w_R^{h_1}\right), \left(w_L^{h_2}, w_R^{h_2}\right), \ldots, \left(w_L^{h_\upsilon}, w_R^{h_\upsilon}\right)\right\}$$

Using boundary conditions, monotonicity and weak buoyancy conditions of fuzzy weights, we can write

$$w_L^{h_1} \leq w_L^{h_2} \leq \ldots. \leq w_L^{h_\upsilon} \leq w_R^{h_1} \leq w_R^{h_2} \leq \ldots \leq w_R^{h_\upsilon}$$

Thus in FNN with crisp inputs and fuzzy weights, we can define fuzzy MF for each synaptic connection using level sets using WBFWS. If a certain MF consists of k parameters, then total free parameters for a single fuzzy synaptic connection are $2k\upsilon$. For example in a FNN structure $10\times3\times1$ with triangular MF and 100 level sets contains $(10\times3\times3\times2\times100)+(3\times3\times2\times100)=19800$ free parameters to be tuned in each iteration. The computational complexity of even a very small FNN makes them nearly impossible to apply on small scale problems because due to very large number of free parameters of the network, see [8], [12] and [13]. We need huge amount of input data to minimize the performance function (usually Mean Square Error) and to control the degrees of freedom of the network, see [20]. Thus our proposed weight sets with limited possibility space gives faster and more reliable convergence of learning using level sets based FNNs.

2. With the help of our proposed WBFWS, we can decide the architecture of a FNN. If for any connection $w_L^h = w_R^h = 0$ then that connection can be considered as dead connection resulting in a reduced model and thus raising the speed of learning in level sets based FNNs.

6 Future Work

Further improved identification of suitable MF is possible by determining the underlying probability structure of synaptic connections of a crisp neural network using non-parametric statistics like Kernel estimation and learning. Thus based on this idea, we can form fuzzy inference systems with varying rules based on neuron complexity. This may provide new research directions to compare different WBFWS based FNNs. For most of the actuarial problems with non-negative limits, we can propose SANFIS (Skewed Adaptive Neuro-Fuzzy Inference System) to give new impression for actuaries towards this comparatively new field of modeling, forecasting and control. A comparison for most suitable wavelet and optimization algorithms with varying learning parameters is also possible. As future work we will extend this concept on multivariate kernel estimation techniques, and type-II fuzz logic systems as worked by [18].

References

1. Buckley, J.J., Hayashi, Y.: Neural Networks for Fuzzy Systems, Fuzzy Sets and Systems. 71(3) (1995) 265-276.
2. Burney, S.M.A., Jilani, T.A., Saleemi, A.: Approximate Knowledge Extraction Using MRA for Type-I Fuzzy Neural Networks. Proc. of IEEE-ICEIS, (2006) 464-469. Islamabad Pakistan.
3. Burney, S.M.A., Jilani, T.A., Saleemi, A.: Optimizing Neuronal Complexity Using Wavelet Based Multiresolution Analysis for Type-I Fuzzy Neural Networks. Proc. of 4th WSEAS-CIMMACS, (2005) 210-216. Miami, Florida, USA.
4. Burney, S.M.A., Jilani, T.A., Ardil, C.: Approximate Bounded Knowledge Extraction Using Type-I Fuzzy Logic. Proc. of the 5th World Enformatika Conference, WEC'05 (2005). Prague, Czech Republic.
5. Castellano, G., Fanelli, A.M.: Fuzzy Inference and Rule Extraction Using Neural Network. Neural Network World Journal. 3 (2000) 361-371.
6. Cohen, A.: Numerical Analysis of Wavelet Methods. Studies in Mathematics and Its Applications. Vol. 32. Elsevier publishers (2003)
7. Daubechies, I.: Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series. SIAM, Philadelphia (1992)
8. Duch, W.: Uncertainty of Data, Fuzzy Membership Functions, and Multilayer Perceptron. IEEE Trans. on Neural Networks, 16(1) 10-23.
9. Dunyak, J., Wunsch, D.: Training Fuzzy Numbers Neural Networks With Alpha-cut Refinement. Proc. of IEEE Int. Conf. on System, Man, Cybernetics. 1, (1997) 189-194.
10. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall USA (1999)
11. Hee, Y.Y., Chong, F.L., Zhong, B.L.: A Hierarchical Evolutionary Algorithm for Constructing and Training Wavelet Networks. Neural Computing & Applications. 10 (2002) 357-366
12. Ishibuchi, H., Nii, M.: Numerical Analysis of the Learning of Fuzzified Neural Networks from If-Then Rules. Fuzzy Sets and System. 120(2) (2001) 281-307.
13. Ishibuchi, H., Fujioka, R., Tanaka, H.: Neural Networks That Learn from Fuzzy If-Then Rules. IEEE Trans. on Fuzzy System. 1(2) (1993) 85-97
14. Jang, J.-S.R, Sun, C.-T, Mizutani, E.: Neuro-Fuzzy Logic and Soft Computing, a Computational Approach to Learning and Machine Intelligence. Practice-Hall, USA (2003)
15. Kaiser, G.: A Friendly Guide to Wavelets. Birkhauser Boston MA USA (1994)
16. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic, Theory and Applications. Prentice Hall, India (2005)
17. Mallat, S.G.: A Theory of Multiresolution Signal Decomposition: the Wavelet Representation. IEEE Trans. on Pattern Analysis and Machine Intelligence. 11(7) (1989) 674-693
18. Mendel, J.M.: Uncertainly Rule-Based Fuzzy Logic Systems, Introduction and new Directions. Prentice Hall, PTR NJ USA (2001)
19. Mitra, S., Hayashi, Y.: Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework. IEEE Trans. on Neural Networks. 11(3) (2000) 748–768
20. Park, S., Han, T.: Iterative Inversion of Fuzzified Neural Networks. IEEE Trans. on Fuzzy Systems. 8(3) (2000) 266-280.

21. Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis. Cambridge University Press (2002)
22. Puyin, L. Xing, H.: Efficient Learning Algorithms for Three-Layer Regular Feedforward Fuzzy Neural Networks. IEEE Trans. on Fuzzy Systems. 15(3) (2004) 545-558
23. Yager, R.R., Filev, D.P.: Essentials of Fuzzy Modeling and Control. John Wiley and Sons (2002)
24. Zimmerman, H.J.: Fuzzy Set Theory and Its Applications. Kluwer Publishers. Boston MA (2001)

Gear Fault Diagnosis in Time Domains by Using Bayesian Networks

Yuan Kang[1], Chun-Chieh Wang[2], and Yeon-Pun Chang[1]

[1] Department of Mechanical Engineering, Chung Yuan Christian University
yk@cycu.edu.tw
[2] R&D Center for Membrane Technology, Chung Yuan Christian University. 200, Chung Pei Rd., Chung Li 32023, Taiwan

Abstract. Fault detection in gear train system is important in order to transmitting power effectively. The artificial intelligent such as neural network is widely used in fault diagnosis and substituted for traditional methods. In rotary machinery, the symptoms of vibration signals in frequency domain have been used as inputs to the neural network and diagnosis results are obtained by network computation. However, in gear or rolling bearing system, it is difficult to extract the symptoms from vibration signals in frequency domain which have shock vibration signals. The diagnosis results are not satisfied by using artificial neural network, if the training samples are not enough. The Bayesian networks (BN) is an effective method for uncertain knowledge and less information in faults diagnosis. In order to classify the instantaneous shock of vibration signals in gear train system, the statistical parameters of vibration signals in time-domain are used in this study. These statistical parameters include kurtosis, crest, skewness factors etc. There, based on the statistical parameters of vibration signals in time-domain, the fault diagnosis is implemented by using BN and compared with two methods back-propagation neural network (BPNN) and probabilistic neural network (PNN) in gear train system.

1 Introduction

Gearboxes are widely used in rotary machinery in order to transmitting power. The definition of the gear fault diagnosis is differentiating faults from vibration signals by expert knowledge or experiences when the gear system broke down. The vibration signals always carry the important information in the machine. It is important to extract the symptoms of the vibration signals from accelerometers and use these symptoms to classify fault by artificial inference in the gear system. Although the most common method of fault diagnosis in rotary machinery is used frequency spectral analysis, it is difficult to extract the symptoms from vibration signals in frequency domain which have shock vibration signals in gear train system. Dyer and Stewart [1] use statistical analysis of vibration signals in time-domain to fault diagnosis in rolling element bearing. The symptoms of vibration signals in time-domain could display clearly for fault diagnosis in gear train system.

The traditional fault diagnosis may be influenced by human subjectivity and experience, and the weighting parameters also designed by human thought. Because the rules given by various experts are different, the diagnosis results are not consistent.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 741–751, 2007.
springerlink.com

Recently, BPNN is a kind of neural network which is widely used to solve fault diagnosis problems, which has been innovated by Rumelhart and McClelland [2] and developed to be a diagnosis method by Sorsa et al. [3]. Kang et al. [4] extracted frequency symptoms of vibration signals to detect faults of electric motor by using BPNN. However, it is hard to identify the faults by using frequency spectrum analysis in gear train system, because the components of frequency spectrum are complex and ambiguous. And when training samples are not enough, the diagnosis results are not accurate for diagnostic samples which are not existence within trained samples. The variations of vibration signals in time-domain are distinct to differentiate faults of gear train system.

Many artificial neural networks (ANNs) have been proposed before, however, the slow repeated iterative process and poor adaptation capability for data restrains the ANNs applications. An effective and flexible probabilistic neural network (PNN) could overcome these drawbacks. PNN presented in [5] is different from other supervised neural network, since the weightings don't alternate with training samples. The output values of PNN are obtained by once forward network computation. Lin et al. [6] used PNN to identify faults in dissolved gas content, while using the gas ratios of the oil and cellulosic decomposition to create training examples.

Based on Bayesian principle, the probability can estimate by prior probability of previous samples. This method is well used to these problems when the information is not enough. BN have proven useful for a variety of monitoring and predictive purposes. Applications have been documented mainly in the medical treatment [7], gas turbine engines [8], and industrial fault diagnosis [9]. Chien et al. [10] constructs a BN on the basis of expert knowledge and historical data for fault diagnosis on distribution feeder. On the other hand, Giagopulos et al. [11] use Bayesian statistical framework to estimate the optimal values of the gear and bearing model parameters and be achieved by combining experimental information from vibration measurements with theoretical information to build a parametric mathematical model.

The statistical parameters of vibration signals in time-domain including waveform, crest, impulse, allowance, kurtosis and skewness parameters are proposed by Heng and Nor [12]. Due to the statistical parameters could represent the explicit symptoms for gear train system and the classification for uncertainty information by using BN, in this study, combined these methods to diagnosis faults in gear train system. These six statistical parameters of vibration signals in time-domain are used for fault diagnosis with BPNN, PNN and BN methods. Also, BN results are compared with both methods BPNN and PNN in fault diagnosis of gear train system.

2 Statistical Parameters of Vibration

The six statistical parameters of time history have been defined by Heng and Nor [12] as follows:

(a) The waveform factor shows the indication of shift in time waveform and determined by

$$Waveform\ (X_W) = \frac{r.m.s\ value}{mean\ value} = \frac{X_{rms}}{\left|\overline{X}\right|} \tag{1}$$

Table 1. The training samples of statistical parameters of vibration signals

Training sample	Parameter / Test No.	Waveform	Crest	Impulse	Allowance	Kurtosis	Skew-ness
1	1	1.28748	4.16475	5.36206	6.42607	2.94202	-1.38656
	2	1.33577	4.42759	5.91424	7.23393	3.45039	-1.43746
	3	1.26744	4.90187	6.21281	7.36275	2.8115	-1.36702
	4	1.26649	5.35327	6.77986	8.01719	2.93045	-1.39855
	5	1.28831	4.25197	5.47787	6.55841	3.03643	-1.39166
	6	1.27132	5.16768	6.56977	7.80874	2.89777	-1.38012
	7	1.25515	4.72966	5.93644	6.99951	2.77261	-1.38922
	8	1.2625	6.26067	7.9041	9.34875	2.88097	-1.39203
	9	1.27562	4.9493	6.31342	7.50522	2.95553	-1.39151
	10	1.27298	5.63142	7.16872	8.52778	2.92079	-1.3766
2	1	1.27916	4.82934	6.17751	7.34722	3.03417	-1.39694
	2	1.30721	8.2694	10.80986	12.91118	5.376	-1.48271
	3	1.32963	10.52056	13.98841	16.84124	5.03708	-1.43339
	4	1.27883	5.11315	6.53884	7.79305	3.1635	-1.4335
	5	1.25876	6.35699	8.0019	9.45222	2.94453	-1.42286
	6	1.29752	10.30627	13.37261	15.88942	6.08271	-1.46424
	7	1.36039	11.89035	16.17552	19.47433	7.94966	-1.39051
	8	1.3236	9.8308	13.01206	15.63566	4.64462	-1.3523
	9	1.30498	9.86201	12.86971	15.36947	4.70865	-1.39747
	10	1.26212	5.49845	6.93971	8.19017	2.92291	-1.39796
3	1	1.68451	11.44801	19.28425	25.26632	14.76671	-0.99271
	2	1.68947	10.91075	18.43343	23.91418	16.44452	-0.95952
	3	1.70213	10.60673	18.05407	23.92433	14.38681	-1.01801
	4	1.58438	10.36547	16.42281	21.0378	11.10712	-1.05348
	5	1.53747	11.87461	18.2568	22.97949	12.0315	-1.10121
	6	1.7177	11.00257	18.89916	25.00289	15.44512	-1.07667
	7	1.63984	11.55684	18.95141	24.56436	14.65595	-1.05854
	8	1.54895	10.27717	15.91886	20.20572	9.62844	-1.08054
	9	1.60122	10.12086	16.20569	20.79312	11.47277	-0.99938
	10	1.67369	10.68908	17.89025	23.13652	16.18965	-0.99548

(b) The crest factor shows the indication of peak height in time waveform and determined by

$$Crest\ (X_C) = \frac{max\ peak}{r.m.s\ value} = \frac{max|X|}{X_{rms}} \quad (2)$$

(c) The impulse factor shows the indication of shock in time waveform and determined by

$$Impulse\ (X_I) = \frac{max\ peak}{mean\ value} = \frac{max|X|}{|\bar{X}|} \quad (3)$$

(d) The allowance factor shows the indication of plenty in time waveform and determined by

$$Allowance\ (X_A) = \frac{max\ peak}{X_r} = \frac{max|X|}{X_r},\ X_r = \left((1/N)\sum_{n=1}^{N} x(n)^{1/2}\right)^2 \quad (4)$$

where $x(n)$ is time waveform of vibration signal.

(e) The skewness and kurtosis factors are both sensitive indicators for the shape of the signal, which are relative to third and fourth moment of signal distribution in time-domain, respectively. The skewness factor corresponds to the moment of third order norm of vibration signal, which is determined by

$$Skewness\ (X_S) = \frac{M_3}{\sigma_3} \quad (5)$$

$$M_3 = (1/N)\sum_{n=1}^{N}(x(n) - \bar{X})^3,\ \sigma = ((1/N)\sum_{n=1}^{N}(x(n) - \bar{X})^2)^{1/2} \quad (6)$$

where M_3 is the third order moment and σ is standard deviation. The kurtosis factor corresponds to the moment of fourth order norm of vibration signal, which is determined by

$$Kurtosis\ (X_K) = \frac{M_4}{\sigma_4},\ M_4 = (1/N)\sum_{n=1}^{N}(x(n) - \bar{X})^4 \quad (7)$$

where M_4 is the fourth order moment. For kurtosis parameter, the values are 1.5 and 1.8 for sine and triangle wave signals in time history, respectively; for crest parameter, the values are 1.7 and 1.0 for triangle and square wave signals in time history,

Table 2. The diagnostic samples of statistical parameters of vibration signals

Diagnostic sample	Parameter / Test No.	Waveform	Crest	Impulse	Allowance	Kurtosis	Skewness
1	1	1.28083	4.9557	6.34742	7.57411	2.92901	-1.37241
	2	1.27573	4.24116	5.41058	6.45568	2.86223	-1.38532
	3	1.2797	4.25527	5.44549	6.50587	2.91428	-1.39213
	4	1.25417	3.88207	4.86876	5.75323	2.73299	-1.39766
	5	1.25061	4.49917	5.62672	6.62779	2.65127	-1.36925
2	1	1.28311	7.37059	9.45725	11.26441	3.39816	-1.41329
	2	1.31473	7.68456	10.10314	12.19807	3.79101	-1.43346
	3	1.29129	6.41499	8.28359	14.92663	3.12193	-1.38309
	4	1.29454	6.84103	8.85602	10.61262	3.44279	-1.42621
	5	1.29993	7.1357	9.27591	11.12575	3.66325	-1.45471
3	1	1.36765	11.41049	19.02875	24.67334	15.31862	-0.95438
	2	1.53771	10.51913	16.17538	20.65024	8.64434	-1.0645
	3	1.54835	11.44328	17.71818	22.44819	11.37889	-1.12111
	4	1.62486	11.48614	18.66335	24.16117	12.91246	-1.06472
	5	1.60378	11.81417	18.94734	24.71282	10.43235	-1.05344

respectively. Although the simple signals could be checked easily with theoretical values of statistical parameters, the vibration signals of gear train system is complex and could not checked for different gear faults.

In this study, the training samples are obtained by simulating corresponding gear faults on experimental rotor-bearing system. Besides the normal gear (O_1) signals, there are two kinds of faults which including tooth breakage fault (O_2) and wear fault (O_3). There are ten training samples for each fault and is listed in Table 1. Each training sample have six statistical parameter of vibration signals which including waveform, crest, impulse, allowance, kurtosis and skewness parameters. Similarly, there are five diagnostic samples for each fault and is listed in Table 2.

3 Bayesian Networks

According to Bayes' theorem [13] for a hypothesis H and an evidence E been giving the conditional (posterior) probability can be determined by

$$P(H|E)=\frac{P(E|H)\times P(H)}{P(E)} \tag{8}$$

where $P(H|E)$ is the posterior probability for H giving the information that the evidence E is true, $P(H)$ is the prior probability of H, and $P(E|H)$ is the probability for evidence E giving the information that the hypothesis H is true.

Perrot et al. [14] used Gaussian probability distribution law to conditional probability density function. The equation of a Gaussian probability law generalized to n dimensions is

$$P(E_i|H_j)=\frac{1}{(2\pi)^{i/2}(det\,\Gamma[j])^{1/2}}exp[-\frac{1}{2}(X_i-E_j)\Gamma[j]^{-1}(X_i-E_j)^T] \tag{9}$$

where X_i and E_j are i-component row vector for statistical parameters and mean values for j-th category. $\Gamma[j]$ is the $i\times i$ covariance matrix for j-th category. In this study, a Bayesian network is designed with six statistical parameters ($i=1,2,\ldots,6$) and three gear faults($j=1,2,3$). The equation of covariance matrix $\Gamma[j]$ as follow:

$$\Gamma[j]=\begin{bmatrix} L_{11} & \cdots & L_{1n} \\ L_{21} & \cdots & L_{2n} \\ & & \\ L_{m1} & \cdots & L_{mn} \end{bmatrix},\quad L_{mn}=\sum_{k=1}^{10}x_m^{(j)}(k)\bullet x_n^{(j)}(k)-\frac{(\sum_{k=1}^{10}x_m^{(j)}(k)\bullet x_n^{(j)}(k))^2}{10} \tag{10}$$

where $m=n=1,2,\ldots,6$. L_{mn} is the covariance value between x_m and x_n. Three probability density values $P(E|H_j)$ are obtained by Eq. (9) with statistical parameters of diagnostic sample. The max probability density value $P(E|H_j)$ represents the decision classification belongs to j-th category. And the posterior for j-th category as follow:

$$P(H_j|E)=\frac{P(E|H_j)}{\sum_{j=1}^{3}P(E|H_j)} \tag{11}$$

4 Back-Propagation Neural Network

A BPNN [3] is designed to fault diagnosis in gear train system which is composed by six input nodes, forty hidden nodes and three output nodes. The inputs and outputs variables are arrayed into a vector as express by

$$(X_W, X_C, X_I, X_A, X_K, X_S, O_1, O_2, O_3) \tag{12}$$

where six elements of inputs represent waveform (X_W), crest (X_C), impulse (X_I), allowance (X_A), kurtosis (X_K) and skewness (X_S) factors, and three elements of outputs including normal gear (O_1), tooth breakage fault (O_2) and wear fault (O_3).

Since these statistical parameters have been used as inputs to the BPNN which used sigmoid function to activated function, the inputs and outputs of diagnostic network can be normalized into a range from 0 to 1 and are classified into several levels except for the skewness factor which is normalized into a range from -1 to 0. Too many levels are chosen caused the redundant computation, contrariwise, too few levels are chosen caused the ambiguous classification. In order to obtain the explicit classification, in this study, these statistical parameters are classified into five levels, one fifth, two fifths, three fifths and four fifths of total sample can be designated to be level limits of very small(VS), small(S), medium(M), large(L) and very large(VL) levels. Thus, the five levels of membership functions are expressed in Table 3. Thus, there are thirty neurons(F_i) in fuzzy layer.

Table 3. The membership function of statistical parameters of vibration signals in time-domain

Level / Parameter	VS (**1** 0 0 0 0)	S (0 **1** 0 0 0)	M (0 0 **1** 0 0)	L (0 0 0 **1** 0)	VL (0 0 0 0 **1**)
Waveform	$X_W<1.2665$	$1.2665 \le X_W<1.2749$	$1.2749 \le X_W<1.2942$	$1.2942 \le X_W<1.3296$	$X_W \ge 1.3296$
Crest	$X_C<4.8293$	$4.8293 \le X_C<5.6314$	$5.6314 \le X_C<7.9098$	$7.9098 \le X_C<10.1209$	$X_C \ge 10.1209$
Impulse	$X_I<6.1474$	$6.1474 \le X_I<7.1687$	$7.1687 \le X_I<10.1717$	$10.1717 \le X_I<13.3726$	$X_I \ge 13.3726$
Allowance	$X_A<7.2594$	$7.2594 \le X_A<8.5278$	$8.5278 \le X_A<12.1927$	$12.1927 \le X_A<15.8894$	$X_A \ge 15.8894$
Kurtosis	$X_K<2.8810$	$2.8810 \le X_K<2.9820$	$2.9820 \le X_K<3.5942$	$3.5942 \le X_K<5.3926$	$X_K \ge 5.3926$
Skewness	$X_S<-1.4205$	$-1.4205 \le X_S<-1.3920$	$-1.3920 \le X_S<-1.3847$	$-1.3847 \le X_S<-1.3215$	$X_S \ge -1.3215$

The connections between the fuzzy and hidden layers, the hidden and output layers are weighting coefficient w_{ij} and w_{jk}, respectively. The input and output of the j-th neuron in the hidden layer respectively can be expressed by

$$net_j = \sum_{i=1}^{30} w_{ij} F_i\,,\ O_j = f(net_j) = 1/(1+exp(-net_j)) \tag{13}$$

where F_i is the i-th fuzzy neuron, w_{ij} is the weighting coefficient between the i-th neuron in the fuzzy layer and the j-th neuron in the hidden layer. $f(\cdot)$ is the sigmoid function adopted to be the activation function. The output of BPNN can be expressed by

$$net_k = \sum_{j=1}^{m} O_j w_{jk}\ ,\ k=1,2,3,\ O_k = f(net_k) = 1/(1+exp(-net_k)) \tag{14}$$

where m is the neuron number in the hidden layer, w_{jk} is the weighting coefficient between the j-th neuron in the hidden layer and the k-th neuron in the output layer.

The BPNN is trained using the error between the actual output and the ideal output, to modify w_{ij} and w_{jk} until the output of BPNN is close to the ideal output with an acceptable accuracy. On the basis of the gradient descent method for the minimization of error, the correction increments of weighting coefficients are defined to be proportional to the slope, related to the changes between the error estimator and the weighting coefficients as

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}, \ \Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} \tag{15}$$

where η is the learning rate used for adjusting the increments of weighting coefficients and controls the convergent speed of the weighting coefficients. E is the error estimator of the network and is defined by

$$E = \frac{1}{2}\sum_{p=1}^{N} E_p = \frac{1}{2}\sum_{p=1}^{N}\sum_{k=1}^{3}\left(T_k - O_k\right)_p^2 \tag{16}$$

where N is the total number of training samples. $(T_k)_p$ is the ideal output of the p-th sample and $(O_k)_p$ is the actual output of the p-th sample. Substituting equation (16) into equation (15) and executing derivations give the increments of weighting coefficients as

$$\Delta w_{jk} = \eta \delta_k O_j = \eta \cdot f'\left(net_k\right) \cdot \left(T_k - O_k\right)_p \cdot O_j \tag{17}$$

for the output layer, and

$$\Delta w_{ij} = \eta \delta_j O_i = \eta \cdot w_{jk} \cdot \delta_k \cdot f'(net_j) \cdot O_i \tag{18}$$

where $f'(\cdot)$ is the first derivation of $f(\cdot)$.

5 Probabilistic Neural Network

A PNN [5] is designed to fault diagnosis in gear train system and the input and output nodes are the same as the BPNN. PNN is different with other supervised neural network, since the weighting coefficients values don't alternate with training samples. Because the numbers of the hidden nodes represents the training samples, the connections between the input and hidden layers are initial weighting coefficient w_{ij} which indicates the symptom sets of the training samples. The connections between the hidden and output layers are weighting coefficient w_{jk} which indicates the fault sets of the training samples. The input of the j-th neuron in the hidden layer can be expressed by

$$net_j = \sum_{i=1}^{6}(X_i - w_{ij})^2, \ j=1,2\ldots,30 \tag{19}$$

where X_i is the i-th input which including waveform (X_W), crest (X_C), impulse (X_I), allowance (X_A), kurtosis (X_K) and skewness (X_S) factors, sequentially. w_{ij} is the weighting coefficient between the *i*-th input in the input layer and the *j*-th neuron in the hidden layer. The output of the *j*-th neuron in the hidden layer can be obtained by

$$O_j = exp(-net_j/2\sigma^2) \quad (20)$$

where σ is the smoothing parameter of gauss function. The output of PNN can be expressed by

$$net_k = \frac{1}{N_K}\sum_{j=1}^{30} O_j w_{jk}\ , k=1,2,3 \quad (21)$$

where w_{jk} is the weighting coefficient between the *j*-th neuron in the hidden layer and the *k*-th neuron in the output layer, and also indicates the fault sets of the training samples. Then, the output O_k of the PNN can be obtained by

$$net_k = \max_r net_r,\ O_k = 1\ , r=1,2,3 \quad (22)$$

6 Case Studies

6.1 Experiment Set-Up

The experiment equipment of gear train system is shown in Figure 1. It consists of a motor, a converter and a pair of spur gears in which the transmitting gear has 46 teeth and the passive gear has 30 teeth. The vibration signals are measured in vertical direction from two accelerometers that mounted on the bearing housing of gear train system.

6.2 Neural Network Diagnosis

When the BPNN accomplish training, using the six statistical parameters and normalized by membership function to compute with the trained weighting coefficients w_{ij} and w_{jk} and obtained output values. Each value represents a kind of gear fault and the output value represents the degree of certainty for corresponding fault. If the value is closed to 1, which represent the possibility of the fault is high.

In Table 4, the diagnosis results by using BPNN are obtained with the trained weighting coefficients w_{ij} and $w_{jk.}$ For fifteen diagnostic samples, there are two and one wrong diagnosis results by using BPNN and PNN, respectively. Although, there are several advantages of artificial neural network such as the weightings of neural network are obtained from neural computation and the diagnosis results by using artificial neural network are objective than traditional expert's experiences, the diagnosis results are not accurate completely when the diagnostic samples are not within the range of trained samples.

Because the membership function is designed with subjective experiences, the diagnosis results by using BPNN with membership function could not obtained the accuracy results completely.

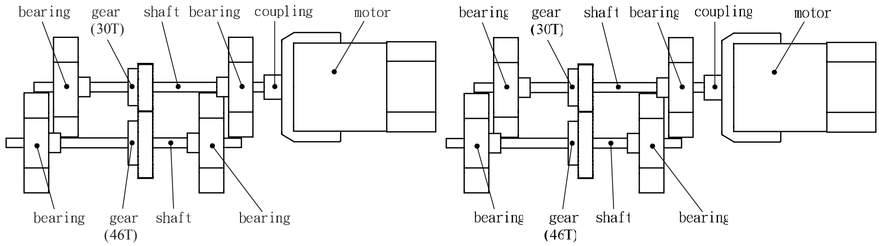


Fig. 1. The experimental setup of gear train system

Table 4. The certainty of diagnosis result due to BN, BPNN and PNN

Diagnostic sample	Faults / Test No.	O_1			O_2			O_3		
		BN	BPNN	PNN	BN	BPNN	PNN	BN	BPNN	PNN
1	1▲	0.942	0.088	1	0.058	0.943	0	0	0.133	0
	2	0.971	0.582	1	0.029	0.068	0	0	0.177	0
	3	0.965	0.527	1	0.035	0.042	0	0	0.055	0
	4	0.982	0.971	1	0.018	0.030	0	0	0.031	0
	5	0.977	0.958	1	0.023	0.499	0	0	0.022	0
2	1▲	0.229	0.935	0	0.771	0.110	1	0	0.008	0
	2	0.109	0.047	0	0.891	0.830	1	0	0.001	0
	3*	0.001	0.230	1	0.999	0.525	0	0	0.013	0
	4	0.305	0.255	0	0.695	0.579	1	0	0.026	0
	5	0.145	0.270	0	0.855	0.575	1	0	0.006	0
3	1	0	0.029	0	0	0.024	0	1	0.981	1
	2	0	0.029	0	0	0.024	0	1	0.981	1
	3	0	0.029	0	0	0.024	0	1	0.981	1
	4	0	0.029	0	0	0.024	0	1	0.981	1
	5	0	0.029	0	0	0.024	0	1	0.981	1

BN: Bayesian networks, BPNN: Back-propagation neural network, PNN: Probabilistic neural network.
▲: wrong result by using BPNN, ∗: wrong result by using PNN.

6.3 Bayesian Networks Diagnosis

According to Eq. (9), the diagnosis results by using BN are listed on Table 4. In BN classifier, for the three kinds of diagnostic samples, the posterior probabilities for gear faults are calculated and the decision is made for the maximum one. There is no wrong diagnosis result by using BN. It is accurate and easy to classify each gear fault for diagnostic samples by using BN method than both methods BPNN and PNN.

7 Conclusions

This study used six statistical parameters of vibration signals in time-domain to fault diagnosis in gear train system. The diagnosis results by using BN are better accurate than BPNN and PNN in Table 4.

Besides, it spends little time to obtain the diagnosis results than BPNN and PNN, because the connections weightings between inputs and outputs in BN are probability computation which differs from BPNN and PNN. Thus, the fault diagnosis results in gear train system by using BN not only represent high accuracy but also have fewer calculations than both methods BPNN and PNN.

Acknowledgements. This study was supported by grant number NSC92-2623-7-033-003-ET from National Science Council and by the Center-of-Excellence Program on Membrane Technology from Ministry of Education, Taiwan, R.O.C.

References

1. Dyer, D., Stewart, R. M.: Detection of Rolling Element Bearing Damage by Statistical Vibration Analysis. Journal of Mechanical Design 100 (1978) 229-235
2. Rumelhart, D. E. and McClelland, J. L.: Parallel Distributed processing, Explorations in the Microstructure of Cognition, MIT Press (1986)
3. Sorsa, T., Koivo, H. N., Koivisto, H.: Neural Networks in Process Fault Diagnosis. IEEE Transactions on Systems 21 (1991) 815-825
4. Kang, Y., Wang, C. C., Chang, Y. P., Hsueh, C. C., Chang, M. C.: Certainty Improvement in Diagnosis of Multiple Faults by Using Versatile Membership Functions for Fuzzy Neural Networks. Lecture Notes in Computer Science 3973 (2006) 370-375
5. Specht, D.: Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. IEEE International Conference on Neural Networks (1988) 525-532
6. Lin, W. M., Lin, C. H., Tasy, M. X.: Transformer-Fault Diagnosis by Integrating Field Data and Standard Codes with Training Enhancible Adaptive Probabilistic Network. IEE Proceedings Generation, Transmission and Distribution 152 (2005) 335-341
7. Kahn Jr., C. E., Laur, J. J., Carrera, G. F.: A Bayesian Network for Diagnosis of Primary Bone Tumors. Journal of Digital Imaging 14(2001) 56-57
8. Mast, T. A., Reed, A. T., Yurkovich, S., Ashby, M., Adibhatla, S.: Bayesian Belief Networks for Fault Identification in Aircraft Gas Turbine Engines. IEEE Conference on Control Applications 1(1999) 39-44
9. Romessis, C., Mathioudakis, K.: Bayesian Network Approach for Gas Path Fault Diagnosis. Journal of Engineering for Gas Turbines and Power 128(2006) 64-72
10. Chien, C. F., Chen, S. L., Lin, Y. S.: Using Bayesian Network for Fault Location on Distribution Feeder. IEEE Transactions on Power Delivery 17 (2002) 785-793
11. Giagopulos, D., Salpistis, C., Natsiavas, S.: Effect of Non-Linearities in the Identification and Fault Detection of Gear-Pair Systems. International Journal of Non-Linear Mechanics 41 (2006) 213-230

12. Heng, R. B. W., Nor, M. J. M.: Statistical Analysis of Sound and Vibration Signals for Monitoring Rolling Element Bearing Condition. Applied Acoustics 53(1998) 211-226
13. Berger, J. O.: Statistical Decision Theory and Bayesian Analysis, 2nd edition, New York, Springer-Verlag (1985)
14. Perrot, N., Trystram, G., Le, G. D., Guely, F.: Sensor Fusion for Real Time Quality Evaluation of Biscuit During Baking. Comparison Between Bayesian and Fuzzy Approaches. Journal of Food Engineering 29(1996) 301-315

Improving a Fuzzy ANN Model Using Correlation Coefficients*

Yanet Rodriguez[1], Bernard De Baets[2], Maria M. Garcia Ricardo Grau[1], Carlos Morell[1], and Rafael Bello[1]

[1] Universidad Central de Las Villas, Carretera a Camajuani km 51/2, Santa Clara, Cuba
{yrsarabia,mmgarcia,rgrau,rbellop}@uclv.edu.cu
[2] Ghent University, Coupure links 653, B-9000 Gent, Belgium
Bernard.DeBaets@UGent.be

Abstract. An improvement of a fuzzy Artificial Neural Network model based on correlation coefficients is presented. As aggregation operator to compute the net input to target neuron a uninorm is used, instead of the sum of all the influences that the neuron receives usually used in typical artificial neurons. Such combination allows increasing the model performance in problem solving. While the natural framework and the interpretability presented in the former model are preserved by using fuzzy sets, experimental results show the improvement can be accomplished by using the proposed model. Significant differences of performance with the previous model in favor of the new one, and comparable results with a traditional classifier were statistically demonstrated. It is also remarkable that the model proposed shows a better behavior in presence of irrelevant attributes than the rest of tested classifiers.

1 Introduction

Neuro-Fuzzy computing constitutes one of the best-known visible hybridizations encompassed in soft computing, capturing the merits of both fuzzy set theory and ANN. There exist various ways to perform this integration and many recent papers demonstrate their usefulness for developing efficient methodologies and algorithms for handling real life ambiguous situations efficiently [1]. Although most of the works reported use a framework to include expert knowledge in the form of fuzzy rules, the direct use of representative cases might be a useful approach as well.

Artificial Neural Networks (ANN) and Case-Based Reasoning (CBR) are two approaches of Artificial Intelligence that use examples stored in a case base. A hybrid model combining these approaches was presented in [2]. The associative ANN is used for suggesting the value of the target attribute for a given query. Afterwards, the case-based module justifies the solution provided by the ANN using a similarity function, which incorporates the ANN weights. Recently, an extension of this hybrid model modeling predictive attributes in terms of fuzzy sets was presented in [3]. The topology and learning of the associative ANN are based on representative values, which

* This work was supported in part by VLIR (Vlaamse InterUniversitaire Raad, Flemish Interuniversity Council, Belgium) under the IUC Program VLIR-UCLV.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 752–761, 2007.
springerlink.com

are considered fuzzy sets in case of numeric attributes. Later, the similarity function uses the ANN weights and the membership degrees in the fuzzy sets to justify the solution given by the fuzzy ANN model.

The fuzzy ANN model in [3] computes the ANN weights based on the notion of Relative Frequency extended to a fuzzy environment. When linguistic terms are used, a more interpretative and natural framework to include expert knowledge is provided than using the crisp model; while accuracy and robustness are preserved. Although, this ANN model can be used for information retrieval, the accuracy of its problem solver should be improved in order to obtain more comparable performance in classification task with respect to other classifiers.

On the other hand, it is very common in statistical analysis of data to find the correlation between variables or attributes. Correlation Coefficients (CC) can be used as distance functions. For example, in [4] two measures of correlation to use in learning systems are mentioned. Besides, its application has been extended to a fuzzy environment based on Zadeh's extension principle. For example, in [5] this value between two fuzzy sets is achieved in the interval [-1,1], while in [6] a methodology to calculate the CC as a fuzzy number is proposed.

This paper presents an improvement of the fuzzy ANN model used in the referred hybrid model, based on correlation coefficients. More specifically, Pearson's correlation coefficient is used to compute the ANN weight between two neurons in the fuzzy ANN model. But, while the Relative Frequency used before only takes values in the interval [0,1], the CC takes values in the interval [-1,1]. Afterwards, how the ANN model now aggregates influences from predictive attributes by using uninorms is proposed too. The following section gives an overview of the fuzzy ANN model using relative Frequency. Section 3 explains in details the new model based on CC and uninorms, while experimental results and their discussion are shown in section 4. Conclusions and future work are presented in section 5.

2 Overview of Fuzzy Neural Net Based on Relative Frequency

Fuzzy-SIAC is the problem solver used in the fuzzy CBR-ANN model proposed in [3]. It is a fuzzy implementation of the Simple Interactive Activation and Competition model exhibited in [2], capturing the merits of both fuzzy set theory and this ANN model. Numeric attributes are modeling using fuzzy sets.

Let *CB* denotes a case base and $A = \{a_1, a_2, .., a_m\}$ the set of m attributes that describe an example e of *CB*. Each example e_k of *CB* is denoted by $e_k = (e_{1k}, e_{2k}, .., e_{mk})$, but if the index k is immaterial then an example will be written as $e = (e_1, e_2, .., e_m)$. Let Da_i be the set of values that takes the attribute a_i for each e of *CB*, and later a finite universe Da_1x Da_2 x.. x Da_m (attribute space) is defined for a case base. If the attribute is numeric, representative values are linguistic terms, otherwise they are linguistic labels.

The Fuzzy-SIAC topology includes a group of neurons for each attribute a, where a single neuron is allocated for each representative value A defined for an attribute a. Let e_a be the value given to attribute a in the example e. Later, the value $f_A(e_a)$ is a measure of "how close" the value e_a is represented by the representative value A . It is a measure between 0 and 1 depending on expressions (1) or (2). When the attribute is numeric, expression (1) is used. Expression (2) is considered for symbolic attributes. The expression $\mu_A(x)$ denotes the membership degree of the value x to the linguistic term A.

$$f_A(x) = \mu_A(x) \tag{1}$$

$$f_A(x) = \begin{cases} 1 & , if \ x = A \\ 0 & , otherwise \end{cases} \tag{2}$$

Only relationships between neurons of different groups are considered. Each arc has associated a weight based on the examples in the case base. An extension of the Relative Frequency to fuzzy modeling is used. Let a and b be two attributes that describe an example e (or instance) of a training set (or case base CB). Let A and B representative values of the attributes a and b respectively. Further, the value labeled by $w_{A,B}$ represents the weight associated to the direct arc between representative values A and B, which is computed as follows:

$$w_{A,B} = \frac{\sum_{e \in CB} f_A(e_a) f_B(e_b)}{\sum_{e \in CB} f_A(e_a)} \tag{3}$$

Before presenting a query q to the trained ANN the set of attributes should be divided into two sets, the set of predictive attributes P and the set of target attributes. Let t be a target attribute and T one of its representative values. The Fuzzy-SIAC model completes the pattern corresponding to the query q taking into account the degree of activation of the neurons in the group of the target attribute t as follows:

$$Net_T(q) = \sum_{a \in P} \sum_{A_i \in R_a} w_{A_i,T} f_{A_i}(q_a) \tag{4}$$

Note that this value is obtained by summing the weighted activations from the neurons corresponding with predictive attributes. Afterwards, the representative value T associated with the node of greatest activation will be the value assigned to this attribute.

3 Fuzzy Neural Net Based on Correlation Coefficients

Correlation coefficients are mathematical forms of measuring the degree of association or intensity between two random variables. So for example, the linear correlation coefficient, or Pearson's coefficient, measures the strength of a possible linear relationship between two continuous variables x and y. Spearman's correlation coefficient is a non-parametric alternative to Pearson's correlation coefficient, valid when these variables have an ordinal nature.

All of these coefficients measure the strength of the linear dependence between the ranked values of the variables. Other coefficients, such as Kendall's Tau, also valid for ordinal variables, try to measure if pairs of cases in which an increase fo x is associated to an increase or decrease of y predominate, etc. But all of these measures are lacking of any sense of causality (they do not express cause-effect relationships) and in fact, the majority of correlation coefficients have a totally symmetrical character [7].

After representative values for all attributes are selected in the Fuzzy-SIAC model, a random variable associated to each representative value can be defined. In other words, a random variable associated with the representative value *A* taking the value $f_A(e_a)$ in the interval [0,1] for each example *e* of *CB* is defined. Thus, it is possible in general to use correlation coefficients between membership degrees (continuous variables derived of fuzzy variables), between membership degrees and binary variables (nominal variables previously dichotomized) and between two of such binary variables.

3.1 Learning of Fuzzy-SIAC Based on Correlation Coefficients

The weight between two neurons of an ANN gives a measure of the strength of the connection between them. Both models Fuzzy-SIAC [3] and the crisp one [3] are associative neural nets, where a Hebbian-like learning mechanism [8] is followed. In this type of learning the modification of the weights takes place based on the states of the excited neurons after the presentation of certain stimulus (information of entrance to the network), without considering whether it was desired to obtain these states of activation or not.

It is appropriate to use the CC as a crisp value between -1 and 1 for the calculation of the ANN weights in the fuzzy ANN model. The strength of the relationship is given by the magnitude, whereas the sign only indicates whether a variable tends to increase (positive sign) or decrease (negative sign) the other one. The values -1 and 1 mean that the variables in the sample have linear relationships, and the value zero indicates that such a relationship does not exist. Besides, it is not dimensional; which implies that the same result can be obtained although any linear transformation of the variables simulating an scale change has been done. Particularly, CC is independent of the choice of the origin for the variables. If the variables are independent, the CC is zero. It should be big when the degree of association between the variables is high and small when it is low.

Correlation coefficients can be extended to a fuzzy environment. More specifically, Pearson's correlation coefficient can be used. The new fuzzy ANN model will be referred to as the Fuzzy-SIACcc. Since CC is symmetrical, that is to say, it measures linear independence without presetting a variable as independent; Fuzzy-SIACcc will have undirected arcs between neurons. Positive values of $w_{A,B}$ indicate a tendency that $f_A(e_a)$ and $f_B(e_b)$ in the training set evolve in the same direction, while negative values of $w_{A,B}$ indicate the opposite. The closer the value of $|w_{A,B}|$ is to 1, the stronger is the tendency.

Let *f* be a function defined by using either expressions (1) or (2) according to the attribute type. The new way to compute the weight $w_{A,B}$ in Fuzzy-SIACcc is proposed according to the expressions (5) and (6) where the value *n* is the number of examples *e* of *CB* where the attributes *a* and *b* have known values.

$$w_{A,B} = \frac{\sum_{i=1}^{n}\left[f_A(x_i) - \overline{f_A}(x)\right]\left[f_B(x_i) - \overline{f_B}(x)\right] \Big/ (n-1)}{S_A S_B} \tag{5}$$

$$S_A^2 = \frac{\sum_{i=1}^{n} \left[f_A(x_i) - \overline{f_A}(x) \right]}{n-1} \tag{6}$$

3.2 Computing the Value for a Target Attribute

The value of the net input in typical artificial neurons is obtained as the sum of all the influences that the neuron receives. Nevertheless, while weights based on fuzzy Relative Frequency used before only take values in the interval [0,1], the CC takes values in the interval [-1,1]. It is not semantically the same to add two positive or two negatives values (these cases indicate the same tendency), than a positive and a negative value. In the latter case a simple additive operation is not the most adequate one.

The influence of neurons associated with predictive attributes in the proposed model should be aggregated by regarding several points of view. Thus, it can be possible to use uninorms [9],[10] as the aggregation operator instead of the arithmetic sum. Uninorms generalize both t-norms and t-conorms, having a neutral element lying anywhere in the unit interval. For example, in [11] a special kind of aggregation is used to handle the certainty in the MYCIN expert system. More specifically, the uninorm (on [-1,1]) proposed in expression (7) was used [11]. Additionally, its use to combine the shapes of single-variable functions as threshold functions in artificial neurons is proposed too.

$$U(x,y) = \begin{cases} x + y - xy & \textit{if} \;\; x, y \geq 0, \\ x + y + xy & \textit{if} \;\; x, y < 0, \; |x|,|y| \leq 1 \\ (x+y)/(1-\min(|x|,|y|)) & \text{otherwise} \end{cases} \tag{7}$$

Note that the previous definition is not clear in the points (-1,1) and (1,-1), though it is understood that the result is -1.

Finally, net input in the Fuzzy-SIACcc model is based on the concept of *strength* defined as follows:

Definition 1. The *strength* of the value of predictive attribute a in an object x with the representative value T for the target attribute t, is a measure of the activation received by the processing neuron N_T only considering this predictive attribute:

$$S(x_a, T) = \underset{A_i \in R_a}{U} \; w_{A_i,T} f_{A_i}(x_a) \tag{8}$$

This means that all of the influences the neuron receives from a predictive attribute p attempt to move the neuron from its current activation, as if there is a feedback connection of the neuron to itself. But, this influence is hidden in the aggregation of all influences depending on their direction.

Later on, the use of the uninorm defined in (7) is closely related with the meaning of this concept. In other words, when the new influences and the current activation have the same purpose (the same sign), the new activation is the result of applying the probabilistic sum t-conorm (the new activation excites the neuron). Otherwise, the activation is modified in the manner previously specified where the sign of the result depends on the highest value and the new activation inhibits the processing neuron.

Note that since the uninorm ***U*** is associative, it can be extended to some number of arguments in an unambiguous way, i.e. to a number of predictive attributes, as follows:

$$\underset{a_i \in P}{U} S(x_{a_i}, T) = U(S_T^{a_1}, U(S_T^{a_2}, U(S_T^{a_3}, \ldots))) \\ where: S_T^{a_i} = S(x_{a_i}, T) \tag{9}$$

Likewise, strengths received from all predictive attributes are aggregated using the same aggregation operator. Later, expression (4) used in Fuzzy-SIAC to achieve the net input to a representative value *T* for the target attribute *t* is replaced in the proposed model by the following expression:

$$Net_T(q) = \underset{a \in P}{U} S(x_a, T) \tag{10}$$

Finally, like in the former model, the representative value corresponding to the node of the greatest net input will be the value assigned to the target attribute for the corresponding query *q*.

4 Experimental Results and Discussion

Some experiments have been presented to show the improvement of the fuzzy ANN model due to the modifications explained above. Several datasets from UCIMLR [12] having both numeric and symbolic predictive attributes, one target attribute and no missing values were used. The percentage of cases correctly classified (accuracy) and the mean absolute error (error) were considered as measures of the performance. Each experiment was repeated ten times for each dataset using 10-fold cross validation [4].

Trapezoidal membership functions to represent numeric attributes using fuzzy sets were considered. Each numeric attribute is discretized using the supervised discretization method of Fayyad and Irani[13], allowing to define the corresponding membership function. Membership function parameters are set guaranteeing that two consecutive functions have as cut point the limit of the interval and each individual fuzzy set has the support equivalent to the half of the interval.

4.1 Testing the New Fuzzy ANN Model

Table 1 shows the performance obtained on these datasets using the fuzzy ANN model, with both Relative Frequency and Pearson's correlation coefficient to compute the ANN weights.

Note that the general behavior measured as the average value (last row of Table 1) is higher for accuracy and lower for error in the model proposed. Additionally, Fuzzy-SIACcc model shows better performance since it obtained higher classification accuracy on 10 of 17 datasets and lower mean absolute error on almost every dataset considered. On the other hand, if only the absolute value of weights would be considered in order to use arithmetic sum as previous model, FuzzySIACcc would has 46% as

Table 1. Results obtained with Fuzzy-ANN models

Dataset name	*Fuzzy-SIAC model*		*Fuzzy-SIACcc*		*Dataset characteristics*	
	accuracy (%)	*error*	*accuracy (%)*	*error*	*Number of cases*	*Number of classes*
Iris	**94.87**	0.24	93.13	**0.05**	150	3
Diabetes	65.11	0.44	**74.52**	**0.25**	768	2
Glass	53.53	0.23	**63.67**	**0.12**	214	7
LiverDBupa	**57.98**	0.49	44.63	0.55	345	2
Vehicle	60.53	0.35	**61.04**	**0.19**	846	4
Wine	**97.08**	0.35	91.17	**0.06**	178	2
Ionosphere	81.36	0.37	**88.06**	**0.12**	351	2
Sonar	53.38	0.49	**60.90**	**0.39**	208	2
Vowel	**57.74**	0.16	50.27	**0.09**	178	11
Segment	**80.75**	0.19	77.78	**0.06**	2310	7
Kr-vs-kp	61.25	0.49	**75.00**	**0.25**	3196	2
Hayes-roth	77.29	0.38	**80.15**	**0.13**	132	3
Contac-lenses	64.33	0.33	**86.83**	**0.09**	24	3
Monks-1	**75.00**	0.47	**75.00**	**0.25**	124	2
Monks-2	**67.14**	**0.44**	51.85	0.48	169	2
Monks-3	**98.89**	0.44	97.22	**0.03**	122	2
Zoo	61.23	0.18	**94.84**	**0.01**	101	7
Average	71.03	0.35	**74.47**	**0.18**		

accuracy's average. Thus, the use of uninorms as aggregation operator is significantly useful in the improved obtained.

4.2 Statistical Validation

The C4.5 algorithm [14], a traditional classifier based on decision trees, was included in the experiments; in order to compare the model proposed with an alternative approach traditionally used to solve classification task. The general behavior of accuracy by using C4.5 was 81%. The mean absolute error produced by C4.5 on these datasets is showed in the second column of Table 2.

On the other hand, the learning of the classifier is inherently determined by the attribute values. In theory, more attributes should provide more discriminating power, but in practice, excessive attributes will confuse the learning algorithm. Afterwards, attribute selection was carried out in order to test the behavior of the model proposed without irrelevant attributes. After that, experiments were repeated showing the mean absolute in the Table 2.

A method named Subset Evaluator [15] was used, searching in the space of all feature subsets the one that maximizes a numeric criterion. In other words, it evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. The last columns of the table above displayed the attribute's reduction occurred after applying this method.

Statistical analysis comparing paired samples were carried out using the mean absolute error computed both prior and after feature selection in each model considered.

Table 2. Results of the models without irrelevant attributes

Dataset name	*C4.5 error*	*Mean absolute error after Attribute Selection*			*Number of attributes*	
		C4.5	*Fuzzy-SIAC*	*Fuzzy-SIACcc*	*Original*	*Relevant*
Iris	0.04	0.04	0.13	0.04	4	2
Diabetes	0.31	0.32	0.42	0.23	8	4
Glass	0.12	0.12	0.23	0.12	9	8
LiverDBupa	0.38	0.46	0.48	0.55	6	1
Vehicle	0.14	0.17	0.00	0.21	18	11
Wine	0.05	0.05	0.34	0.04	13	11
Ionosphere	0.11	0.11	0.35	0.14	34	14
Sonar	0.27	0.21	0.48	0.41	60	19
Vowel	0.04	0.04	0.16	0.08	13	7
Segment	0.01	0.01	0.17	0.06	19	7
Kr-vs-kp	0.01	0.11	0.46	0.11	36	7
Hayes-roth	0.19	0.22	0.36	0.12	4	3
Contac-lenses	0.14	0.23	0.23	0.24	4	1
Monks-1	0.23	0.36	0.45	0.28	6	4
Monks-2	0.46	0.47	0.47	0.49	6	3
Monks-3	0.12	0.12	0.39	0.07	6	3
Zoo	0.02	0.02	0.17	0.02	16	10
Average	0.16	0.18	0.31	0.18		

Table 3. Results of Wilcoxon Test (confidence)

Test Statistics[a]

			error_fsiaccc - error_fsiacfr	error_C4.5 - error_fsiaccc
Monte Carlo Sig. (2-tailed)	Sig.		.000	.225
	99% Confidence Interval	Lower Bound	.000	.209
		Upper Bound	.000	.241

a. Based on 10000 sampled tables with starting seed 334431365.

Results of Table 3 indicate a significant improvement of the model proposed in this paper with respect to the former model (significance of the test 0.00), while a comparable behavior in comparison to the C4.5 algorithm on these datasets is shown (significance of the test 0.225).

On the other hand, the statistical analysis (see Table 4) proved that after selecting the relevant attributes, a significant improvement on the result is obtained with both Fuzzy-SIAC model and C4.5 algorithm. Nevertheless, the comparison of the same measure in the Fuzzy-SIACcc does not show significant differences (test significance: 0.491). That is to say, there are not enough elements to conclude the same as the rest of models; indicating that the model proposed would be tolerant to irrelevant features because the knowledge encoded in its weights is related with the relevance or redundancy of this connection in order to predict a target attribute.

Table 4. Statistical results prior and after selection of irrelevant attributes

Test Statistics[a,b]

			error_fsiacfr after - error_fsiacfr	error_fsiaccc after - error_fsiaccc	error_c45 after - error_C4.5
Monte Carlo Sig. (2-tailed)	Sig.		.002	.491	.045
	99% Confidence Interval	Lower Bound	.001	.469	.038
		Upper Bound	.004	.514	.053

a. Wilcoxon Signed Ranks Test

b. Based on 10000 sampled tables with starting seed 1556559737.

5 Conclusions and Future Work

This paper has presented an improvement of a fuzzy ANN model based on correlation coefficients. Linguistic terms are used to represent numeric attributes in the associative ANN model like in the previous one. The new fuzzy ANN model uses the Pearson's correlation coefficient to achieve the ANN weights, instead of the relative frequency criterion used in the former one. Secondly, a new operator to aggregate the weighted activations of the neurons corresponding to predictive attributes is applied. Afterwards, when a query is presented, the influences of predictive attributes are now aggregated by using a uninorm operator instead of the simple arithmetic sum.

Experimental results showed the improvement that can be accomplished by using the proposed model compared to the previous one, statistically demonstrating significant differences of the mean absolute error. Additionally, while the natural framework and the interpretability presented in the former model are preserved by using fuzzy sets; comparable results with the C4.5 algorithm are shown by the model presented. On the other hand, the model suggested behaves well in presence of irrelevant features whereas the rest of the models had significantly improved with attribute selection.

As a future work, the experimental evaluation should be extended to datasets with missing values. Additionally, the new way used to compute the activation of the neuron associated with a target attribute can be considered in the case based module of the former model to justify the solution provided by the fuzzy ANN.

References

1. Pal S. K., Dillon T. S., Yeung D. S. (Eds). Soft Computing in Case Based Reasoning. Springer-Verlag London (2001)
2. García, MM. and Bello, P.R.: A Model and its different applications to Case-Based reasoning. Knowledge-based systems 9 (1996) 465-473
3. Rodríguez, Y., Garcia, M., De Baets, B., Bello, R. and Morell, C.: Extending a Hybrid CBR-ANN Model by Modeling Predictive Attributes using Fuzzy Sets. Lecture Notes on Artificial Inteligence 4140, Springer Verlag (2006) 238-248
4. Wilson, D. R, and Martinez T. R.: Improved Heterogeneous Distance Functions. Journal of Artificial Intelligence Research, vol. 6, no. 1, (1997) 1-34
5. Shiang, D. and Lin, Nancy P.: Correlation of fuzzy sets. Fuzzy Sets and Systems 102 (1999) 221-226
6. Shiang-Tai, L. and Chiang, K.: Fuzzy measures for correlation coefficient of fuzzy numbers. Fuzzy Sets and Systems 128 (2002) 267-275
7. Siegel S, Castellan N.J.: Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York. (1988)
8. Sima, J. : Neural Expert System. Neural Networks. Vol 8 #2 (1995)
9. De Baets, B. and Fodor, J.: Residual Operators of Uninorms. Soft Computing 3 (1999) 89-100
10. Yager, Ronald R. and Rybalov, A.: Uninorm Agregation Operators. Fuzzy Sets and Systems 80 (1996) 111-120
11. Tsadiras, A.K and Margaritis, K.G.: The MYCIN certainty factor handling function as uninorm operator and its use as a threshold function in artificial neurons. Fuzzy Sets and Systems 93 (1998) 263-274
12. Murphy, P.M., Aha, D.W.: UCI Repository of Machine Learning Databases, http://www.ics.uci.edu/~mlearn/mlrepository.htm
13. Fayyad, U. M. and Irani, K. B. Multi-interval discretization of continuous-valued attributes for classification learning. *IJCAI-93*, (1993) 1022-1027
14. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA. (1993)
15. Hall, M.A., and Smith, L.A. Practical feature subset selection for machine learning. Proceedings of the 21st ACSC. Springer. (1998) 181-191

A Sliding Mode Control Using Fuzzy-Neural Hierarchical Multi-model Identifier

Ieroham Baruch, Jose-Luis O. Guzman, Carlos-Roman Mariaca-Gaspar, and Rosalba Galvan Guerra

CINVESTAV-IPN, Department of Automatic Control, Av. IPN No2508, A.P. 14-740, 07360 Mexico D.F., Mexico
{baruch,lolivares,cmariaca,rgalvan}@ctrl.cinvestav.mx

Abstract. A Recurrent Trainable Neural Network (RTNN) with a two layer canonical architecture learned by a dynamic Backpropagation learning algorithm is incorporated in a Hierarchical Fuzzy-Neural Multi-Model (HFNMM) identifier, combining the fuzzy model flexibility with the learning abilities of the RTNNs. The local and global features of the proposed HFNMM identifier are implemented by a Hierarchical Sliding Mode Controller (HSMC). The proposed HSMC scheme is applied for 1-DOF mechanical plant with friction control, where the obtained comparative results show that the HSMC with a HFNMM identifier outperforms the SMC with a single RTNN identifier.

1 Introduction

In the last decade, the Computational Intelligence (CI), including Nero-Fuzzy (N-F) and Fuzzy-Neural (F-N) systems became universal tools for many applications. Similarly to the static Neural Networks (NN), the Fuzzy Systems (FS) could approximate static nonlinear plants where structural plant information is needed to extract the fuzzy rules [1], [2]. The difference between them is that the NN models are global models where training is performed on the entire pattern range and the FS models perform a fuzzy blending of local models space based on the partition of the input space. So the aim of the N-F (F-N) model is to merge both NN and FS approaches so to obtain fast adaptive models possessing learning, [2]. The fuzzy-neural networks are capable of incorporating both numerical data (quantitative information) and expert's knowledge (qualitative information), and describe them in the form of linguistic IF-THEN rules. During the last decade considerable research has been devoted towards developing recurrent N-F models. The first attempt was made in [3] where an ANFIS with external feedback is used as an N-F controller. Through Backpropagation (BP) learning, ANFIS is adapted to refine, or derive the fuzzy IF-THEN rules using system input-output data. Due to the recurrent N-F model with internal dynamics, the recurrent fuzzy rules introduced the feedback in the antecedent and the consequent part of the model, [4]-[6], which is in fact a computational procedure. To resolve dynamic problems, in [7], [8] a Compensation-based Recurrent Fuzzy Neural Network (CRFNN) is proposed. This is a recurrent multilayer connectionist network for fuzzy reasoning constructed from a set of different fuzzy rules (normal, compensatory, recurrent) learned by different BP

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 762–771, 2007.
springerlink.com

learning laws. The main disadvantage here is the lack of universality where different layers with different concepts, different membership functions and different BP learning laws should be applied. Furthermore, the CRFNN require using a lot of fuzzy rules so to perform a good approximation. The error cost function used for a BP learning of the Gaussian membership function parameters in some cases is not convex which makes the learning convergence very time consuming, [9]. To reduce the number of IF-THAN rules, the hierarchical approach could be used [10]. A promising approach of recurrent N-F systems with internal dynamics is the application of the Takagi-Sugeno (T-S) fuzzy rules with a static premise and a dynamic function consequent part, [11]-[13]. The application of T-S fuzzy models permits to improve the interpretability applying a combination of global and local learning, [14]. So the T-S models are capable to describe a highly nonlinear systems using small number of rules, [14]. To improve the flexibility, some more works like [15]-[17], [11], [18] proposed as a dynamic function in the consequent part of the T-S rules to use a Recurrent NN (RNN). An application of the T-S approach for identification of dynamic plants with distributed parameters is given in [19]. The difference between the used in [11], [18] fuzzy neural model and the approach used in [16]-[18] is that the first one uses the Frasconi, Gori and Soda (FGS) [20] RNN model, which is sequential one, and the second one uses the Recurrent Trainable NN (RTNN) model, [21], which is completely parallel one. But it is not still enough because the neural nonlinear dynamic function ought to be learned, and the BP learning algorithm is not introduced in the T-S fuzzy rule. The paper [22] proposed to extend the T-S fuzzy rules, using in its consequent part a learning procedure instead of dynamic nonlinear function and to organize the defuzzyfication part as a second RNN hierarchical level incorporated in a Hierarchical Fuzzy-Neural Multi-Model (HFNMM) architecture. In the present paper we proposed to extend the results obtained in [22], applying a HSMC, using the information issued by the HFNMM identifier.

2 RTNN Model and SMC Description

2.1 Architecture and Learning of the RTNN

The RTNN architecture and learning, [21], [23], are summarized by the equations:

$$X(k+1) = JX(k) + BU(K); J = block - diag(J_{ii}); |J_{ii}| < 1 . \quad (1)$$

$$Z(k) = \Gamma[X(k)]; Y(k) = \Phi[CZ(k)] \quad (2)$$

$$W_{ij}(k+1) = W_{ij}(k) + \eta\Delta W_{ij}(k) + \alpha\Delta W_{ij}(k-1) \quad (3)$$

$$\Delta C_{ij}(k) = [Y_{d,j}(k)\text{-}Y_j(k)]\Phi'_j[Y_j(k)]Z_i(k), \quad (4)$$

$$\Delta J_{ij}(k) = RX_i(k-1); \Delta B_{ij}(k) = RU_i(k), \quad (5)$$

$$R = C_i(k)[Y_d(k)\text{-}Y(k)]\Gamma'_j[Z_i(k)] \quad (6)$$

Where: *Y, X,* and *U* are, respectively, output, state and input vectors with dimensions *l, n, m*; *J* is a (*nxn*)- state block-diagonal weight matrix; J_{ii} is an *i*-th diagonal block of *J* with (*1x1*) dimension; *Γ(.), Φ(.)* are vector-valued activation functions like saturation, sigmoid or hyperbolic tangent, which have compatible dimensions; W_{ij} is a general weight, denoting each weight matrix element (C_{ij}, A_{ij}, B_{ij}) in the RTNN model, to be updated; ΔW_{ij} (ΔC_{ij}, ΔJ_{ij}, ΔB_{ij}), is the weight correction of W_{ij}; while; η and α are learning rate parameters; ΔJ_{ij}, ΔB_{ij}, ΔC_{ij} are weight corrections of the weights J_{ij}, B_{ij}, C_{ij}, respectively; (Y_d -*Y*) is an error vector of the output RTNN layer, where Y_d is a desired target vector and *Y* is a RTNN output vector, both with dimensions *l*; X_i is an *i*-th element of the state vector; *R* is an auxiliary variable; Φ_j', Γ_j' are derivatives of the activation functions. Stability proof of the learning algorithm is given in [23]. The equations (1)-(6) forms a BP-learning procedure, where the functional algorithm (1), (2) represented the forward step, executed with constant weights, and the learning algorithm (3)-(6) represented the backward step, executed with constant signal vector variables. This learning procedure, denoted by Π *(L, M, N,* Y_d*, U, X, J, B, C, E)*, could be executed on-line or off-line. It uses as input data the RTNN model dimensions *l, m, n,* and the learning data vectors Y_d, *U*, and as output data – *X, J, B, C* of RTNN.

2.2 SMC Adaptive Neural Control Systems Description

Let us suppose that the studied nonlinear plant possess the following structure:

$$X_p(k+1) = F[X_p(k), U(k)]; Y_p(k) = G[X_p(k)] . \tag{7}$$

Where: $X_p(k)$, $Y_p(k)$, $U(k)$ are plant state, output and input vector variables with dimensions *np, l* and *m*, where $l = m$ is supposed; *F* and *G* are smooth, odd, bounded nonlinear functions. The block diagram of the control scheme is shown on Fig. 1a. It contains identification and state estimation RTNN and an indirect adaptive sliding mode controller. The stable nonlinear plant is identified by a RTNN with topology, given by equations (1), (2) which is learned by the stable BP-learning algorithm, given by equations (3)-(6), where the identification error $E_i(k) = Y_p(k) - Y(k)$ tends to zero ($E_i \to 0$, $k \to \infty$). The linearization of the activation functions of the learned simplified identification RTNN model (1), (2) leads to the following linear local plant model:

$$X(k+1) = JX(k) + BU(K); Y(k) = CX(k) . \tag{8}$$

Where $l = m$, is supposed. Let us define the following sliding surface with respect to the output tracking error, [24]:

$$S(k+1) = E(k+1) + \sum_{i=1}^{p} \gamma_i \mathrm{E(k\text{-}i\text{+}1)}];| \gamma_i |<1 ; \tag{9}$$

Where: *S(.)* is the sliding surface error function; *E(.)* is the systems output tracking error; γ_i are parameters of the desired error function; p is the order of the error function. The tracking error is defined as:

$$E(k) = R(k) - Y(k) . \tag{10}$$

Where: *R(k)* is a l-dimensional reference vector and *Y(k)* is an output vector with the same dimension. The objective of the sliding mode control systems design is to find a control action which maintains the systems error on the sliding surface which assure that the output tracking error reaches zero in *p* steps, where $p<n$. So, the control objective is fulfilled if:

$$S(k+1)=0\ . \tag{11}$$

The iteration of the error (10) gives:

$$E(k+1)=R(k+1)-Y(k+1)\ . \tag{12}$$

Now, let us to iterate (8) so to obtain the input/output local plant model which yields:

$$Y(k+1)=CX(k+1)=C[JX(k)+BU(K)]\ . \tag{13}$$

From (9), (11), and (12), it is easy to obtain:

$$R(k+1)-Y(k+1)+\sum_{i=1}^{p}\gamma_i \mathrm{E(k\text{-}i\text{+}1)=0}\ ; \tag{14}$$

The substitution of (13) in (14) gives:

$$R(k+1)-CJX(k)+CBU(K)+\sum_{i=1}^{p}\gamma_i \mathrm{E(k\text{-}i\text{+}1)=0} \tag{15}$$

As the local approximation plant model (8) is controllable, observable and stable, see [23], [24], the matrix *J* is diagonal, and $l = m$, the matrix product (*CB*) is nonsingular, and the plant states *X(k)* are smooth non-increasing functions. Now, from (15) it is easy to obtain the equivalent control capable to lead the system to the sliding surface which yields:

$$U_{eq}(k)=(CB)^{-1}[-CJX(k)+R(k+1)+\sum_{i=1}^{p}\gamma_i \mathrm{E(k\text{-}i\text{+}1)}];|\ \gamma_i|<1\ ; \tag{16}$$

$$U(k)=\begin{cases}\mathrm{Ueq(k)}, & \text{, if } \|\mathrm{Ueq(k)}\| < \mathrm{Uo} \\ \mathrm{-Uo\ Ueq(k)}/\|\mathrm{Ueq(k)}\|, & \text{if } \|\mathrm{Ueq(k)}\| \geq \mathrm{Uo}\end{cases} \tag{17}$$

Where: *Uo* is an upper bound of the control level; $p<N$ is the derivative order of the desired error model (sliding surface); γ_i are parameters of the desired stable error model. The proposed SMC cope with the characteristics of the wide class of plant model reduction neural control with reference model and represents in fact an indirect adaptive neural control. The block-diagram of the SMC adaptive neural control system is given on the Fig.1a, [24]. The control scheme contains a RTNN identifier, which issue a parameter and state information to the sliding mode controller. The equations (16), (17) represented also a control procedure, which is given as: $\Pi(L, M, N, X, R, E_1,\ldots, E_p, J, B, C, Uo)$.

3 HFNMM Identifier and HSM Controller Description

3.1 Hierarchical Fuzzy-Neural Multi-model Identifier Description

The papers of *Baruch et al* [15], [16], [17] proposed to use in the consequent part of the T-S rule instead of nonlinear function - a RTNN model (1), (2), given by:

$$R_i\text{:}\textbf{\textit{If}}\ \mathrm{x(k)}\ \textbf{\textit{is}}\ \mathrm{J_i}\ \textbf{\textit{and}}\ \mathrm{u(k)}\ \textbf{\textit{is}}\ B_i\ \textbf{\textit{then}}\ \mathrm{y_i(k+1) = N_i\ [x_i(k),u(k)],\ i =1,2,..,P} \tag{18}$$

Where: the function $y_i(k+1) = N_i\ [x_i(k),u(k)]$ represents the RTNN, given by the equations (1), (2); i - is the number of the function and P is the total number of RTNN approximation functions. The biases, obtained in the process of BP learning of the RTNN model could be used to form the membership functions, as they are natural centers of gravity for each variable, [17]. The number of rules could be optimized using the Mean-Square Error (MSE) of RTNN's learning, ($MSE < 2.5\%$). As the local RTNN model could be learned by the local error of approximation $E_i = Y_{di} - Y_i$, the rule (18) could be derived using the learning procedure $Y = \Pi\ (L, M, N, Y_d, U, X, J, B, C, E)$. Then (18) obtained the form, [22]:

$$R_i\text{:}\textbf{\textit{If}}\ \mathrm{x(k)}\ \textbf{\textit{is}}\ \mathrm{J_i}\ \textbf{\textit{and}}\ \mathrm{u(k)}\ \textbf{\textit{is}}\ \mathrm{B_i}\ \textbf{\textit{then}}\ \mathrm{Y_i = \Pi_i\ (L,M,N_i,Y_{di},U,X_i,J_i,B_i,C_i,E_i),\ i=1,2,..,P} \tag{19}$$

The output of the fuzzy neural multi-model system, represented by the upper hierarchical level of defuzzyfication is proposed to be given by the equation, [22]:

$$Y = \Pi\ (L, M, N, Y_d, Yo, X, J, B, C, E) \tag{20}$$

Where: the input vector Yo is formed from the vectors $y_i(k),\ i = 1,\dots, P;\ E =.\ Y_d - Y$ is the error of learning; $\Pi(.)$ is a RTNN learning procedure, given by equations (1)-(6). So, the output of the upper hierarchical defuzzyfication procedure (20) is a filtered weighted sum of the outputs of the T-S rules. As the RTNN is a universal function approximator, the number of rules P could be rather small, e.g. $P = 3$ (negative, zero, and positive) in the case of overlapping membership functions and $P = 2$ (negative and positive), in the case of not overlapping membership functions. The structure of the entire identification system (see Fig. 1b) contains a Fuzzyfier, a Fuzzy Rule-Based Inference System (FRBIS), containing up to three T-S rules (19), and a defuzzyfier. The system uses a RTNN model as an adaptive, upper hierarchical level defuzzyfier (20). The local and global errors used to learn the respective RTNNs models are $E_i(k) = Y_{di}(k) - Y_i(k);\ E(k) = Y_d(k) - Y_i(k)$. The HFNMM identifier has two levels – Lower Hierarchical Level of Identification (LLI), and Upper Hierarchical Level of Identification (ULI). It is composed of three parts: 1) Fuzzyfication, where the normalized plant output signal $Y_d(k)$ is divided in three intervals (membership functions - μ_i): positive [1, -0.5], negative [-1, 0.5], and zero [-0.5, 0.5]; 2) Lower Level Inference Engine, which contains three T-S fuzzy rules, given by (10), and operating in the three intervals. The consequent part of each rule (the consequent learning procedure) has the L, M, N_i RTNN model dimensions, Y_{di}, U, E_i inputs and Y_i (used as entry of the defuzzyfication level), X_i, J_i, B_i, C_i outputs, used for control. The T-S fuzzy rule has the form:

$$R_i\text{:}\ \textbf{\textit{If}}\ \mathrm{Yd(k)}\ \textbf{\textit{is}}\ A_i\ \textbf{\textit{then}}\ \mathrm{Y_i = \Pi_i\ (L, M, N_i, Y_{di}, U, X_i, J_i, B_i, C_i, E_i),\ i =1,2, 3} \tag{21}$$

3) Upper Level Defuzzyfication, which consists of one RTNN learning procedure, doing a filtered weighted summation of the outputs Y_i of the lower level RTNNs. The defuzzyfication learning procedure (20) has *L, M, N* RTNN model dimensions, Y_i ($P=3$), *E*, inputs, and *Y(k)* - output.

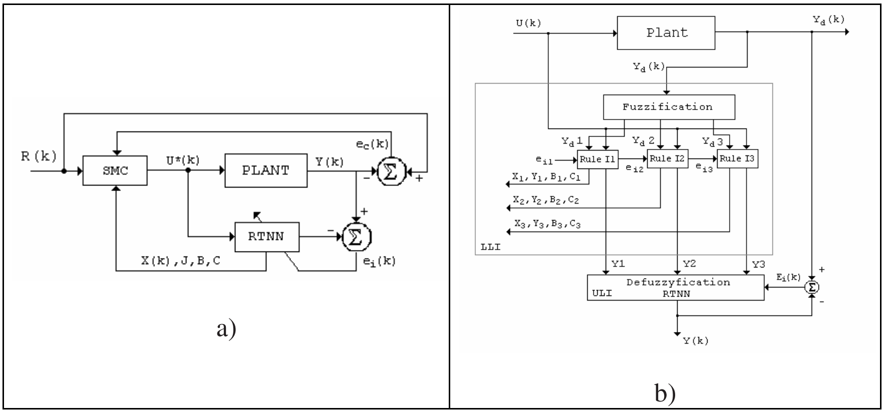


Fig. 1. Block-diagrams; a) Block-diagram of the SMC adaptive neural system; b) Detailed block-diagram of the HFNMM identifier

3.2 Indirect Adaptive Control by Means of a HSM Controller

The main objective of the HFNMM identifier is to issue states and parameters for the HFNNMM controller when its output follows the output of the plant with a minimum error of approximation. The objective of the indirect adaptive HFNMM controller is to reduce the error of control, so the plant output to track the reference signal. The block-diagram of this control is schematically depicted in Fig.2a. The identification part on the right contains three RTNNs, corresponding to the three rules, fired by the fuzzyfied plant output and taking part of the FRBIS HFNMM identifier, and the RTNN DF1 represents the defuzzyfier of the HFNMM identifier. The detailed structure of the HFNMM identifier could be seen on Fig.1b. The control part on the left contains three Sliding Mode (SM) control blocks. The SM control block represented a control T-S rule, fired by the fuzzyfied reference, and its entries are the corresponding states, and parameters issued by the HFNMM identifier. The RTNN DF2 represents the defuzzyfier. The detailed structure of the indirect adaptive HSM controller is given on Fig.2b. The structure of the entire control system has a Fuzzyfier, a Fuzzy Rule-Based Inference System (FRBIS), containing up to three T-S rules, and a defuzzyfier. The system uses a RTNN model as an adaptive, upper hierarchical level defuzzyfier, given by a learning procedure. The HFNMM controller has two levels – Lower Hierarchical Level of Control (LLC), and Upper Hierarchical Level of Control (ULC). It is composed of three parts: 1) Fuzzyfication, where the normalized reference signal *R(k)* is divided in three intervals (membership functions - μ_i): positive [1, -0.5], negative [-1, 0.5], and zero [-0.5, 0.5]; 2) Lower Level Inference Engine, which contains three T-S fuzzy rules, operating in the corresponding intervals. The

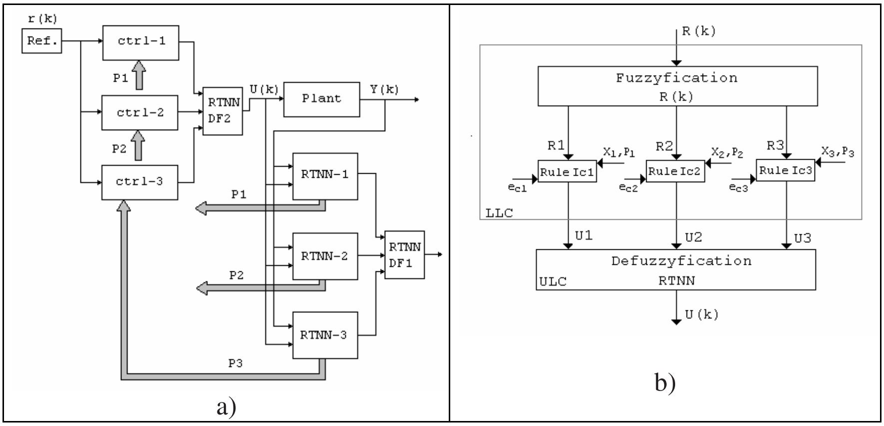


Fig. 2. Block-diagrams; a) Block-diagram of the indirect adaptive fuzzy-neural multi-model control system; b) Detailed block-diagram of the HFNMM controller

consequent part of each SM control rule realizes the SM control (16), (17), using the state, and parameter information, issued by the corresponding RTNN of the identifier. The corresponding computational control procedure has the M, L, N_i dimensions, R_i, Y_{di}, E_{ci}, X_i, J_i, B_i, C_i inputs and parameters needed. The T-S fuzzy rule has the form:

$$R_i\text{: } \textbf{\textit{If}} \text{ R(k) } \textbf{\textit{is}} \ B_i \ \textbf{\textit{then}} \ \text{U}_\text{i} = \Pi_\text{i} (\text{M, L, N}_\text{i}, \text{R}_\text{i}, \text{Y}_\text{di}, \text{X}_\text{i}, \text{J}_\text{i}, \text{B}_\text{i}, \text{C}_\text{i}, \text{E}_\text{ci}), \text{ i} = 1,2,3 \quad (22)$$

The defuzzyfication learning procedure described by:

$$U = \Pi\,(M, L, N, Y_d, Uo, X, J, B, C, E) \quad (23)$$

4 Simulation Results

Let us consider a DC-motor - driven nonlinear 1-DOF mechanical system with friction, [17], to have the following system and friction parameters: α = 0.001 m/s; Fs+ = 4.2 N; Fs- = - 4.0 N; ΔF+ = 1.8 N ; ΔF- = - 1.7 N ; vcr = 0.1 m/s; β = 0.5 Ns/m.; the period of discretization is To = 0.1s; the system gain ko = 8; the mass m = 1 kg; the load disturbance depends on the position and the velocity, (d(t) = d1q(t) + d2v(t); d1 = 0.25; d2 = - 0.7). So the discrete-time model of the 1-DOF mass mechanical system with friction is obtained in the form:

$$x_1(k+1) = x_2(k);\; x_2(k+1) = -0.025x_1(k) - 0.3x_2(k) + 0.8u(k) - 0.1fr(k); \quad (24)$$

$$v(k) = x_2(k) - x_1(k);\; y(k) = 0.1\, x_1(k) \quad (25)$$

Where: $x_1(k)$, $x_2(k)$ are system states; $v(k)$ is shaft velocity; $y(k)$ is shaft position; $fr(k)$ is a friction force, taken from [16], with given up values of friction parameters. The graphics of the simulation results, obtained with the indirect adaptive HSM control system, are shown on Fig. 3.a-d.

The topology of the identification and FF control RTNNs is (1, 5, 1), and that – of the FB control RTNN is (5, 5, 1). The topology of the defuzzyfication RTNN is (1, 3, 1). The learning rate parameters are $\eta = 0.01$, $\alpha = 0.9$ and the period of discretization of the control is $To = 0.01$ *sec*. The reference signal is $r(k) = sat\ [0.5\ sin(\pi k) + 0.5\ sin\ (\pi k/2)]$ with a saturation level as ± 0.8. The results of control show that the MSE% of control has final values which are: *0.25%* for the indirect adaptive HSM control with HFNMM identifier; *0.4%* for the SMC with single RTNN identifier. From the graphics of Fig. 3a-d, we could see that the indirect adaptive HSM control with HFNMM identifier is better than the SMC using single RTNN identifier.

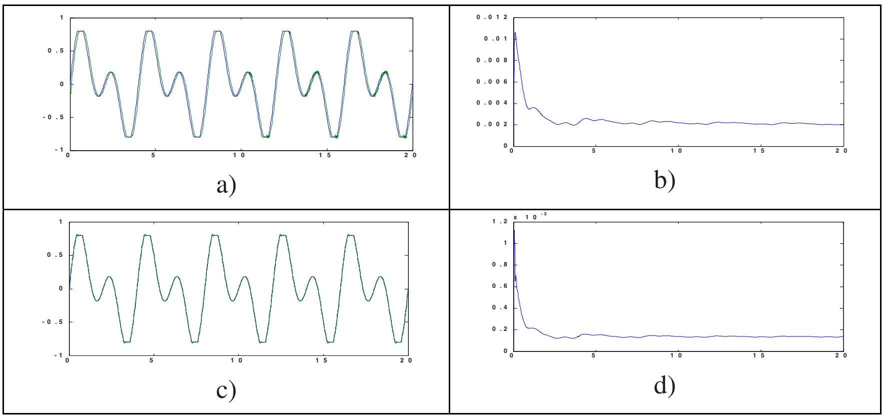


Fig. 3. Comparative control system results; a) reference signal vs. plant output (SMC and a single RTNN identifier); b) MSE% of SMC; c) reference signal vs. plant output (HSMC a HFNMM identifier); d) MSE% of HSM control

5 Conclusions

The present paper proposed a new identification and HSM adaptive control system based on the Hierarchical Fuzzy-Neural Multi-Model identifier.

The HFNMM identifier has three parts: 1) fuzzyfication, where the output of the plant is divided in intervals μ (here they are three - positive [1, -0.5], negative [-1, 0.5], and zero [-0.5, 0.5]); 2) inference engine –the lower hierarchical level, which contains three T-S rules corresponding to three RTNN models operating in three overlapping intervals μ; 3) defuzzyfication – the upper hierarchical level, which consists of one RTNN doing a filtered weighted summation of the outputs of the lower level RTNNs. The learning and functioning of both hierarchical levels is independent. The main objective of the HFNMM identifier is to approximate the plant output with a minimum error. The HFNMM identifier is incorporated in a HSM control scheme, using a HSM controller. The proposed HFNMM-based control scheme is applied for a 1-DOF mechanical plant with friction control. The comparative simulation results show the superiority of the proposed HSM control scheme with respect to others.

Acknowledgements. The MS students Jose-Luis Olivares Guzman, and Rosalba Galvan Guerra, and the Ph.D. student Carlos-Roman Mariaca-Gaspar, are thankful to CONACYT-MEXICO for the scholarship received during his studies in the Department of Automatic Control, CINVESTAV-IPN, Mexico.

References

1. R. Babuska, Fuzzy Modeling for Control, Norwell, MA, Kluwer, NY, 1998.
2. C.T. Lin, C.S.G.Lee, Neural Fuzzy Systems: A Neuro - Fuzzy Synergism to Intelligent Systems, Prentice Hall, Englewood Cliffs, NJ, 1996.
3. J.S.R. Jang, ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man and Cybernetics 23 (3) 1993, 665-685.
4. Y. Gao, M. J. Er, NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches, Fuzzy Sets and Systems, 150 (2) (2005) 331-350.
5. J.B. Theocharis, A high-order recurrent neuro-fuzzy system with internal dynamics: Application to the adaptive noise cancellation Fuzzy Sets and Systems, 157 (4) (2006) 471-500.
6. S.M. Zhou, L.D. Xu, A new type of recurrent fuzzy-neural network for modeling dynamic systems, Knowledge Based Systems, 14 (5-6) (2001) 243-251.
7. C.J. Lin, C.H. Chen, A compensation-based recurrent fuzzy neural network for dynamic system identification, European J. of Operational Research, 172 (2) (2006) 696-715.
8. C.J. Lin, C.H. Chen, Identification and prediction using recurrent compensatory neuro-fuzzy systems, Fuzzy Sets and Systems, 150 (2) (2005) 307-330.
9. S. Haykin, Neural Networks, A Comprehensive Foundations, 2-nd Edition, Prentice Hall, Englewood Cliffs, NJ, 1999.
10. A. Nürnberger, A hierarchical recurrent neuro-fuzzy model for system identification, International Journal of Approximate Reasoning, 32 (2-3) (2003) 153-170.
11. P.A. Mastorocostas, J.B. Theocharis, A recurrent fuzzy-neural model for dynamic system identification, IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, 32 (2002) 176-190.
12. T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Trans. Systems, Man, and Cybernetics, 15 (1985) 116-132.
13. M. Teixeira, S. Zak, Stabilizing controller design for uncertain nonlinear systems using fuzzy models, IEEE Trans. on Systems, Man, and Cybernetics, 7 (1999) 133-142.
14. J. Yen, L. Wang, Ch. W. Gillespie, Improving the Interpretability of TSK fuzzy models by combining global learning and local learning, IEEE Trans. on Fuzzy Systems, 6 (4) (1998) 530-537.
15. I. Baruch, E. Gortcheva, Fuzzy neural model for nonlinear systems identification, Proc. of the AARTC IFAC Workshop, Cancun, Mexico, April 15-17, 1998, 283-288.
16. I. Baruch, F. Thomas, R. Garrido, E. Gortcheva, A hybrid multi-model neural network for nonlinear systems identification, Proc. of the Int. Joint Conf. on Neural Networks, IJCNN'99, Washington D.C., USA, July 10-16, 1999, 4278-4283.
17. I. Baruch, J.M. Flores, R. Garrido, A fuzzy-neural recurrent multi-model for systems identification and control, Proc. of the European Control Conference, ECC'01, Porto, Portugal, Sept. 4-7, 2001, 3540-3545.
18. P.A. Mastorocostas, J.B. Theocharis, An orthogonal least-squares method for recurrent fuzzy-neural modeling, Fuzzy Sets and Systems, 140 (2) (2003) 285-300.

19. A. Nürnberger, A. Radetzky, R. Kruse, Using recurrent neuro-fuzzy techniques for the identification and simulation of dynamic systems, Neurocomputing 36 (2001) 123-147.
20. P. Frasconi, M. Gori, G. Soda, Local feedback multilayered networks, Neural Computation, 4 (1992) 120-130.
21. I. Baruch, I. Stoyanov, E. Gortcheva, Topology and learning of a class RNN, ELEKTRIK 4-Supplement (1996) 35-42.
22. I. Baruch, J.L. Olivares, C.R. Mariaca-Gaspar, R. Galvan, A fuzzy-neural hierarchical multi-model for systems identification and direct adaptive control, Proc. of IFSA 2006, June, Cancun, Mexico (submitted for presentation).
23. I. Baruch, J.M. Flores, F. Nava, I.R. Ramirez, B. Nenkova, An advanced neural network topology and learning applied for identification and control of a D.C. motor, Proc. of the 1-st Int. IEEE Symp. Intel. Syst., Varna, Bulgaria, Sept., 2002, 289-295.
24. I. Baruch, L.A. Hernandez, J. Barrera-Cortes, Adaptive discrete-time sliding mode control using recurrent neural networks, Proc. of the 2-nd IFAC Symposium on System, Structure and Control, Oaxaca, Mexico, Dec. 8-10, 2004, file [SSSC112.pdf].

A Method for Creating Ensemble Neural Networks Using a Sampling Data Approach

Miguel López[1, 2], Patricia Melin[2], and Oscar Castillo[2]

[1] PhD Student of Computer Science in the Universidad Autonoma de Baja California, Tijuana, B. C., México
danym23@aol.com
[2] Computer Science in the Graduate Division Tijuana, Institute of Technology Tijuana, B. C., Mexico
pmelin@tectijuana.mx,ocastillo@tectijuana.mx

Abstract. Ensemble Neural Networks are a learning paradigm where many neural networks are used together to solve a particular problem. In this paper, the relationship between the ensemble and its component neural networks is analyzed with the goal of creating of a set of nets for an ensemble with the use of a sampling-technique. This technique is such that each net in the ensemble is trained on a different sub-sample of the training data.

1 Introduction

The defining characteristic of an ensemble is that it involves combining a set of nets each of which essentially accomplishes the same task.

The use of an ensemble can provide an effective alternative to the tradition of generating a population of nets, and then choosing the one with the best performance, whilst discarding the rest. The idea to ensemble neural networks is to find ways of exploiting instead the information contained in these redundant nets.

Combining estimators to improve performance has recently received more attention in the neural networks area, in terms of neural computing, there are two mains issues: first creation, or selection, of a set of nets to be combined in an ensemble; and second, the methods by which the outputs of the members of the ensemble are combined[2].

2 Methods for Creating Ensemble Members

There is no advantage to combining a set of nets which are identical; identical that is, in that they generalize in the same way. In principle, a set of nets could vary in terms of their weights, the time they took to converge, and even their architecture and yet constitute the same solution, since they resulted in the same pattern of errors when tested on a test set. There are many parameters which can be manipulated an efforts, to obtain a set of nets which generalize differently.

These include the following: initial conditions, training data, the topology of the nets, and the training algorithm.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 772–780, 2007.
springerlink.com

The main methods which have been employed for the creation of ensemble members are; changing the set of initial random weights, changing topology, changing the algorithm employed, changing the data[1].

The last one method, changing the data, is to be most frequently used for the creation of ensembles are those which involve altering the training data.

There are a number of different ways in which this can be done, which include: sampling data, disjoint training sets, boosting and adaptive resampling, different data sources, and preprocessing. These can be considerated individually, and it should be noted that ensemble could be created using a combination of two or more these techniques.

3 Sampling Data

A common approach to the creation of a set of nets for an ensemble is to use some form of sampling technique, such that each net in the ensemble is trained on a different subsample of the training data. Resampling methods which have been used for this purpose include cross-validation (Krogh and Vedelsby[2],1995) and bootstrapping(Breiman[4]), although in statistics the methods are better known as techniques for estimating the error of a predictor from limited sets of data[3].

In bagging (Breiman[5]) a training set containing N cases is perturbed by sampling with replacement (bootstrap) N times from the training set. The perturbed data set may contain repeats. This procedure can be repetead several times to create a number of different, although overlapping, data sets. Such statistical resampling techniques are particularly useful where there is a shortage of data.

Disjoint training sets, this method sampling without replacement, there is then no overlap between the data used to train different nets.

Boosting and adaptive resampling, the base of this algorithm is that training sets adaptively resampled , such that the weights in the resampling are incremented for those case which are most often misclassified [3].

4 Sampling Data Method for Creating Ensemble Neural Networks

The goal of the algorithm in the sampling data method is to find the data set of training for the creation of ensemble neural networks.

To train the network we used the Rastrigin´s function, wich has two dependent variables, and the function is defined as:

```
Ras(p)=20+(p1.^2+p2.^2-10.0*(cos(2*pi.*p1)
      +cos(2* pi .* p2)))
```
(1)

5 The Algorithm Sampling Data Method

1. - The collected data of Rastrigin´s function are used to train the neural network.

$$P=[p(i)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots p(n)] \quad (2)$$

2. - In the process of training of the neural network the data that generate an error of greater or equal to the mean square error are stored like the new data set of training.

$$\text{If } e(i) > RMSE \text{ then } Psampling = [p(i) \ldots\ldots\ldots\ldots p(n)] \quad (3)$$

3. - The stored data represents the data set of sampling that is going away to use in the creation of the neural networks ensemble.

$$Psampling = [p(i) \ldots\ldots\ldots\ldots p(n)] \quad (4)$$

4. - The output of the modular neural networks is evaluated by, wich is average of the output of the modules.

$$Oneural\_mod = prom[(Mnn(i) + \ldots.. Mnn(n)] \quad (5)$$

5.1 Sampling Data Method

The neural networks in this study were trained using the back propagation algorithm with sigmoid activation function. During the learning phase, the output error by each data of training is verified, and if it's greater to the mean square error, input data is stored to be used like the group of sampling data that were used for the creation of the neural networks ensemble. Rastrigin Function in figure 1 is used for the training the neural network using the data generated by the Rastrigin´s function without using sampling data method.

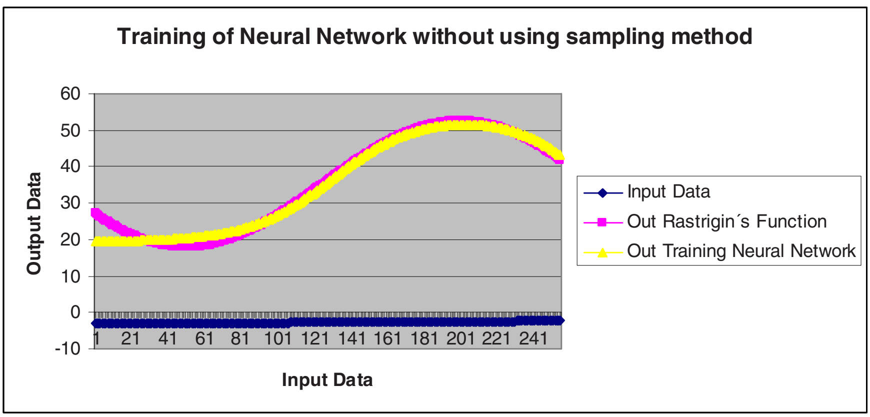


Fig. 1. Training of Neural Network without sampling data method

We show in Figure 2 the results of repeating 30 times the training of data and the average error achieved.

Rastrigin´s Function in figure 3 is used for training the neural network, using the data generated by Rastrigin´s function with a sampling data method.

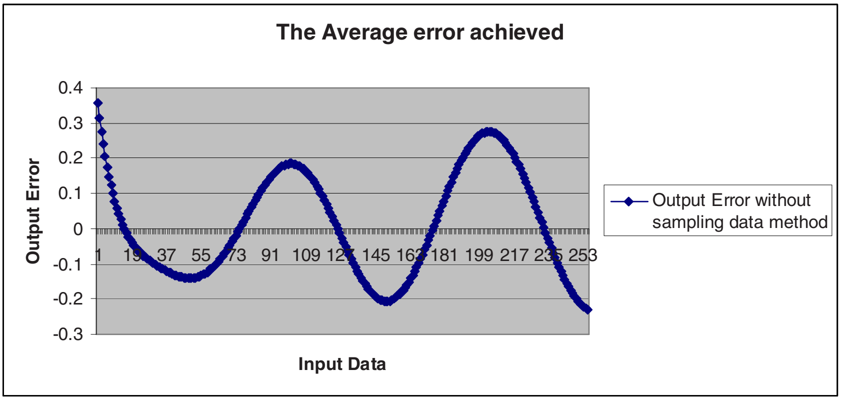


Fig. 2. The average error achieved without sampling data method

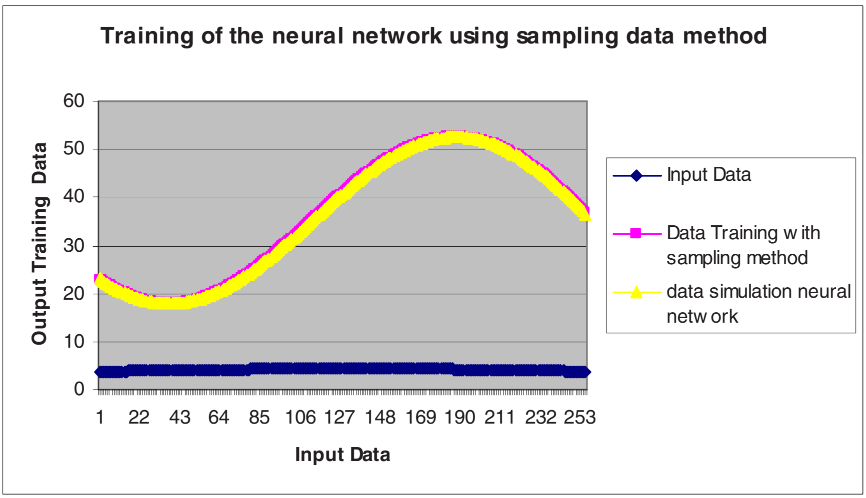


Fig. 3. Training of the neural network using sampling data method

Executing thirty times the training process and showing the average error achieved in figure 4, using data sampling method.

In Figure 5 we show a comparison between the Output Error using the Sampling Data Method and without in sampling data method.

5.2 Creating Modular Neural Networks

In the experimental results shown previously with the sampling method, for the creation of modular neural networks, we used a neural network of 1 hidden layer and 40 neurons, learning rate of .001, Rastrigin´s function, 2000 input data, training function

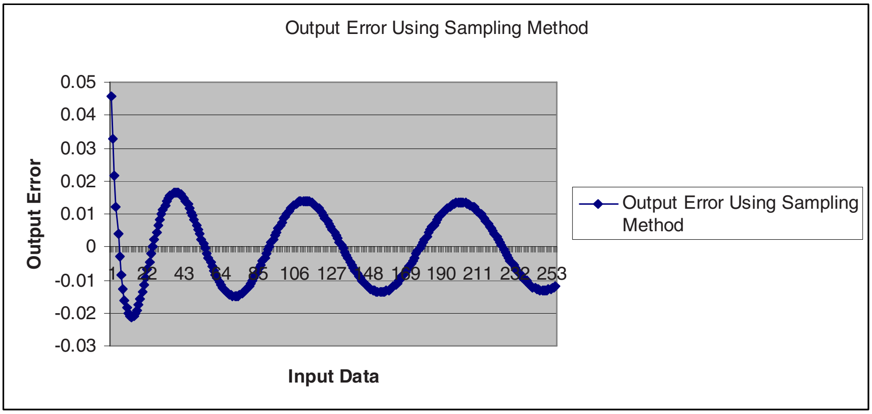


Fig. 4. The average error achieved using sampling data method

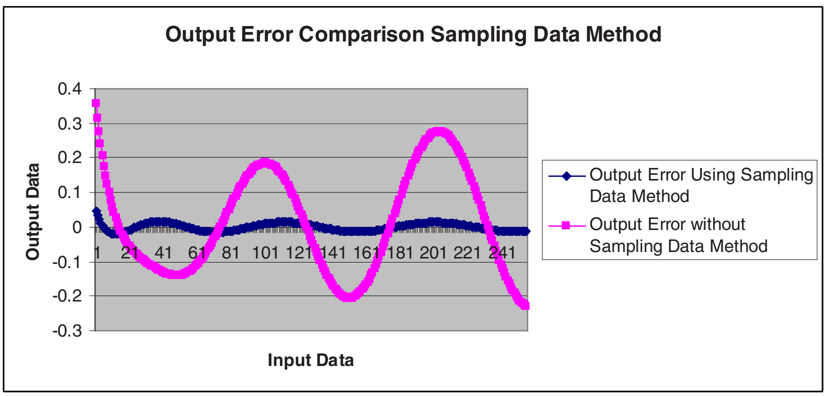


Fig. 5. Output Error Comparison sampling data method

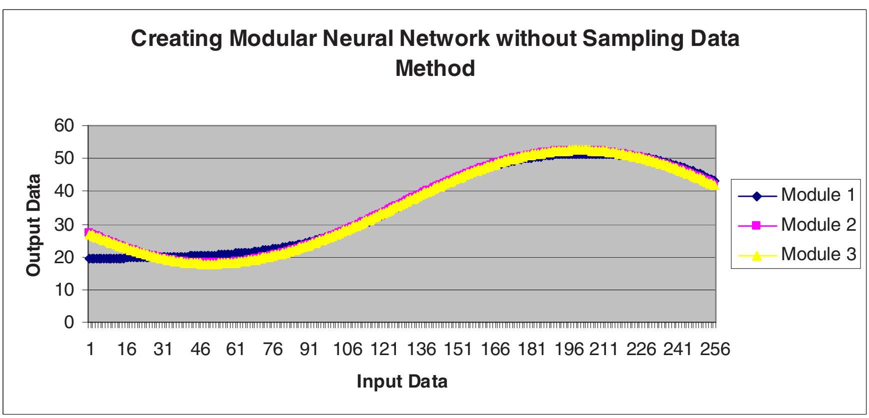


Fig. 6. Modular Neural Network without sampling data method

Levenberg-Marquardt backpropagation. The First three modular neural networks created without using the sampling method are the following.

We show in Figure 6 the experimental results creating modular neural networks without sampling data.

We show in Figure 7 the experimentals results creating modular neural networks using sampling data method.

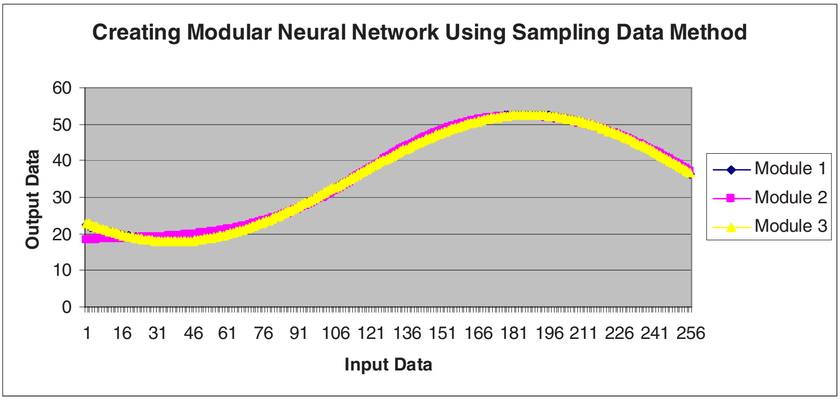


Fig. 7. Modular Neural Network using sampling data method

The previous results do not have a great difference using sampling method and without using it. Subsequently results of the output error of the modular neural networks, are presented and their average for the modular neural network.

We show in Figure 8 the experimentals results creating modular neural network without sampling data method with modular average.

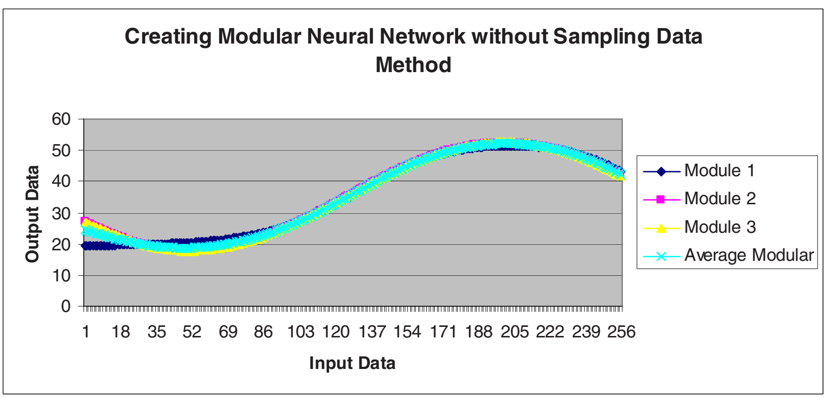


Fig. 8. Modular Neural Networks without sampling data method with modular average

We show in Figure 9 the experimentals results creating modular neural networks using sampling data method with modular average.

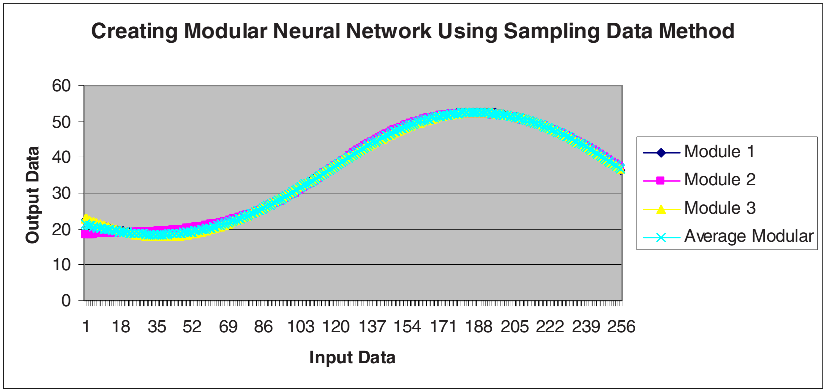


Fig. 9. Modular Neural Networks using sampling data method with modular average

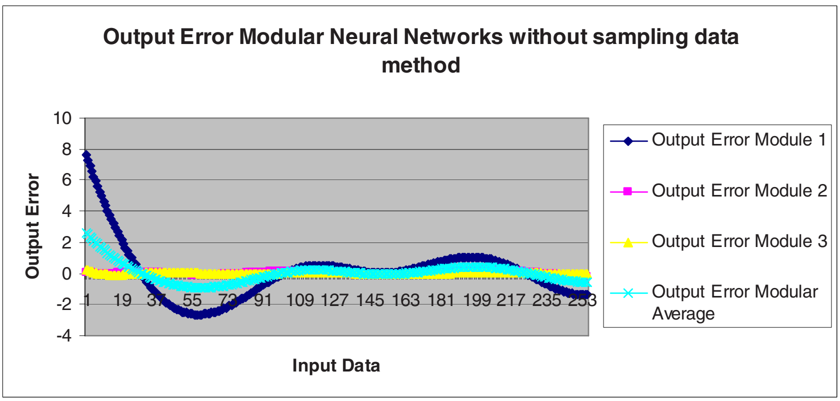


Fig. 10. Output Error of the Modular Neural Network without sampling data method

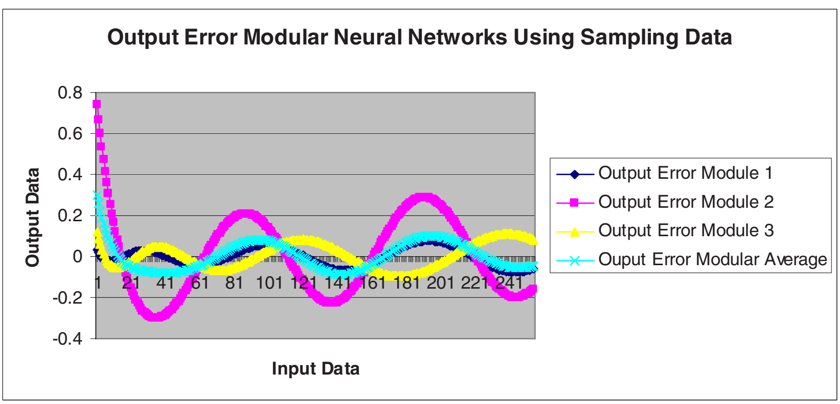


Fig. 11. Output Error of the Modular Neural Network using sampling data method

We show in Figure 10 the results of the output error without the sampling data method.

We show in Figure 11 the results the output error using sampling data method.

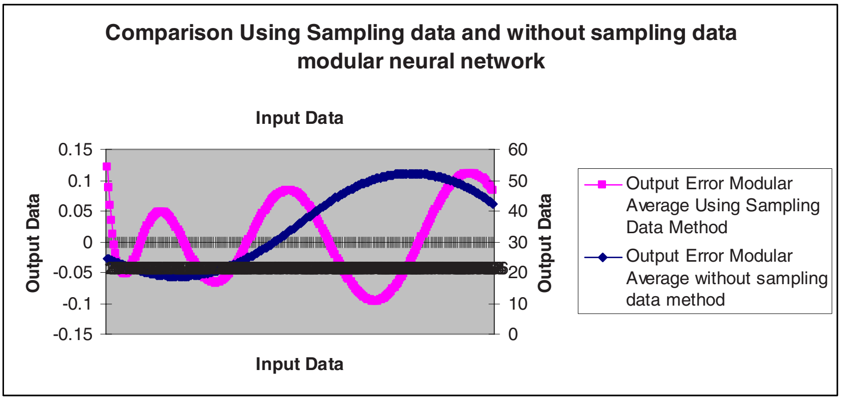


Fig. 12. Comparison using and without sampling data Modular Neural Networks

6 Conclusions

In conclusion, it's difficult to identify the best statistical sampling method to estimate the predictive accuracy of a neural network. The decision strongly depends on the complexity of the problem, and the number and variability of the available cases. Based on the experimental results obtained, we can comment that to utilize less data for the training of the modular neural networks is minimized the output error of the neural network, its necessary to prove a sampling method with different benchmark functions to establish a criterion that verifies it's operation to do a comparative with other sampling methods. It can be concluded that to a create modular neural networks is a good option to compare sampling data methods that allow the process of training modular neural network to work satisfactorily, without themselves making changes to with the architecture and parameters of the modular network neuronal.

We show in figure 12 comparison results between using and without sampling data method.

References

1. The effect of data sampling on the performance evaluation of artificial neural networks in medical diagnosis, Georgia D. Tourassi, Carey E. Floyd, Med Decis Making. 1997 Apr-Jun;17(2):186-92. PMID: 9107614 [PubMed - indexed for MEDLINE].
2. Feature Selection for Ensembles, David W. Opitz, Sixteenth National Conference on Artificial/ Intelligence (AAAI), (379-384). Orlando, FL.
3. On combining artificial neural nets, Sharkey, A., Connection Science, Vol. 8, pp.299-313, 1996.
4. *Bagging predictors* ,Leo Breiman., Machine Learning, 24(2):123--140, 1996.

5. Stacked Regressions ,Leo Breiman., Machine Learning 24(1): 49-64 (1996).
6. Generating Accurate and Diverse Members of a Neural-Networks Ensemble, David W. Opitz , Jude W. Shavlik, Advances in Neural Information Processing Systems 8, D. S. Touretzky, M. Mozer, and M. Hasselmo, eds, MIT Press, Cambridge, MA, 1996, pp. 535-541.
7. Overfitting and Diversity in Classification Ensembles based on Feature Selection, Padraig Cunningham, TCD Computer Science Technical Report, TCD-CS-2000-07.
8. An Ensemble Neural Networks strategy for time series and spatio-temporal forecasting applications, Auroop R Ganguly, American Geophysical Union Spring Meeting, Washington, DC, Bras (2002).
9. Face Recognition Using Ensembles of Netrworks, S. Gutta, J. Huang, B. Takacs and H. Wechsler, *icpr*, p. 50, 13th International Conference on Pattern Recognition (ICPR'96) - Volume 4, 1996. Vienna, Austria.
10. Popular Ensemble Methods: An Empirical Study, Opitz, D. and Maclin, R. (1999), Journal of Artificial Intelligence Research, Volume 11, pages 169-198.
11. Neural Nets and Diversity. Proceedings of the 14th International Conference on Computer Safety, Reliability and Security, Sharkey, A.J.C., Sharkey, N.E. and Chandroth, G.O., Belgirate, Italy, 1995, pp 375-389.
12. Searching weight space for backpropagation solutions type, Sharkey, A.J.C., Sharkey, N.E., Neary, J., 1995, In L.F. Niklasson and M.B. Boden (Eds) Current Trends in Connectionism, Lawrence Erlbaum Associates, Hillsdale, N.J., pp 103-121.
13. Generating accurate and diverse members of a neural networkensemble, D. W. Opitz and J. W. Shavlik ,in: Advances in Neural Information Processing Systems 8, D. S. Touretzky, M. Mozer, and M. Hasselmo, eds, MIT Press, Cambridge, MA, 1996, pp.535-541.
14. An Empirical Evaluation of Bagging and Boosting. Richard Maclin, David W. Opitz:, AAAI/IAAI 1997: 546-551.
15. Modularity, combining and artificial neural nets Amanda.J. C. Sharkey. Connection Science, Vol. 9, No. 1, pages 3-10, 1997.

Part XVI

Soft Computing

Using Fuzzy Sets for Coarseness Representation in Texture Images

J. Chamorro-Martínez[1], E. Galán-Perales[1], J.M. Soto-Hidalgo[1], and B. Prados-Suárez[2,⋆]

[1] Department of Computer Science and Artificial Intelligence, University of Granada, C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
{jesus,elena,soto}@decsai.ugr.es

[2] Department of Computer Science, University of Jaen, C/ Alfonso X El Sabio 28, 23700, Linares, Jaen, Spain
belenps@ujaen.es

Abstract. Texture is a visual feature frequently used in image analysis that has associated certain vagueness. However, the majority of the approaches found in the literature do not either consider such vagueness or they do not take into account human perception to model the related uncertainty. In this paper we model the concept of "coarseness", one of the most important textural features, by means of fuzzy sets and considering the way humans perceive this kind of texture. Specifically, we relate representative measures of coarseness with its presence degree. To obtain these "presence degrees", we collect assessments from polls filled by human subjects, performing an aggregation of such assessments. Thus, the membership function corresponding to the fuzzy set "coarseness" is modelled by using as reference set the representative measures and the aggregated data.

Keywords: Image features, texture features, fuzzy texture, visual coarseness.

1 Introduction

Texture is being increasingly recognized as an important cue for the analysis of natural imagery. It is one of the most difficult visual features to be characterized due to the imprecission of the concept itself. In fact, there is not an accurate definition for the concept of texture but some widespread intuitive ideas. In this way, texture is described by some authors as local changes in the intensity patterns or gray tones. Other authors consider texture as a set of basic items called *texels* (or texture primitives), arranged in a certain way [1]. Moreover, it is usual for humans to describe visual textures according to some "textural concepts" like *coarseness*, *orientation*, *regularity* [2]. To describe such concepts, linguistic labels are used (e.g. coarse or fine can be used to describe coarseness).

The own imprecision of the concept of texture suggests to use representation models that incorporate the uncertainty. Nevertheless, the majority of the

⋆ This work has been partially supported by "Consejera de Innovacin Ciencia y Empresa" under TIC-1570 project.

Fig. 1. Some examples of images with different degrees of fineness

approaches that can be found are crisp proposals [3] where uncertainty is not properly taken into account. To deal with the imprecission relative to visual texture, there are some approaches which introduce the use of fuzzy logic. However, in many of these approaches, fuzzy logic is usually applied just during the process but the output do not habitually model the imprecission (being often a crisp one). Examples of this fact are frequently found in the literature, like those approaches that use texture to perform image segmentation. A lot of these approaches use fuzzy clustering [4], fuzzy rules [5], etc. but most of them have in common that the obtained result is crisp, no matter the intermediate representation for uncertainty used.

In this paper we focus our study on coarseness, one of the textural properties most used in the literature which allows to distinguish between fine and coarse textures. In fact, the concept of texture is usually associated to the presence of fineness. A *fine* texture can be considered as small texture primitives with big gray tone differences between neighbour primitives (e.g. the image in figure 1(A)). On the contrary, if texture primitives are bigger and formed by several pixels, it is a *coarse* texture (e.g. the image in figure 1(I)).

In our approach, we propose to model fineness by means of fuzzy sets to deal with the problem of imprecision found in texture characterization. To do this, two questions will be faced: what reference set should be used for the fuzzy set, and how to obtain the related membership functions. To solve the first question, a set of measures will be automatically computed from the texture image. To answer the second question, functional relationship between a certain measure and the presence degree of a textural concept related to it will be found considering human perception.

The rest of the paper is organized as follows. In section 2 we introduce our methodology to obtain the fuzzy sets related to fineness textural concept. In section 3 we show the results of applying the model and the main conclusions and future work are sumarized in section 4.

2 Fineness Modelling

There are many measures over the literature that, given an image, capture the fineness (or coarseness) presence in the sense that the greater the value given by the measure, the greater the perception of texture. However, there is no perceptual relationship between the value given by these measures and the degree in which the humans perceive the texture. Thus, given a certain value calculated by applying a measure to an image, there is not an immediate way to decide whether there is a fine texture, a coarse texture or something intermediate (i.e. there is not a textural interpretation).

To face this problem, we propose to model the fineness perception as a fuzzy set defined on the domain of a given measure. Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ be a set of measures of fineness (e.g. $\mathcal{P} = \{EdgeDensity, Variance, Entropy\}$) and let $\mathcal{T}_k$ be a fuzzy set defined on the domain of $P_k \in \mathcal{P}$ representing the concept of "fineness". Thus, the membership function associated to $\mathcal{T}_k$ will be defined as[1]

$$\mathcal{T}_k : \mathbb{R} \rightarrow [0, 1] \tag{1}$$

where a value of 1 will mean fineness presence while a value of 0 will mean no fineness presence (i.e. coarseness presence).

In this section, given a measure $P_k \in \mathcal{P}$, we propose to obtain $\mathcal{T}_k$ by finding a functional relationship between P_k and the perception degree of fineness. To do it, we will use a set $\mathcal{I} = \{I_1, \ldots, I_N\}$ of N images that fully represent the different degrees of fineness. Thus, for each image $I_i \in \mathcal{I}$, we will obtain (a) a human assessment of the fineness degree perceived, noted as v^i, which will be collected by means of a poll with human subjects (section 2.1), and (b) a value calculated applying the measure $P_k \in \mathcal{P}$ to the image I_i, noted as m_k^i (section 2.2). From the multiset $\{(m_k^1, v^1), \ldots, (m_k^N, v^N)\}$, the membership function $\mathcal{T}_k$ will be estimated (section 2.3).

2.1 Assessment Collection

In this section, the way to obtain a vector $\Gamma = [v1, \ldots, v^N]$ of the assessments of the perception degree of fineness from the image set $\mathcal{I} = \{I_1, \ldots, I_N\}$ will be described. Thus, firstly the image set $\mathcal{I}$ will be selected. After that, a poll which allows to get assessments of the perception degree of fineness will be designed. These assessments will be obtained for each image in $\mathcal{I}$, so an aggregation of the different assessments will be performed.

The texture image set

A set $\mathcal{I} = \{I_1, \ldots, I_N\}$ of $N = 80$ images representative of the concept of *fineness* has been selected. Figure 1 shows some images extracted from the set $\mathcal{I}$. The selection was done to cover the different perception degrees of fineness with a representative number of images. Furthermore, the images have been chosen so that as far as possible, just one perception degree of fineness is perceived.

[1] To simplify the notation, as it is usual in the scope of fuzzy sets, we will use the same notation $\mathcal{T}_k$ for the fuzzy set and for the membership function that defines it.

The poll. Given the image set $\mathcal{I}$, the next step is to obtain assessments about the perception of fineness from a set of subjects. From now on we shall note as $\Theta^i = [o_1^i, \ldots, o_L^i]$ the vector of assessments obtained from L subjects for the image I_i. To get Θ^i, subjects will be asked to assign images to classes, so that each class has associated a perception degree of fineness.

In particular, 20 subjects have participated in the poll and 9 classes have been considered. The first nine images in figure 1 show the nine representative images for each class used in this poll. It should be noticed that the images are decreasingly ordered according to the presence degree of the fineness concept. The first class (Figure 1(A)) represents a presence degree of 1 while the ninth class (Figure 1(I)), represents a presence degree of 0. The rest of the classes (Figure 1(B)-(H)) represent degrees in the interval (0,1).

As result, a vector of 20 assessments $\Theta^i = [o_1^i, \ldots, o_{20}^i]$ is obtained for each image $I_i \in \mathcal{I}$. The degree o_j^i associated to the assessment given by the subject S_j to the image I_i is computed as $o_j^i = (9 - k) * 0.125$, where $k \in \{1, \ldots, 9\}$ is the index of the class C_k to which the image is assigned by the subject.

Assessment aggregation. Our aim at this point is to obtain, for each image in the set $\mathcal{I}$, one assessment v^i that summarizes the assessments Θ^i given by the different subjects about the presence degree of fineness.

To aggregate opinions we have used an OWA operator guided by a quantifier [6]. Concretely, the quantifier "the most" has been employed, which allows to represent the opinion of the majority of the polled subjects. This quantifier is defined as

$$Q(r) = \begin{cases} 0 & \text{if } r < a, \\ \frac{r-a}{b-a} & \text{if } a \le r \le b, \\ 1 & \text{if } r > b \end{cases} \tag{2}$$

with $r \in [0, 1]$, $a = 0.3$ and $b = 0.8$. Once the quantifier Q has been chosen, the weighting vector of the OWA operator can be obtained following Yager [6] as $w_j = Q(j/L) - Q((j-1)/L), j = 1, 2, ..., L$. According to this, for each image $I_i \in \mathcal{I}$, the vector Θ^i obtained from L subjects will be aggregated into one assessment v^i as follows:

$$v^i = w_1 \hat{o}_1^i + w_2 \hat{o}_2^i + ... + w_L \hat{o}_L^i \tag{3}$$

where $[\hat{o}_1^i, \ldots, \hat{o}_L^i]$ is a vector obtained by ranking in nonincreasing order the values of the vector Θ^i.

2.2 Fineness Measures

In our proposal, the fuzzy set $\mathcal{T}_k$ will be defined on the domain of a certain measure of fineness P_k. Due to the fact that there are many measures in the literature that characterize the presence of fine texture, a selection of a suitable set $\mathcal{P} = \{P_1, \ldots, P_K\}$ is needed. In this paper, we propose to use the 18 measures shown in the first column of table 1 (that includes classical statistical measures well known in the literature, measures in the frequency domain, etc.).

From the measures shown in table 1, some will have better ability to represent fineness while others will be worse. Thus, the question of what ability a measure has to discriminate different presence degrees of fineness needs to be solved, i.e. how many classes can P_k actually discriminate. To face this question, we propose to analyze each $P_k \in \mathcal{P}$ by applying a set of multiple comparison tests following the algorithm 1. This algorithm starts with an initial partition and iteratively joins clusters until a partition in which all classes are distinguishable is achieved. In our proposal, the initial partition will be formed by the 9 classes used in our poll (where each class will contain the images assigned to it by the majority of the subjects), as δ the Euclidean distance between the centroids of the involved classes will be used, as ϕ a set of 5 multiple comparison tests will be considered (concretely, the tests of Scheffé, Bonferroni, Duncan, Tukey's least significant difference, and Tukey's honestly significant difference [7]), and finally the number of positive tests to accept distinguishability will be fixed to $NT = 3$.

From now on, we shall note as $C_1^k, C_2^k, \ldots, C_{NC_k}^k$ the NC_k classes that can be discriminated by P_k. For each C_r^k, we will note as $\bar{c}_r^k$ the class representer value and as $\bar{v}_r^k$ the presence degree of fineness associated to C_r^k. In this paper, we propose to compute $\bar{c}_r^k$ as the mean of the measure values in the class C_r^k and $\bar{v}_r^k$ as the mean of the presence degrees of fineness associated to the classes grouped into C_r^k.

In the case of fineness, table 1 shows the results obtained by applying the proposed algorithm 1 with the different measures considered in this paper. The second column of this table shows the number of classes that each measure can discriminate and the third column shows how the initial classes have been grouped. Moreover, the columns from fourth to seventh show the representer values $\bar{c}_r^k$ and $\bar{v}_r^k$ associated to each cluster. It can be noticed that the measure that can discriminate a higher number of classes is ED, while the majority of the measures discern 2 or 3 classes. Furthermore, the measures in the last 4 rows of table 1 cannot discriminate more than one class, being discarded for later analysis because of their little representativeness of fineness.

In reference to the the classes joint, note that for most cases the initial classes 1 and 9 have not been grouped into a greater cluster except in the case of the measures FD and Tamura. For these measures, the classes 1, 2 and 3 were grouped into one cluster, which means they cannot discriminate properly the presence degree of fineness when the texture is getting fine.

2.3 Obtaining the Membership Function

In this section we will deal with the problem of obtaining the membership function for the fuzzy set $\mathcal{T}_k$. Concretely, we propose to define such membership

Algorithm 1. Obtaining the distinguishable clusters

Input:
  $Part^0 = C_1, C_2, \ldots, C_n$: Initial Partition
  δ: distance function between clusters
  ϕ: Set of multiple comparison tests
  NT: Number of positive tests to accept distinguishability
1.-Initialization
  $k = 0$
  $distinguishable = false$
2.- While (distinguishable = false) and (k < n)
  Apply the multiple comparison tests ϕ to $Part^k$
  If for each pair C_i, $C_j \in Part^k$ more than NT of the multiple comparison tests ϕ show distinguishability
    $distinguishable = true$
  Else
    Search for the pair of clusters C_r, C_{r+1}, verifying
      $\delta(C_r, C_{r+1}) = \min\{\delta(C_i, C_{i+1}), C_i, C_{i+1} \in Part^k\}$
    Join C_r and C_{r+1} on a cluster $C_u = C_r \cup C_{r+1}$
    $Part^{k+1} = Part^k - C_r - C_{r+1} + C_u$
    $k = k + 1$
3.- Output: $\widetilde{Part}_k = C_1, C_2, \ldots, C_{n-k}$

function as a linear spline that associates the values given by a certain measure with the assessments given by the human subjects, i.e.:

$$\mathcal{T}_k(x) = \begin{cases} 0 & x \leq x_1 \\ f^1(x) & x \in (x_1, x_2] \\ f^2(x) & x \in (x_2, x_3] \\ \vdots & \vdots \\ 1 & x > x_{NC_k} \end{cases} \tag{4}$$

with $f^r(x)$ being a straight line defined as $f^r(x) = a_1^r x + a_0^r$.

To obtain the parameters of equation 4, the classes resulting from the application of the algorithm 1 (that in fact, were obtained considering the ability of the measure to distinguish the classes originally given by the subjects) will be used. In this paper, the spline knots will be chosen considering the class representers, i.e., $x_r = \bar{c}_r^k$ with x_r being the centroid of C_r^k. The function $f^r(x)$ will be obtained as the straight line defined between the points $(\bar{c}_r^k, \bar{v}_r^k)$ and $(\bar{c}_{r+1}^k, \bar{v}_{r+1}^k)$, with $\bar{v}_r^k$ being the fineness degree of presence related to the cluster C_r^k. Thus, the parameters a_1^r and a_0^r of $f^r(x)$ are computed as $a_1^r = \frac{\bar{v}_{r+1}^k - \bar{v}_r^k}{\bar{c}_{r+1}^k - \bar{c}_r^k}$ and $a_0^r = \bar{v}_r^k - \bar{c}_r^k a_1^r$, respectively.

The way the spline is defined allows to ensure that, for the representer values $\bar{c}_r^k$ of each class, the membership function returns the mean assessment given by the subjects to that class (i.e., $\mathcal{T}_k(\bar{c}_r^k) = \bar{v}_r^k$). From this point of view, $f^r(x)$ may

Table 1. Number of classes, grouped clusters, representers of the obtained classes and RMSE found by applying the membership function related to each measure

Measures	NC_k	Classes	$(\bar{c_4}, \bar{v_4})$	$(\bar{c_3}, \bar{v_3})$	$(\bar{c_2}, \bar{v_2})$	$(\bar{c_1}, \bar{v_1})$	RMSE
Tamura[8]	3	{1-3,4-8,9}	-	(2.76,1)	(3.28,0.375)	(3.61,0)	0.239
FD[9]	3	{1-3,4-8,9}	-	(3.04,1)	(2.65,0.38)	(2.39,0)	0.239
ED[10]	4	{1,2-4,5-8,9}	(0.374,1)	(0.34,0.75)	(0.31,0.31)	(0.269,0)	0.280
Correlation[11]	3	{1,2-8,9}	-	(0.154,1)	(0.579,0.5)	(0.817,0)	0.283
Weszka[12]	3	{1,2-8,9}	-	(0.153,1)	(0.102,0.5)	(0.051,0)	0.364
Wu[13]	3	{1,2-8,9}	-	(31.5,1)	(21.2,0.5)	(10.5,0)	0.366
Amadasun[14]	2	{1-8,9}	-	-	(11.1,1)	(30.8,0)	0.469
LH[11]	2	{1-8,9}	-	-	(0.102,1)	(0.202,0)	0.481
DGD[15]	2	{1,2-9}	-	-	(14.5,1)	(4.1,0)	0.487
SRE[16]	2	{1-8,9}	-	-	(0.986,1)	(0.933,0)	0.487
SNE[17]	2	{1-8,9}	-	-	(0.785,1)	(0.564,0)	0.493
Newsam[18]	2	{1-8,9}	-	-	(19.1,1)	(17.4,0)	0.512
Entropy[11]	2	{1-8,9}	-	-	(9.02,1)	(8.42,0)	0.532
Uniformity[11]	2	{1-8,9}	-	-	(0.0002,1)	(0.0004,0)	0.566
Contrast[11]	1	-	-	-	-	-	-
Variance[11]	1	-	-	-	-	-	-
FMPS[19]	1	-	-	-	-	-	-
Abbadeni[2]	1	-	-	-	-	-	-

be considered as a function that represents the transition between the classes C_r^k and C_{r+1}^k.

The above approach has been used to define the membership functions $\mathcal{T}_k$, with $k = 1, \ldots, 18$ corresponding to the 18 measures considered in this paper. To do it, the class representer values $\bar{c}_r^k$ and the corresponding assessment values $\bar{v}_r^k$ shown in table 1 have been used. These functions have been applied to each image $I_i \in \mathcal{I}$ and the obtained value has been compared with the one assessed by human subjects. Table 1 shows the RMSE obtained for the different fuzzy sets considered in this paper. This table has been sorted according to least RMSE, where the less the RMSE, the better the performance of the fuzzy set.

3 Results

The function $\mathcal{T}_k$ obtained for each measure (defined by the parameter values shown in table 1) has been applied to different real images. Table 2 shows three real images with different perception degree of fineness. For each image and each measure, this table shows the value obtained by the related function and the error obtained when comparing this value with the assessment value given by subjects (by computing the difference between both of them). It can be noticed by looking the results shown in this table that our model allows to represent appropriately the perception of fineness. Note that the values given by the different functions are similar to the corresponding assessment degree for most of the obtained functions. This is particularly noticed for extreme fineness degrees of presence

Table 2. Estimated and error values obtained by applying the proposed model to three real images

	 Assessment Value=0		 Assessment Value=0.5		 Assessment Value=1	
Measure	Value	Error	Value	Error	Value	Error
Tamura	0	0	0.510	0.010	1	0
FD	0	0	0.641	0.141	1	0
ED	0	0	0.406	0.094	1	0
Correlation	0	0	0.519	0.019	1	0
Weszka	0	0	0.562	0.062	0.959	0.041
Wu	0	0	0.620	0.120	1	0
Amadasun	0	0	0.976	0.476	1	0
LH	0	0	1.000	0.500	1	0
DGD	0	0	0	0.500	0.952	0.048
SRE	0	0	1.000	0.500	1	0
SNE	0	0	0.929	0.429	1	0
Newsam	0	0	0.983	0.483	1	0
Entropy	0.030	0.030	1.000	0.500	1	0
Uniformity	0.518	0.518	1.000	0.500	1	0

(first and last columns in table 2) while for intermediate degrees (central column in table 2) higher errors are obtained, specially for those measures found in the last rows (corresponding to measures with high RMSE in table 1).

Table 3 shows a comparative between our model and the assessments obtained from subjects for a mosaic image. The first column of this table shows the mosaic image made by several images, each one with a different increasing perception degree of fineness. The second column shows the number of the subimage considered from left to right and from top to bottom. The third column shows the assessments given by humans for the different images. The fourth column shows the perception degree of fineness obtained by applying our model using the Tamura measure (the one with least RMSE according to table 1). The fifth column shows the difference between the computed degree and the human assessment. In the case of the sixth column we calculate the differences between the assessment given by each subject and the computed degree, and we obtain as error measure the mean from these 20 differences. Finally, the average errors shown in the last row with values of 0.011 and 0.110 show the goodness of our approach to represent the subjectivity found in fineness perception. It can be noticed that our model captures the evolution of the perception degrees of fineness. Let us remark that when very fine textures are found, the assigned value is

Table 3. Errors obtained from a mosaic image for the Tamura measure

	Image	Human Assessment(H)	Estimated Value (V)	Error #1 ($\|H-V\|$)	Error #2
	1	0	0	0	0
	2	0.125	0.089	0.036	0.057
	3	0.250	0.237	0.013	0.023
	4	0.375	0.376	0.001	0.094
	5	0.500	0.501	0.001	0.011
	6	0.625	0.626	0.001	0.053
	7	0.750	0.948	0.198	0.562
	8	0.875	1	0.125	0.135
	9	0.960	1	0.040	0.052
				Avg: 0.011	Avg: 0.110

1 which is due to the fact that the Tamura measure had grouped the first classes into one cluster.

4 Conclusions and Future Works

In this paper, a methodology to represent the fineness concept by means of fuzzy sets has been proposed. To define the membership function associated to the fuzzy set, the functional relationship between a given measure (automatically computed over the image) and the perception degree of fineness has been achieved. In order to obtain the perception degree of fineness, a group of human subjects have been polled and their assessments have been aggregated by means of OWA operators. The functions presented in this paper allow to decide whether there is fine, coarse or intermediate texture, splitting the domain into some perceptual meaningful intervals. The results given by our approach show a high level of connection with the assessments given by subjects.

As future work, we will extend the proposed approach to obtain membership functions on $\mathbb{R}^n$, i.e. functions on vectors of measures. Furthermore, the performance of the fineness functions will be analyzed in applications like textural classification or segmentation.

References

1. Shapiro, L.G., Stockman, G.: Image Segmentation. In: Computer Vision. (2001) 297–301
2. Abbadeni, N., Ziou, N., Wang, D.: Autocovariance-based perceptual textural features corresponding to human visual perception. In: Proc. of 15th International Conference on Pattern Recognition. Volume 3. (2000) 901–904
3. Reed, T.R., Buf, J.H.D.: A review of recent texture segmentation and feature extraction techniques. CVGIP: Image Understanding **57**(3) (1993) 359–372

4. Wang, C., Wang, H., Mei, Q.: Texture segmentation based on an adaptively fuzzy clustering neural network. In: Proc. of 2004 International Conference on Machine Learning and Cybernetics. Volume 2. (2004) 1173–1176
5. Potrebić, M.: Iterative fuzzy rule base technique for image segmentation. In: Proc. of 7th seminar on Neural Network Applications in Elecrical Engineering. (2004) 221–224
6. Yager, R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man and Cybernetics **18**(1) (1988) 183–190
7. Hochberg, Y., Tamhane, A.: Multiple Comparison Procedures. Wiley (1987)
8. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetics **8** (1978) 460–473
9. Peleg, S., Naor, J., Hartley, R., Avnir, D.: Multiple resolution texture analysis and classification. IEEE Transactions on Pattern Analysis and Machine Intelligence (4) (1984) 518–523
10. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **8**(6) (1986) 679–698
11. Haralick, R.: Statistical and structural approaches to texture. Proceedings IEEE **67**(5) (1979) 786–804
12. Weszka, J., Dyer, C., Rosenfeld, A.: A comparative study of texture measures for terrain classification. IEEE Transactions on Systems, Man and Cybernetics **6** (1976) 269–285
13. Wu, C., Chen, Y.: Statistical feature matrix for texture analysis. CVGIP: Graphical Models and Image Processing **54**(5) (1992) 407–419
14. Amadasun, M., King, R.: Textural features corresponding to textural properties. IEEE Transactions on Systems, Man and Cybernetics **19**(5) (1989) 1264–1274
15. Kim, S., Choi, K., Lee, D.: Texture classification using run difference matrix. In: Proc. of IEEE 1991 Ultrasonics Symposium. Volume 2. (1991) 1097–1100
16. Galloway, M.: Texture analysis using gray level run lengths. Computer Graphics and Image Processing **4** (1975) 172–179
17. Sun, C., Wee, W.: Neighboring gray level dependence matrix for texture classification. Computer Vision, Graphics and Image Processing **23** (1983) 341–352
18. Newsam, S., Kammath, C.: Retrieval using texture features in high resolution multi-spectral satellite imagery. In: Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, SPIE Defense and Security. (2004)
19. Yoshida, H., Casalino, D., Keserci, B., Coskun, A., Ozturk, O., Savranlar, A.: Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images. Physics in Medicine and Biology **48** (2003) 3735–3753

Pattern Classification Model with T-Fuzzy Data

Cao Bing-yuan

School of Mathematics and Information Science
Guangzhou University, Guangzhou 510006, China
caobingy@163.com, bycao@gzhu.edu.cn

Abstract. First, a new pattern classification model with T-fuzzy data is built on the basis of definition in T-fuzzy data and of its operation properties. Besides, T-fuzzy data are determined to measure T-fuzzy numbers by using a distance formula, so that a model is also decided. Meanwhile four methods to the model are presented, which are applied to pattern classification and recognition of environmental quality. The result gained from a sample coincides with practice, providing a valid method for a pursuit of indeterminacy problems in human imitation identification, because the sample contains more information, where a little training results from the sample is integrated with estimation from expert's experience, and information resources are made full use of in analysis of confirmation. Finally, we give some method with which T-fuzzy data is obtained.

1 Introduction

A pattern classification means pointing out a similar sort of object to a given object, which needs not only a good command of mathematical knowledge, but also possession of special knowledge and certain experiments as well. It is difficult for us to give an exact description of the sample in complicated phenomenon in the realistic world, and the data obtained by means of quantifying are of approximate value to a certain degree, therefore, there is no doubt that a classical mathematical method may lose a lot of information. It is well known that the concept of a fuzzy set, first originated from the study of problems, is related to pattern classification and thereafter, there appear many recognition methods by the inspiration of fuzzy sets. Many authors [2,5,11] have developed fuzzy matching pattern and various distance models between fuzzy sets (including fuzzy data). Because $T-$fuzzy data are fuzzy sets with points fixed and circles round them changed, it can perfectly imitate person's thoughts. Once,author applied the pattern classification method to recognition of the humanity fossil and children's health growth, respectively, the results of which prove satisfactory. This paper aims at inducing T-fuzzy data [1,6] into pattern classification model before developing a new model different from the methods gained from traditional ones and from any other fuzzy methods to pattern classification [2]. It will be of very value in the research of classification and recognition in complex artificial intelligence and nerve network, based on classification systems designed by some principles, and examples testify it in the paper.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 793–802, 2007.
springerlink.com

2 Some Definitions and Properties

Definition 1. *Suppose* $\tilde{x} = (x, \alpha, \beta)_{LR}$ *to be* $L - R$ *fuzzy numbers [1,4,7], where* x *represents main value of* $\tilde{x}$*, and* α,β *the left and the right extension of* $\tilde{x}$*, respectively, and if both* L,R *are functions*

$$T(u) = \begin{cases} 1-u, & if \quad 0 \le u \le 1, \\ 0, & if \quad otherwise, \end{cases}$$

i.e.

$$\mu_{\tilde{x}}(u) = \begin{cases} 1 - \dfrac{x-u}{\alpha}, & if \quad x - \alpha \le u \le x, \\ 1 - \dfrac{u-x}{\beta}, & if \quad x \le u \le x + \beta, \quad \alpha, \beta > 0 \\ 0, & if \quad otherwise, \end{cases}$$

then $\tilde{x}$ *is called* T*-fuzzy number, it is represented by* $\tilde{x} = (x, \alpha, \beta)_T$ *and the whole is represented by* $\mathcal{T}(\mathcal{R})$*, where* $\mathcal{R}$ *is a set of real numbers.*

Definition 2. T*-fuzzy number* $\tilde{x}$ *is convex and regular fuzzy subset on real number axis* $\mathcal{R}$*, such that,*

(a) $\exists : u_0 \in U, \mu_{\tilde{x}}(u_0) = 1;$

(b) $\mu_{\tilde{x}}(u)$ *is a function of piecewise continuity.*

Definition 3. *Let* $\tilde{x}_1 = (x_1, \underline{\xi}_1, \overline{\xi}_1)_T$, $\tilde{x}_2 = (x_2, \underline{\xi}_2, \overline{\xi}_2)_T$ *be* $T-$*fuzzy numbers, where* $x_1, x_2 \in \mathcal{R}; \underline{\xi}_1, \overline{\xi}_1, \underline{\xi}_2, \overline{\xi}_2 \in \mathcal{R}^+$*. Then their operation can be defined as:*

1) $\tilde{x}_1 + \tilde{x}_2 = (x_1 + x_2, \underline{\xi}_1 + \underline{\xi}_2, \overline{\xi}_1 + \overline{\xi}_2)_T;$

2) *To* $\forall \lambda \in \mathcal{R}$, $\lambda\tilde{x} = \begin{cases} (\lambda x, \lambda\underline{\xi}, \lambda\overline{\xi})_T, & if \quad \lambda \ge 0, \\ (\lambda x, -\lambda\overline{\xi}, -\lambda\underline{\xi})_T, & if \quad \lambda < 0; \end{cases}$

3) $\tilde{x}_1 - \tilde{x}_2 = \tilde{x}_1 + (-\tilde{x}_2) = (x_1 - x_2, \underline{\xi}_1 + \overline{\xi}_2, \overline{\xi}_1 + \underline{\xi}_2)_T.$

Definition 4. *[6] Let* $\mathcal{R}^n$ *be* $n-$*dimensional Euclidean space. Then the space of nonempty compact convex subsets in* $\mathcal{R}^n$ *can be embedded in a Banach space by identification with support function and a defined* L_2*-metric. If* $\tilde{A} = [\underline{A}, \overline{A}]$, $\tilde{B} = [\underline{B}, \overline{B}]$ *are two compact intervals, this metric reduces to*

$$D_2(\tilde{A}, \tilde{B})^2 = (\underline{A} - \underline{B})^2 + (\overline{A} - \overline{B})^2.$$

Suppose that $\tilde{x} = (m(x), L_1, R_1)$, $\tilde{y} = (m(y), L_2, R_2)$*, the metric* d *on* $\mathcal{T}(\mathcal{R})$ *is defined as*

$$d(\tilde{x}, \tilde{y})^2 = D_2(\text{supp}x, \text{supp}y)^2 + (m(x) - m(y))^2,$$

where $\text{supp}(\cdot)$ *denotes the support of* $(\cdot)$ *and* $m(\cdot)$ *denotes the main value of* $(\cdot)$*.*

Especially, when $\tilde{x} = (x, \underline{\xi}, \overline{\xi})_T$, $\tilde{y} = (y, \underline{\eta}, \overline{\eta})_T,)$, the metric of $\tilde{x}, \tilde{y}$ is written d and defined as

$$d(\tilde{x}, \tilde{y}) = [(x - y - (\underline{\xi} - \underline{\eta}))^2 + (x - y + (\overline{\xi} - \overline{\eta}))^2 + (x - y)^2]^{\frac{1}{2}}. \qquad (1)$$

Let $\mathcal{P}(\mathcal{R})$ be that subspace of $\mathcal{T}(\mathcal{R})$ consisting of all those elements having non-negative support: for each $(x, \underline{\xi}, \overline{\xi}) \in \mathcal{P}(\mathcal{R})$, $x - \underline{\xi} \geq 0$. Then $\mathcal{P}(\mathcal{R})$ is a cone in $\mathcal{T}(\mathcal{R})$ and a closed convex subset of $\mathcal{T}(\mathcal{R})$ with respect to the topology induced by metric d. Here $\tilde{x}, \tilde{y} \in \mathcal{P}(\mathcal{R}), x, y \in \mathcal{R}$.

Proposition 1. *Suppose that $\mathcal{T}(\mathcal{R})$ represents the whole T-fuzzy point sets defined on an ordinary set $\mathcal{R}$, then $d(\tilde{x}, \tilde{y})$ is a distance.*

Proof. From Formula (1),
(a) $d(\tilde{x}, \tilde{y}) > 0$ and (b) $d(\tilde{x}, \tilde{y}) = d(\tilde{y}, \tilde{x}), \forall \tilde{x}, \tilde{y} \in \mathcal{T}(\mathcal{R})$ are obvious.
We only prove (c) and (d) as follows:

$$(c) d(\tilde{x}, \tilde{y}) = 0 \Longleftrightarrow x = y, \underline{\xi} = \underline{\eta}, \overline{\xi} = \overline{\eta}.$$

From (1) we know, $\quad (x - y - (\underline{\xi} - \underline{\eta}))^2 + (x - y - (\overline{\xi} - \overline{\eta}))^2 + (x - y)^2 = 0,$

then (c) is hold.
(d) $\quad d(\tilde{x}, \tilde{y}) \leq d(\tilde{x}, \tilde{z}) + d(\tilde{y}, \tilde{z})$.

$\forall \tilde{x}, \tilde{y}, \tilde{z} \in \mathcal{P}(\mathcal{R})$ and for $\tilde{x} \subseteq \tilde{y} \subseteq \tilde{z}$.

$$\begin{aligned}
\because \quad d(\tilde{x}, \tilde{y}) &= [(x - y - (\underline{\xi} - \underline{\eta}))^2 + (x - y + (\overline{\xi} - \overline{\eta}))^2 + (x - y)^2]^{\frac{1}{2}} \\
&= [(x - y - z + z - (\underline{\xi} - \underline{\eta}) - \underline{\zeta} + \underline{\zeta})^2 \\
&\quad + (x - y - z + z + (\overline{\xi} - \overline{\eta}) - \overline{\zeta} + \overline{\zeta})^2 + (x - y - z + z)^2]^{\frac{1}{2}} \\
&\leq \{[(x - z) - (\underline{\xi} - \underline{\zeta})]^2 + [(y - z) + (\underline{\eta} - \underline{\zeta})]^2 + (x - y)^2 \\
&\quad + [(x - z) + (\overline{\xi} - \overline{\zeta})]^2 + [(y - z) - (\overline{\eta} - \overline{\zeta})]^2 + (y - z)^2\}^{\frac{1}{2}} \\
&\leq \{[(x - z) - (\underline{\xi} - \underline{\zeta})]^2 + [(x - z) + (\overline{\xi} - \overline{\zeta})]^2 + (x - z)^2\}^{\frac{1}{2}} \\
&\quad + \{(y - z + (\underline{\eta} - \underline{\zeta}))^2 + (y - z - (\overline{\eta} - \overline{\zeta}))^2 + (y - z)^2\}^{\frac{1}{2}} \\
&= d(\tilde{x}, \tilde{z}) + d(\tilde{y}, \tilde{z}),
\end{aligned}$$

$\therefore$ (d) holds.
Hence, $d(\tilde{x}, \tilde{y})$ is a distance.
Especially, when $\tilde{d} \in [0, 1]$, we have

$$\tilde{d}(\tilde{x}, \tilde{y}) \leq \max\{\tilde{d}(\tilde{x}, \tilde{z}), \tilde{d}(\tilde{z}, \tilde{y})\}.$$

3 Constitution of Pattern Classification Model with *T*-Fuzzy Data

As for a pattern classification model, we have methods such as statistics and language. As for model building, we have fuzzy set methods such as threshold value, experience and dual opposite comparative functions [2], whose steps are concluded as follows.

Step 1. Feature collection

To the object set $u_i = (u_{i1}, u_{i2}, \cdots, u_{in}) \in U (i = 1, \cdots, m)$, we collect and recognize the collective peculiarities concerned, test the data in its feature description: $\tilde{x}_{ij} = (x_{ij}, \underline{\xi}_{ij}, \overline{\xi}_{ij})_T (i = 1, \cdots, m; j = 1, \cdots, n)$, where the extension can be chosen as functions $\max\{0, 1 - x_{ij}\}$.

Step 2. Variation pattern

Change u_i into T-fuzzy number pattern $p(u_i) = (u_i^1, u_i^2, \cdots, u_i^n)$, meanwhile, we give a standard object v_0 and determine its pattern, which is a vector $p(v_0) = (v_0^1, v_0^2, \cdots, v_0^n)$ with respect to v_0, and its assigned value with corresponding feature is $\tilde{y}_0 = (y_0, \underline{\eta}_0, \overline{\eta}_0)_T$.

Step 3. Nonfuzzyfication

We handle nonfuzzyfication of sample $\tilde{x}_{ij}$ and $\tilde{y}_0$, calculate metric value d_{ij} between $\tilde{x}_{ij}$ and $\tilde{y}_0$ by the aid of distance Formula (1).

Step 4. Methods

A. Methods of experience value

Properly determine an experience value $\alpha \in \mathcal{R}$. If d_{ij} is as Formula (1), when $d_{ij} \leq \alpha$, a recognition is accepted, and when $d_{ij} > \alpha$, a recognition is refused.

B. Method of a dual opposite comparison function [9]

Suppose

$$\mu_{v_0}(u_{ij}) = \tilde{d}_{ij}(\tilde{x}_{ij}/\tilde{y}_0) \triangleq \frac{d_{\tilde{x}_{ij}}(\tilde{y}_0)}{[d_{\tilde{x}_{ij}}(\tilde{y}_0) + d_{\tilde{y}_0}(\tilde{x}_{ij})]}, (i \neq j); \tag{2}$$

or

$$\mu_{v_0}(u_{ij}) = \frac{d_{\tilde{x}_{ij}}(\tilde{y}_0)}{[d_{\tilde{x}_{ij}}(\tilde{y}_0) \bigvee d_{\tilde{y}_0}(\tilde{x}_{ij})]}, (i \neq j), \tag{3}$$

where $\mu_{v_0}(u_{ij}) \in [0,1], (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$ and list the reaching relation matrix orders, we calculate

$$\theta = \max\{\mu_{v_0}(u_{i1}), \mu_{v_0}(u_{i2}), \cdots, \mu_{v_0}(u_{in}), (i = 1, 2, \ldots, m)\}.$$

If $\mu_{v_0}(u_{k_0}) = \theta$, we can determine the k_0-th sample u_{k_0} most similar to a standard one v_0.

C. Methods of threshold value

Consider Formal (2) or (3), properly select determination of a threshold value $\lambda \in [0,1]$. If $\mu_{v_0}(u_{ij}) \geq \lambda$, a recognition is accepted, and when $\mu_{v_0}(u_{ij}) < \lambda$, a recognition is refused.

If there exists m feature influencing u_i corresponding to v, feature values of u_i and v_i are $\tilde{x}_{ij}(i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$ and $\tilde{y}$, respectively, then the metric in $\tilde{x}_{ij}$ and $\tilde{y}$ shall be weighted, i.e.,

$$d_i = \begin{cases} \sum\limits_{j=1}^{m} k_{ij} d(\tilde{x}_{ij}, \tilde{y}), & i \neq j; \\ 0, & i = j \end{cases}$$

where $k_{ij} \geq 0$ and $\sum_{j=1}^{m} k_{ij} = 1 (i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$.

The author advances another new recognition classification method here, and now it is introduced as follows.

D. Concrete pattern classification

Obviously, we consider and use variety of information in it, the steps of which are shown as follows.

Step 1'. Feature collection and Step 2'. Variation pattern, similar to Step 1 and Step 2 above (omitted).

Step 3'. Nonfuzzyfication

We calculate metric value between v_0 and u_i, that is, we calculate metric value d_{ij} by the aid of distance Formula (1), then

$$(d_{ij})_{m\times n} = \begin{pmatrix} d(\tilde{x}_{11},\tilde{y}_0) & d(\tilde{x}_{12},\tilde{y}_0) & \cdots & d(\tilde{x}_{1n},\tilde{y}_0) \\ d(\tilde{x}_{21},\tilde{y}_0) & d(\tilde{x}_{22},\tilde{y}_0) & \cdots & d(\tilde{x}_{2n},\tilde{y}_0) \\ \cdots & \cdots & \cdots & \cdots \\ d(\tilde{x}_{m1},\tilde{y}_0) & d(\tilde{x}_{m2},\tilde{y}_0) & \cdots & d(\tilde{x}_{mn},\tilde{y}_0) \end{pmatrix}, \tag{4}$$

here $d_{ij} = d(\tilde{x}_{ij},\tilde{y}_0)(i = 1,\cdots,m; j = 1,\cdots,n)$ is distance between $\tilde{x}_{ij}$ and $\tilde{y}_0$.

Step 4'. Optimum decision [10].

Take maximum, minimum and average value to $d(\tilde{x}_{ij},\tilde{y}_0)$ in (4) according to line and we have

$$\begin{aligned} &d_i^+ = \sup\{d_{i1}, d_{i2},\cdots,d_{in}\}, d_i^- = \inf\{d_{i1}, d_{i2},\cdots,d_{in}\}, \\ &\bar{d}_i = \frac{d_{i1}+d_{i2}+\cdots+d_{in}}{n}, (i = 1,\cdots,m), \end{aligned} \tag{5}$$

then an assessable matrix from Formal (5) will be obtained:

$$\begin{matrix} \text{I} \\ \text{II} \\ \vdots \\ \text{m} \end{matrix} \begin{pmatrix} d_1^+ & d_1^- & \bar{d}_1 \\ d_2^+ & d_2^- & \bar{d}_2 \\ \vdots & \vdots & \vdots \\ d_m^+ & d_m^- & \bar{d}_m \end{pmatrix}.$$

Again, suppose

$$f(d_i^+ + d_i^- + \bar{d}_i) = \alpha_1 d_i^+ + \alpha_2 d_i^- + \alpha_3 \bar{d}_i,$$

here, $\alpha_l \in [0,1](l = 1,2,3)$ represents a weight number obtained by Analytic Hierarchy Process or Delphi method.

Finally, we take $f_k = \min f(*_i)$, where " $*_i$ " represents "$d_i^+ + d_i^- + \bar{d}_i$", $(i = 1,\cdots,m)$. This indicates k-object u_k presses closest to a standard one v_0.

Step 5'. Further decision

Sometimes, in order to describe the similar degree between u_k and v_0 at length, we can adopt the following decision.

Compare to all $d_{ij}(i = 1,2,\ldots,m; j = 1,2,\ldots,n)$, suppose $D^+ = \sup_{i,j} d_{ij}$, $D^- = \inf_{i,j} d_{ij}$, then

$$\mu_{p(u_1)}(v_0) = 1 - \frac{f(*_1) - D^-}{D^+ - D^-},\cdots,\mu_{p(u_m)}(v_0) = 1 - \frac{f(*_m) - D^-}{D^+ - D^-}.$$

Take $\mu_{p(u_k)}(v_0) = \max\{\mu_{p(u_1)}(v_0),\ldots,\mu_{p(u_m)}(v_0)\}$, then, $\mu_{p(u_k)}(v_0)$ is the membership degree in k-object u_k belonging to a standard one v_0.

4 Practical Example

In environmental protection, we always distinguish the grade of environmental quality [8]. Now, we divide it into I–V grades as follows. $U = \{I, II, III, IV, V\} = \{\text{clean, less clean, less polluted, more polluted, most polluted}\}$, where U is a universe. We choose atmosphere, water on and under the earth as u_1, u_2 and u_3 respectively for environmental factors according to monitoring data at 10 observation points in some city, and their target sets are

$u_1 = \{SO_2, NO_x, TSP\}; u_2 = \{COD, NH_{3-}N, DO, NO_{3-}N, Cr^{+6}, CN\};$

$u_3 = \{SO_4^{3-}, Cl^-, NO_{3-}N, \text{earth hardness}, Cr^{+6}, CN\},$

respectively. Reference [8] gives a standard value in distinguishing the grade of environmental quality as Table 1.

Table 1. Standard value in distinguish the grade of environmental quality

		distinguish the grade				
factors	index	I	II	III	IV	V
atmos-phere	SO_2	≤ 0.05	$> 0.05 \backsim\leq 0.15$	$> 0.15 \sim\leq 0.25$	$> 0.25 \sim\leq 0.50$	> 0.50
	NO_x	≤ 0.05	$> 0.05 \backsim\leq 0.10$	$> 0.10 \sim\leq 0.15$	$> 0.15 \sim\leq 0.30$	> 0.30
	TSP	≤ 0.15	$> 0.15 \backsim\leq 0.30$	$> 0.30 \sim\leq 0.50$	$> 0.50 \sim\leq 0.75$	> 0.75
water on the earth	COD	≤ 2	$> 2 \backsim\leq 6$	$> 6 \sim\leq 12$	$> 12 \sim\leq 25$	> 25
	$NH_{3-}N$	≤ 0.25	$> 0.25 \backsim\leq 0.50$	$> 0.50 \sim\leq 1$	$> 1 \sim\leq 3$	> 3
	DO	≥ 8	$\geq 4 \backsim< 8$	$\geq 3 \backsim< 4$	$\geq 2 \backsim< 3$	< 2
	$NO_{3-}N$	≤ 10	$> 10 \backsim\leq 20$	$> 20 \sim\leq 40$	$> 40 \sim\leq 80$	> 80
	C_r^{+6}	≤ 0.01	$> 0.01 \backsim\leq 0.05$	$> 0.05 \sim\leq 0.10$	$> 0.10 \sim\leq 0.25$	> 0.25
	CN	≤ 0.01	$> 0.01 \backsim\leq 0.05$	$> 0.05 \sim\leq 0.10$	$> 0.10 \sim\leq 0.25$	> 0.25
water under the earth	SO_4^{3-}	≤ 120	$> 120 \backsim\leq 250$	$> 250 \sim\leq 750$	$> 750 \sim\leq 1000$	> 1000
	Cl^-	≤ 120	$> 120 \backsim\leq 250$	$> 250 \sim\leq 750$	$> 750 \sim\leq 1000$	> 1000
	$NO_{3-}N$	≤ 10	$> 10 \backsim\leq 20$	$> 20 \backsim\leq 40$	$> 40 \backsim\leq 80$	> 8
	hardness	≤ 250	$> 250 \backsim\leq 450$	$> 450 \sim\leq 650$	$> 650 \sim\leq 900$	> 900
	C_r^{+6}	≤ 0.01	$> 0.01 \backsim\leq 0.05$	$> 0.05 \sim\leq 0.10$	$> 0.10 \sim\leq 0.25$	> 0.25
	CN	≤ 0.01	$> 0.01 \backsim\leq 0.05$	$> 0.05 \sim\leq 0.10$	$> 0.10 \sim\leq 0.25$	> 0.25

Unit: atmosphere—mg/Nm^3, water on and under the earth —mg/l; hardness— counting by calcium carbonate(mg/l).

Step 1'. Feature collection, Step 2'. Variation pattern.

Let $A = \{u_1, u_2, u_3\}$, and T-fuzzy number sets corresponding to the five standard degrades in U are as follows.

$$\begin{aligned}
A_I = \{&(0.05, 0, 0)_T, (0.05, 0, 0)_T, (0.15, 0, 0)_T; (2, 0, 0)_T, (0.25, 0, 0)_T, \\
&(8, 0, 0)_T, (10, 0, 0)_T, (0.01, 0, 0)_T, (0.01, 0, 0)_T; (120, 0, 0)_T, \\
&(120, 0, 0)_T, (10, 0, 0)_T, (250, 0, 0)_T, (0.01, 0, 0)_T, (0.01, 0, 0)_T\}. \\
A_{II} = \{&(0.1, 0.04, 0.05)_T, (0.08, 0.02, 0.01)_T, (0.2, 0.02, 0.1)_T; (4, 1.8, 2)_T, \\
&(0.35, 0.1, 0.05)_T, (6, 2, 1.5)_T, (15, 4.5, 5)_T, (0.03, 0.02, 0.01)_T,
\end{aligned}$$

$$(0.03, 0.02, 0.01)_T; (190, 65, 60)_T, (190, 65, 60)_T, (15, 4.5, 5)_T,$$
$$(360, 110, 90)_T, (0.03, 0.02, 0.01)_T, (0.03, 0.02, 0.01)_T\}.$$
$$A_{III} = \{(0.2, 0.01, 0.05)_T, (0.12, 0.02, 0.01)_T, (0.4, 0.1, 0.05)_T; (9, 2, 3)_T,$$
$$(0.7, 0.05, 0.2)_T, (3.5, 0.5, 0.3)_T, (30, 9, 10)_T, (0.08, 0.01, 0.01)_T,$$
$$(0.08, 0.02, 0.02)_T; (510, 260, 240)_T, (510, 260, 240)_T, (30, 9, 10)_T,$$
$$(560, 110, 90)_T\}; (0.08, 0.01, 0.01)_T, (0.08, 0.02, 0.02)_T\}$$
$$A_{IV} = \{(0.38, 0.12, 0.1)_T, (0.22, 0.05, 0.08)_T, (0.62, 0.10, 0.12)_T;$$
$$(2, 0.8, 1)_T, (2.5, 0.5, 0.1)_T, (60, 18, 20)_T, (0.18, 0.07, 0.05)_T,$$
$$(0.18, 0.07, 0.05)_T; (880, 130, 120)_T, (880, 130, 120)_T, (60, 18, 20)_T,$$
$$(780, 130, 120)_T, (0.18, 0.07, 0.05)_T, (0.18, 0.07, 0.05)_T\}.$$
$$A_V = \{(0.5, 0, 0)_T, (0.3, 0, 0)_T, (0.75, 0, 0)_T; (25, 0, 0)_T, (3, 0, 0)_T,$$
$$(2, 0, 0)_T, (80, 0, 0,)_T, (0.25, 0, 0)_T, (0.25, 0, 0)_T; (1000, 0, 0)_T,$$
$$(1000, 0, 0)_T, (80, 0, 0)_T, (900, 0, 0,)_T, (0.25, 0, 0)_T, (0.25, 0, 0)_T\}.$$

Then according to Table 1 [8], we have tested the basic index feature value of environmental quality in a city as Table 2.

Table 2. Basic indexes feature value

factors	atmosphere			water on the earth					
index	SO_2	NO_x	TSP	COD	$NH_{3-}N$	DO	$NO_{3-}N$	C_r^{+6}	CN
feature value	0.07	0.05	0.6	19.2	1.5	5.6	10	0.01	0.01
factors	water under the earth								
index	SO_4^{3-}	Cl^-	$NO_{3-}N$	hardness	C_r^{+6}	CN			
feature value	612	625	14	290	0.01	0.01			

$$\begin{aligned} A_0 &= \{u_1^0, u_2^0, u_3^0\} \\ &= \{(0.06, 0.005, 0.015)_T, (0.05, 0.01, 0.02)_T, (0.5, 0.05, 0.15)_T; \\ &\quad (19, 0.05, 0.4)_T, (2, 0.5, 0.1)_T, (5.5, 0.3, 0.4)_T, (9, 1, 2)_T, \\ &\quad (0.01, 0.005, 0.001)_T, (0.02, 0.001, 0.005)_T; (611, 0.5, 1.5)_T, (624, 1, 2)_T, \\ &\quad (15, 1, 0)_T, (290, 1, 2)_T, (0.01, 0.005, 0.001)_T, (0.02, 0.001, 0.005)_T\}. \end{aligned}$$

Step 3'. Nonfuzzyfication

Matrix $(d_{ij})_{5\times 15}$ is obtained for each component of A_j(j=I,II,III,IV,V) by calculating, the distance between component of them and A_0 with (1)

$$(d_{ij})_{5\times 15} = \begin{pmatrix} 0.0110 & 0.0129 & 0.3279 & 16.8845 & 1.9015 & 2.5495 & 1.8257 & 490.6674 \\ 0.0492 & 0.0238 & 0.3084 & 15.1120 & 1.5465 & 1.1902 & 6.4096 & 425.9724 \\ 0.1511 & 0.0635 & 0.1555 & 9.9593 & 1.1332 & 2.1016 & 21.9924 & 230.2662 \\ 0.3207 & 0.1814 & 0.0956 & 4.3152 & 0.5477 & 3.1691 & 53.2854 & 284.0056 \\ 0.4367 & 0.2470 & 0.2327 & 5.8868 & 1.1633 & 3.5449 & 70.6777 & 388.6676 \end{pmatrix}$$

$$\left.\begin{matrix} 503.6682 & 5.3541 & 39.6863 & 0.0029 & 0.0091 & 0.0029 & 0.0091 \\ 438.8379 & 3.5237 & 102.2823 & 0.0205 & 0.0016 & 0.0205 & 0.0016 \\ 236.2915 & 17.3109 & 275.0667 & 0.0716 & 0.0590 & 0.0716 & 0.0603 \\ 271.748 & 48.4218 & 496.6840 & 0.1712 & 0.1591 & 0.1712 & 0.1591 \\ 375.6687 & 65.3350 & 609.6679 & 0.1712 & 0.1591 & 0.2413 & 0.2287 \end{matrix}\right) .$$

Step 4'. Optimum decision

We will obtain an assessable matrix from Formal (5),

$$\begin{matrix} \mathrm{I} \\ \mathrm{II} \\ \mathrm{III} \\ \mathrm{IV} \\ \mathrm{V} \end{matrix} \begin{pmatrix} d_1^+ & d_1^- & \bar{d}_1 \\ d_2^+ & d_2^- & \bar{d}_2 \\ d_3^+ & d_3^- & \bar{d}_3 \\ d_4^+ & d_4^- & \bar{d}_4 \\ d_5^+ & d_5^- & \bar{d}_5 \end{pmatrix} = \begin{pmatrix} 503.6682 & 0.0029 & 70.8609 \\ 438.8379 & 0.0116 & 66.3501 \\ 236.2915 & 0.0590 & 52.9836 \\ 496.6840 & 0.0956 & 77.5623 \\ 609.6679 & 0.1591 & 101.4886 \end{pmatrix}.$$

We might as well let $\alpha_1 = \alpha_2 = \alpha_3 = \frac{1}{3}$, then

$f_I(*_1) = \frac{1}{3}(503.6682 + 0.0029 + 70.8607) \approx 191.5107$,

$f_{II}(*_2) \approx 168.4, f_{III}(*_3) \approx 96.4447, f_{IV}(*_4) \approx 191.447, f_V(*_5) \approx 237.1052$, here, $f_{III}(*_3) = 96.4447$ is smallest, such that the city's environmental quality approximates III grade, i.e, less polluted, matching practice.

Step 5'. Further decision

Let $D^+ = \sup_{i,j} d_{ij} = 609.6679,\ D^- = \inf_{i,j} d_{ij} = 0.00116$. Then

$$\mu_{p(u_1)}(v_0) = 1 - \frac{191.5107 - 0.00116}{609.6679 - 0.00116} = 0.6859, \mu_{p(u_2)}(v_0) = 0.7238,$$
$$\mu_{p(u_3)}(v_0) = 0.8418, \mu_{p(u_4)}(v_0) = 0.686, \mu_{p(u_5)}(v_0) = 0.6111.$$

Take $\mu_{p(u_3)}(v_0) = \max\{\mu_{p(u_i)}(v_0), i = 1, \ldots, 5\} = 0.8418$,then, the membership degree in 3-th object u_3 belonging to a standard one v_0 is 0.8418, i.e., a membership degree belonging to III is 84.18% in environmental quality of the city.

The result coincides with the 2-order fuzzy synthetical evaluation method in Reference [8]. But

Reference [8] evaluation is as follows

Table 3. Evaluation Result

	distinguish the grade				
index	clean(I)	less clean(II)	pollution(III)	more pollution(IV)	most pollution(V)
result			0.328	0.316	

That is, membership degree belonging to III and IV grade is only 0.008 different. If the evaluation rounds, all become 0.32, thus the conclusion will change in evaluation of total city's environment. If the evaluation process is dealt a little bit dissimilarly, that is, if only 0.008 is reduced in 0.324, 0.324 is changed into 0.316, and the total conclusion of this city environment becomes IV grade and it means medium pollution. If we use this paper's method, such circumstance never appears.

This is because of the following reasons

1) The pattern classification model in this paper contains more information by using T-fuzzy data.

2) Since the smaller weight by more division will make each single factor evaluation meaningless under Zadeh operator " $\wedge$ ", a lot of information loses.

It is enough for us to stop at Step 4' by using the recognition model mentioned here. Now there is no need to compress the result into interval [0,1] because we only use three weight factors, we lose no information. In order to have a detail judgement, we can adopt Step 5'.

5 Obtaining Method in T-Fuzzy Data

The so-called "precision" data are almost approximation of a truth value, existing much in the realistic world. Therefore, it is most important for us to obtain fuzzy data. A construction method in T-fuzzy data is introduced, based on fuzzy time series analysis, as follows.

Suppose what we record only is a group of real numbers for $x_1, x_2, \ldots, x_n$.T-fuzzy number can be constructed by this "accurate" numbers, with its construction steps as follows.

1) Suppose that t period data are influenced by the front and back data each (or two each), and if $M_t = \max\{x_{t-1}, x_t, x_{t+1}\}, m_t = \min\{x_{t-1}x_t, x_{t+1}\}$, then $M_t \geqslant m_t$, at $t = 2, 3, \cdots, N-1$, and

$M_t = \max\{x_1, x_2\}, m_t = \min\{x_1, x_2\},$ at $t = 1$;

$M_t = \max\{x_{N-1}, x_N\}, m_t = \min\{x_{N-1}, x_N\},$ at $t = N$.

2) Let

$$\tilde{y}_t(x) = \begin{cases} 1 - \dfrac{1}{c_t}|x - x_t|, & x \in [m_t, M_t], \\ 0, & x \notin [m_t, M_t], \end{cases}$$

where $c_t = \frac{1}{2}(M_t - m_t),\ x_t = \frac{1}{2}(M_t + m_t), t = 1, 2, \cdots, N$. Or let

$$\tilde{y}_t(x) = \begin{cases} 1 - \dfrac{1}{\sigma}|x - x_t|, & x \in [x_t - \sigma, x_t + \sigma], \\ 0, & x \notin [x_t - \sigma, x_t + \sigma], \end{cases}$$

where σ is a fixed positive number. Also let σ change with t, for example, σ can be taken as c_t.

3) $\tilde{y}_t = (x_t, c_t)$ is composed by x_t with c_t, which is a T-fuzzy number.

$c_t > 0$, its value can be chosen by freely-fixed method with t within interval $[1, N]$ varying according to practical situation.

The usual methods of obtainment T-fuzzy data still are as follows.

A. Direct obtainment.

Record experiments or the measurement's data as fuzzy numbers according to its character.

B. Fitting.

Fit the collected fuzzy data into the distributing function with known fuzzy numbers; the close one is what we beg for.

C. The assignment of information.

D. Structure method and etc.

In application, it is difficult for us to obtain exact data, so an approximate value obtained from tests or measure is regarded as fitting of T-fuzzy data. If historical data were noted down in a more overall way, in fact, these data are fuzzy, even though they were done so precisely, T-fuzzy data can be constructed by a fuzzy time series analysis. If they are not overall or inexact, we can improve a distributive histogram by an information distributive method before obtaining a T-fuzzy data.

6 Conclusion

We can prevent original information from distortion if we describe it with T-fuzzy data. If we use the mentioned 'determination', we can lessen the loss of information, so that the paper's result can be used for classification and recognition in more complicated system and engineering.

Acknowledgements. Supported by National Natural Science Foundation of China (No. 70271047) and by Science Foundation of Guangzhou University.

References

1. Cao, B.Y.: Study on non-distinct self-regression forecast model. Chinese Science Bulletin, Vol.35, **13** (1990) 1057-1062
2. Cao, B.Y.: Cluster and recognition model with T-fuzzy data. Mathematical Statistics and Applied Probability, Vol.11, **4**(1996) 317-325
3. Cao, B.Y.: Fuzzy Geometric Programming. Kluwer Academic Publishers, Dordrecht (2002)
4. Cao, B.Y.: Application of Fuzzy Mathematics and Systems. Science Press, Peking (2005)
5. Cao, B.Y., Liu, G.H.: A new fuzzy recognition model for children's health growth. In Cao B.Y. Editor-in-chief: Proceedings of Result Congress on Fuzzy Sets and Systems: Hunan Press of Science and Technology, Changsha China, 1991 198-200.
6. Diamond, P.: Least squares fitting of several fuzzy data. Proc. of IFSA Congress. Vol.1, (1987) 329-332
7. Dubois, D., Prade, H.: Fuzzy Sets and Systems—Theory and Applications. Academic Press, New York,1980.
8. Lao, Q.T.: Fuzzy comprehensive evaluation to city environmental quality. Chinese Environmental Science, 2(1990)
9. Shimura, M.: Fuzzy sets concept in rank-ordering objects. J. Math. Anal. & Appl., Vol.43, (1973) 717-733
10. Yager, R.R.: Fuzzy sets, Probabilities and Decision. J.Cybern, Vol.10, (1980) 1-18
11. Zadeh, L.A.: Fuzzy Sets and their application to pattern classification and cluster analysis. Multivariate Analysis, (1976) 113-161

Adjusting an Environment to Human Behaviors Based on Biopsy Information

Junzo Watada and Zalili Binti Musa

Waseda University
Graduate School of Information, Production & Systems
2-7 Hibikino, Wakamatsu, Kitakyushu 808-0135 Japan
junzow@osb.att.ne.jp, zalili@kuktem.edu.my

Abstract. It is difficult to define a comfortable space for people. It is partly because comfortness relates to many attributes specify a space, partly because all people have different preferences, and also because even the same person changes his/her preference according the state of their health, body conditions, working states and so on. Various parameters and attributes should be controlled to realize such a comfortable space according the data-base of past usages. Information obtained from human bodies such as temperature, blood pressure, α brainwave, heart beats and etc. can be employed to adjust the space to the best condition. In order to realize a comfortable space, first we recognize human states and behabiors, and second we provide an apporpriate environment for a focal person according to his/her state.

The objective of this paper is to achieve a comfortable space in terms of affectie behaviors or actions. We deal with such techneaques as KANSEI engineering. In other words, it is required to realize our comfortable environment suitable to feelings. Kansei Engineering is one to offer better products to the customers by employing human sensitivity.

1 Introduction

In a today's stress society, comfortable space and life are important to cure strong stress on us. It is in such human science as psychology, human factors, physiology, anthropology, hygiene and so on to study comfortable feeling, and it is in such technology and engineering as mechanical engineering, architecture, control engineering, system engineering, and living science, material science, acoustics, social welfare, and design science to build such a comfortable space. It is not sufficient to combine these separated research results. We should create new total technology to realize a comfortable space.

In this paper, the comfortable space means one where a person feels comfortable when he/she stays there although all people have a different feeling about being comfortable. It is hard to realize a comfortable space for all people. Therefore, we should build a comfortable space from a viewpoint of each person's. It is expected that a machine or a system can recognize and evaluate the state of a person from his/her behavior, voice, measurement and by gathering and recognizing data automatically. If we can measure an electroencephalogram

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 803–812, 2007.
springerlink.com

(brainwaves), heart beats, sweat, saliva, etc. by an instrument, the obtained information is employed to realize a comfortable space.

In order to realize a comfortable space, first we recognize human states and behabiors, and second we provide an apporpriate environment for a focal person in his/her state.

The objective of this paper is to achieve a comfortable space in terms of affectie behaviors or actions. We deal with such techneaques as KANSEI engineering. In other words, the objective of this paper is to show the possibility that the measurement of human senses enables us to realize a comfortable space for any person even if the person changes his/her comfortable feeling according the change of the conditions. In the research, we employ the values of features obtained from electrocardiogram that a test object has in a comfortable environment. Fuzzy control and neural network are used to adjust a space to the most comfortable one as possible in controlling home appliances according the working and resting states as shown as Figure 1

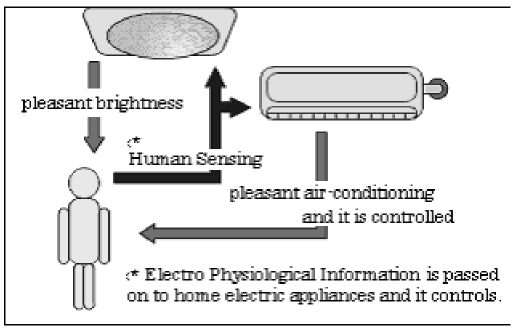


Fig. 1. Control based on Electro Physiological Information

2 Kansei Engineering

Kansei is a Japanese word that means total information of human senses. Kansei Engineering is "a technology, method or theory to translate human Kansei or image to production of real things or to design of objects."It is vague and uncertain that a customer has an image or expectation about some product. Again, Kansei Engineering is a technology to build such Kansei or vague image in product design in some way [4, 5]. The objective of this paper is to employ electrocaridographes obtained from a person so that we can adjust the environment of a space to the most comfortable that he/she feels. These are many biopsy measurements from a human body which can be used for the control of a space. For instance, heart beats, sweat, saliva, etc. are another alternative.

Kansei Engineering is a theory and techniques which employ five senses and use such mechanism in developing products and services. If we successfully employ such senses information in development of products, we can provide more personal use product for individuals.

The most important issue in Kansei Engineering is to measure "Kansei." There are methods of measuring human senses such as sensory test, measurement of a living body and cognitive measurement [4].

2.1 Biometrics

Biometrics is science and technology of measuring and analyzing biological data. In information technology, biometrics refers to technologies that measure and analyze human body characteristics or physical characteristics, such as fingerprints, eye retinas and irises, palm prints, facial structure, and voice recognition are just some of many methods for identification purposes. Therefore, a biometric system is a process using a characteristic of a body as a method for identification since these characteristics are unique to each individual.

Identification by biometric verification is becoming increasingly common in corporate and public security systems, consumer electronics and point of sale (POS) applications [1]. For example, law enforcement agencies for identification of criminals, children, and for licensing of people employed in federally regulated careers such as security brokers. In addition to security, the driving force behind biometric verification has been convenience.

However, there are other available techniques that can be use to identify a human. Based on Roger Clarke, biometric techniques can be dividing into five categories which are 1) appearance (e.g. the familiar passport descriptions of height, weight, color of skin, hair and eyes), 2) social behavior (e.g. habituated body-signals, general voice characteristics and style of speech), 3) bio-dynamics (e.g. the manner in which one's signature is written), 4) natural physiographic(e.g. thumbprint, fingerprint sets and handprints) and 5) imposed physical characteristics (e.g. dog-tags, collars, bracelets and anklets)[2].

2.2 Comfortable Space

Basically, it is hard to realize a comfortable space for all people. Therefore, we should build a comfortable space from a viewpoint of each person's. It is expected that a machine or a system can recognize and evaluate the state of a person from his/her behavior, voice, measurement and by gathering and recognizing data automatically. Therefore, the objective of this paper is to show the possibility that the measurement of human senses enables us to realize a comfortable space for any person even if the person changes his/her comfortable feeling according the change of the conditions [6, 8]. In the research, we employ the values of features obtained from electrocardiogram that a test object has in a comfortable environment. Fuzzy control and neural network are used to adjust a space to the most comfortable one as possible.

There are various spaces in our environment. Also, the space of each room has different objectives even in a residential space. A tall building provides a different space from one in a house. Working space and space in a vehicle as well as residential space are included in an urban space. In future a universe space will be taken into consideration.

It is always essential that such a space provides us with comfort as well as safe and convenience. The objective of our paper is to provide a technique for us such that information obtained from a living body automates the adjustment of temperature, humidity, brightness and wind strength by home appliances. The technique realizes comfortable space for each individual in real time.

3 Electrocardiogram (ECG)

An electrocardiogram is a record that measures electric waves obtained from a heart muscle through electrodes placed on the surface skin of a chest, arms and legs. The wave causes the muscle to squeeze and pump blood from the heart. The activity of the heart is pursued by the sinus node (SN) in the right upper portion of the heart. The electric signal from the sinus node is transferred to a cardiac ventricle from an atrium and makes muscles electrically excite and shrink whose actions pump the blood to flow into blood vessels. The electrocardiogram records such electric actions of the heart as waves.

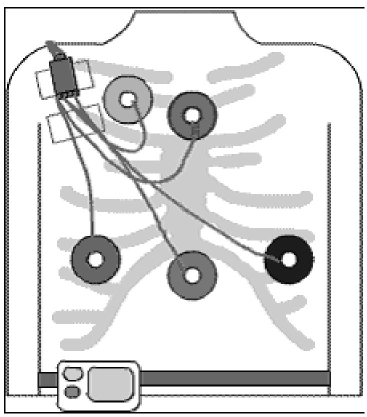


Fig. 2. Position of electrodes

In this experiment Wearable Holster Electrocardiogram which records long term is employed. The holster electrocardiogram can continuously measure and record electrocardiograph. The size of the holster electrocardiogram is light and small which does not weigh an object.

Figure 3 shows an example of a normal electrocardiograph. The electrocardiograph consists of five kinds of waves as PQRST. P waves show excites, QRS waves show excitement of entricles. T waves come out when the entricle cured its excitement. The excitement of an atrium can not measured because their recover waves are quit small and overlapped with QRS waves.

In this experiment, R-R interval that is an interval between QRS waves is employed. One wave of QRS is produced when the heart gives one beat. That is, the R-R interval is a one that the heart gives one beat.

Table 1. Names of Electrode Positions

Code	Location of the leads	
Red	V5 1)	1 Ch
Yellow	Intersection of right mid-clavicular line and first intercostal space.	CM_5 induce
Orange	xiphoid process of sternum	2 Ch
Blue	manubrium of sternum	NASA induce
Black	right V5: 4)	Earth

1)(:At the intersection of left anterior axillary line with a horizontal line through V4)
2)(V4: Intersection of left mid-clavicular line and fifth intercostal space)
3): At the intersection of right anterior axillary line with a horizontal line through right V4
4)(right V4: Intersection of right mid-clavicular line and fifth intercostal space)

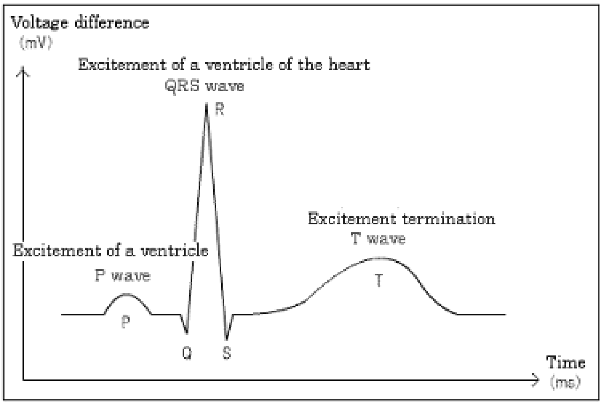


Fig. 3. Illustration of ECG

4 Information Abstract

4.1 Method of Abstracting Information

The experiment consists of rest interval 5 minutes and work interval 10 minutes. These intervals are repeated. In the work interval, test subjects are directed to typewrite some texts in order to give a stress on the test subjects. Figure 4 and Figure 5 shows its results.

Based on Figure 4 and Figure 5, the horizontal axis shows QRS number numbered sequentially which are time-series data. Meanwhile, the vertical axis shows the voltage difference. From that figure, the change between rest and work intervals is done.

4.2 Results

In Figure 4 and Figure 5, vertical axis shows R-R intervals. Horizontal axis shows QRS number that is numbered in the order that QRS wave comes out.

- work interval 10 minutes 1 ∼ 725 steps
- rest interval 5 minutes 726 ∼ 1045 steps
- work interval 10 minutes 1046 ∼ 1743 steps

We can recognize the difference between work and rest intervals in the Figure 4. In order to clarify the difference, the average of QRS numbers in 30 steps is figurized as Figure 5.

The average of 30 steps resulted in the clear difference between work and rest intervals. Using this result, the threshold is decided and the result can be employed to control.

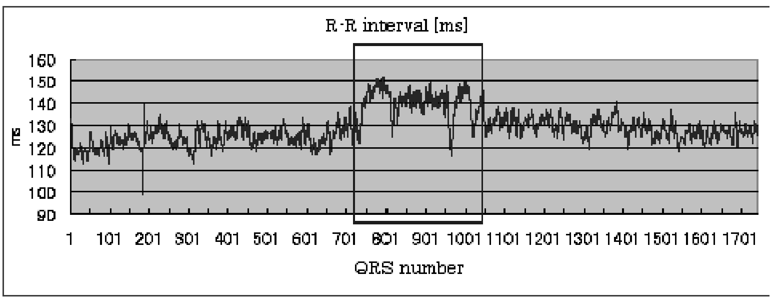


Fig. 4. R-R Interval

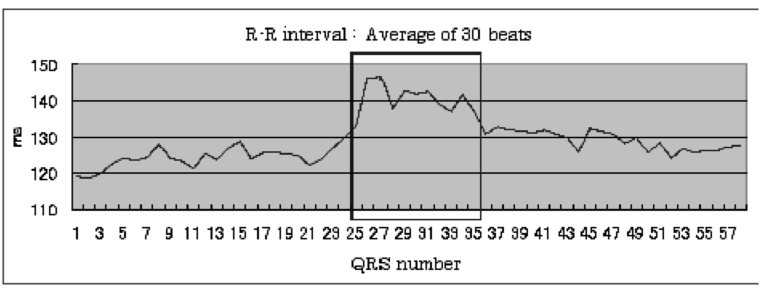


Fig. 5. R-R Interval (average of 30 beats)

5 Fuzzy Control

Fuzzy control is widely employed in industries in early days since the proposal of a fuzzy system in 1965. The fuzzy control is a kind of intelligent control methods. Rules employed in the fuzzy control are written IF-THEN formula and approximately reasoned to decide a control amount. As one rule can cover the wide range of control, it is possible to appropriately control the controlled object using few rules so as that it mimics human operations [7].

Also when the rules are overlapping with each other, the plural rules can compensate the action of control with each other. Even if some rules will not work well, other plural rules can compensate these rules and obtain the expected control.

It is not necessary in a fuzzy controller to model a controlled object, previously. That is, if it is possible to obtain response results, an object can be adjusted perfectly without understanding the controlled object. Of course, it is not possible to write IF-THEN rules without understanding basic response and actions of the controlled object. In this paper we employed the fuzzy control to adjust the room temperature according the measurement of electrocardiograph from a test subject.

6 Fuzzy Control System Based on Electrocardiogram

It is possible to know the present state of a test subject using the electrocardiograph. In this paper a fuzzy controller is employed in adjusting an environment to the most comfortable on the basis of the measurement.

In this paper one attribute is selected out of various features concerned with a comfortable space. That is, in building a comfortable space, room temperature is employed as a controlled parameter and the most comfortable space will be realized for a test subject by adjusting the room temperature. It means to build the Kansei System realizing that a person in a space will be made most comfortable by controlling the temperature of the space.

6.1 Result of Selection of Characteristic Features

1) The interval between steps 1 to 10 can be explained as the interval is the start of working and the object did not feel any stress for the working, then the electrocardiograph was not yet decreasing sufficiently.

2) The interval between steps 11 to 13 can be explained as the subject had rest and the electrocardiograph is appropriate. On the other hand, in step 14 the subject started working. In this term the subject just started the work so the electrocardiograph did not decrease from the relaxed state to the stressed state.

3) The interval between steps 24 to 28 can be explained as the subject had appropriate electrocardiograph between steps 24 to 27 as same as in the interval between steps 11 and 13 but in step 28 the subject just started rest so the state of the subject was not changed sufficiently from the relaxed state to the stressed state. Therefore, the electrocasrdiogram did not decrease as it was expected.

6.2 Rule Table

In the experiment a fuzzy controller is employed to adjust the temperature in the room. The fuzzy controller is configured as a Table 2

The rule is employed as show as in Table 2 where x_1 is denotes to present temperature and Δx_1 denotes to change of the present temperature.

Table 2. Table of Control Rules

		Δx_1				
		PB	PS	ZO	NS	NB
x_1	NB			PB		
	NS			PS		NS
	ZO	PB	PS	ZO	NS	NB
	PS	PS		NS		
	PB			NB		

In Table 2 notations NB to PB are as N, P, B and S denote Negative, Positive, Big and Small, respectively. That is, NB means big in the negative value, PS means small in the positive value and ZO means about zero. The ilustraction NB to PB .

6.3 Algorithm of Control System

STEP1) Input the degree of the present temperature.
STEP2) Change the value into 1 and 0 according to the working and rest of the subject using the characteristic abstraction of electrocardiograph.
STEP3) Using data obtained in STEP2,
- 3-1) As value 1 shows that the subject is in the rest state, set the optimum temperature to 25 C degree in order to make the subject relaxed. So the temperature of the space is controlled as it becomes 25 C degree by a fuzzy controller. Value α is set to 25.
- 3-2) As value 0 shows that the subject is in the working state, set the optimum temperature to 18 C degree in order to make the subject work comfortably. So the temperature of the space is controlled as it becomes 18 C degree by fuzzy control. Value α is set to 18.

6.4 Simulation Results

The simulation is pursued according the algorithm of a control system shown in section 6.3. Figure 6 shows the results. The temperature of the room started from 20 C degree and the temperature of the room is optimally adjusted according the state of the subject using the measurement of the electrocardiograph.

As Figure 6 shows, the temperature of the room is optimally controlled according the state of the subject. In the interval between steps 0 to 100 the

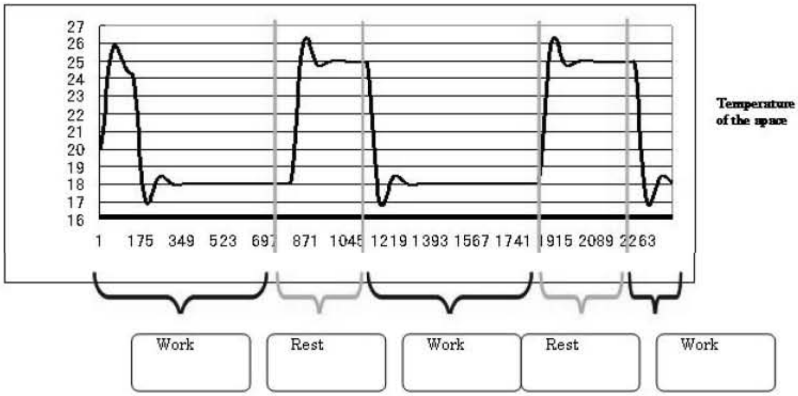


Fig. 6. Resuted room temperature

temperature of the room was adjusted to 25 C degree because the subject just started the work and the electrocardiograph did not decrease before feeling the stress.

Section 6.3 was spent to explain how the space temperature is adjusted to the same value according to the subject state. In this section the optimal temperature is obtained from database which contains the historical change of the temperature in the condition of a day such as time, weather, and the state of a person. It should be explained how to decide value under the lack of data. A neural network is employed to compensate a lack of data. The smart housing adjusts a room temperature to some optimal value which is obtained from data base. When the data base has no appropriate value, the neural network compensates the lacked value using the neighbor values.

7 Conclusions

The temperature of the space is effectively controlled using the measurement of electrocardiograph of a test subject according his/her state such as in working or in rest. The proposed method can be extended to other biopsy human parameters such as humidity, smell and so on. And also we may use other parameters of a human body including electrocardiogram, heart analysis, saliva analysis and so on. This system enables us to control a human environment according the state of a person.

Nevertheless, there are several issues that should be considered. One issue is how the temperature should be set for plural persons in a room. One solution is to set the mean value for all persons there. In this case it will happen that all people would not be satisfied.

As we mentioned before, recently, "Comfort," "Senses," and "Kansei" are used in order to express products in their advertising. The reason is because today we seek products more fitting to ourselves than in the days when manufacturing few kinds of many products. Comfort requires safety as well as good environment, smoothing handling. But it is difficult to measure. Therefore, we believed that our proposed method based on human behaviors or human states is a way to measure a comfortable space for a human.

Acknowledgements

Experimental instruments are provided by Fukuda ME Co. 1) Holster electrocardiogram QR2100 which enables us to record electrocardiograph for 24 hours and 2) HS1000Lite analyzing instrument.

The first author would like to thank to Kolej Universiti Kejuruteraan dan Teknologi Malaysia (KUKTEM) and Kementerian Pengajian Tinggi Malaysia for a study leave.

References

1. *Biometric Insight: http://www.biometricinsight. com/biometrics.html, 2006*
2. Clarke, R.: *Biometrics and Privacy* http://www.anu.edu.au /people /Roger.Clarke /DV /Biometrics.html (2001).
3. Kohritani, M., Watada, J., Hirano, H., Yubazaki, N.: Kansei Engineering for Comfortable Space Management, *Knowleddge-Based Intelligent Information and Engineering Systems (LNAI 3682)*, eds. by Khosla, K. Howlett, R.J., Jain., L.C. **II** (2005) 1291-1297.
4. Nagamachi, M.: *Kansei Engineering, Kaibundo* (1988), 2,3,112-114.
5. Shinya Nagasawa, Development of Products by Kansei, Nippon Publishing Service, pp. 4-13, (2002).
6. Takagi, M., Watada,J., Naoyosi,Y.: Realization of Comfortable Space Based on Senses Information, *Proceedints, IEEE SMC 2004*, Session 77 (2004) 6363-6364.
7. Terano, T., Asai,K., Sugeno, M.: *Applied Fuzzy Systems*, Academic Press (1994) 101-117.
8. Watada,J., Takagi,M., Yubazaki,N., Hirano H.: Realization of comfortable space based on human senses, *Journal of Systems and Control Engineering Part I*, **220(I8)** (2006) 667-675.

Sub-algebras of Finite Lattice Implication Algebra

Yang Xu, Jun Ma*, and Jiajun Lai

Department of Mathematics, Southwest Jiaotong University
Chengdu 610031, China
xuyang@home.swjtu.edu.cn, mj99032@sina.com.cn

Abstract. As a kind of logical algebra, lattice implication algebra has been applied in lattice-valued logic. Study in algebraic structure of sub-algebra of lattice implication algebra can help to construct and apply lattice implication algebra to real applications. In this paper, we studied the structural properties of sub-algebra of a finite lattice implication algebra and proposed a method for extracting a sub-algebra from it.

1 Introduction

The lattice-valued logic based on lattice implication algebra is proposed as an alternative soft computing techniques, which can be applied to describe uncertainty with incomparability [9], which is established on the lattice implication algebra (LIA). Many research have been done on sub-structures of LIA, such as filters, ideas, etc [9,13,11,4], however, few work has been reported on its sub-algebra. Study on sub-algebra of LIA is of great importance. On the one hand, we can learn more essential properties of LIA and relationship between LIA and other logical algebras [8]. On the other hand, we can get a solid foundation for applying lattice-valued logic based on it to resolve real problems.

In this paper, we will focus on the topic of sub-algebra of finite LIA and discuss its properties in structure. The rest of the paper is organized as follows. In Section 2, some results on structure of finite LIA is given. Section 3 shows the main properties of sub-algebra of finite LIA. Future work is discussed in Section 4.

2 Preliminary

Definition 1. *A* lattice implication algebra *is defined to be a bounded lattice* $(L, \vee, \wedge, 0, 1)$ *with order-reversing involution "* $'$ *" and a binary operation "*$\rightarrow$*" satisfying the following axioms:*

1). $x \rightarrow (y \rightarrow z) = y \rightarrow (x \rightarrow z)$,
2). $x \rightarrow x = 1$,
3). $x \rightarrow y = y' \rightarrow x'$,

* Corresponding author.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 813–821, 2007.
springerlink.com

4). $x \to y = y \to x = 1 \Rightarrow x = y$,
5). $(x \to y) \to y = (y \to x) \to x$,
6). $(x \vee y) \to z = (x \to z) \wedge (y \to z)$,
7). $(x \wedge y) \to z = (x \to z) \vee (y \to z)$,

for all $x, y, z \in L$.

In the following, let L be a finite LIA of the form

$$L = \left\{0 < \frac{1}{n} < \cdots < \frac{n-1}{n} < 1\right\}, \tag{1}$$

where $n \in \mathbb{Z}$ (for convenience, we shall denote L as $L(n)$).

For any finite LIA, the following lemmas (Lemma 1, 2, and 3) have been proved by Qin, Liu, and Ma, respectively.

Lemma 1. [9] *If L is a finite chain-type LIA, then it is a finite Łukasiewicz chain.* □

Lemma 2. [1] *A finite LIA is a direct product of finite Łukasiewicz chains.* □

Lemma 3. [3] *A non-chain-type LIA has at least two dual atoms, i.e., if an LIA has unique dual atom, then it is chain-type.* □

Theorem 1. [4] *For any positive number $n \geqslant 2$, if n is not prime number, then there exist more than two different LIAs which has n elements, otherwise if n is a prime number, there exists unique chain-type LIA with n elements.* □

3 Main Results

According to Lemma 2, we shall discuss the structure of finite chain-type LIAs first. For convenience, let $\mathrm{Sub}(L)$ be the set of all sub-algebras of $L(n)$.

Theorem 2. *If $L(n+1)$ has non-trivial lattice implication sub-algebra(s), then n is not a prime number, and*

$$\mathrm{Sub}(L) = \{L(p+1): \ p \mid n, p \in \mathbb{Z}, p \geqslant 1\}. \tag{2}$$

Proof. Without loss of generality, suppose $L(m+1)$ is a lattice implication sub-algebra of $L(n+1)$, and

$$L(m+1) = \left\{0 < \frac{i_1}{n} < \frac{i_2}{n} < \cdots < \frac{i_{m-1}}{n} < 1\right\}.$$

By Lemma 1, we have: for any $k \in \{1, 2, \ldots, m-1\}$,

$$\left(\frac{i_k}{n}\right)' = \frac{i_{m-k}}{n},$$

that is, for any $k \in \{1, 2, \ldots, m-1\}$,

$$i_k + i_{m-k} = n. \tag{3}$$

For any k, $j \in \{1, 2, \ldots, m-1\}$, we have

$$\frac{i_k}{n} \to \frac{i_j}{n} = 1, \quad (k \leqslant j)$$

and

$$\frac{i_k}{n} \to \frac{i_j}{n} = \min\left\{1, \frac{(n - i_k + i_j)}{n}\right\}, \quad (k > j).$$

In fact, by properties of LIA, we have

$$\frac{i_k}{n} \to \frac{i_j}{n} \in \left\{\frac{i_j}{n}, \cdots, \frac{i_{m-1}}{n}\right\}. \tag{4}$$

Now, we shall prove that for any $k \in \{1, 2, \ldots, m-2\}$,

$$\frac{i_{m-1}}{n} \to \frac{i_k}{n} = \frac{i_{k+1}}{n}. \tag{5}$$

Notice that for any $k \in \{1, 2, \ldots, m-2\}$,

$$1 - \frac{i_{m-1}}{n} + \frac{i_k}{n} > \frac{i_k}{n},$$

for $n - i_{m-1} > 0$. Therefore,

$$\frac{i_{m-1}}{n} \to \frac{i_k}{n} > \frac{i_k}{n}.$$

Moreover, because the sequence

$$\frac{i_{m-1}}{n} \to \frac{i_{m-2}}{n}, \quad \frac{i_{m-1}}{n} \to \frac{i_{m-3}}{n}, \quad \ldots, \quad \frac{i_{m-1}}{n} \to \frac{i_k}{n}$$

is an increasing sequence with $m-k-1$ different results, and

$$\frac{i_{k+1}}{n}, \frac{i_{k+2}}{n}, \ldots, \frac{i_{m-1}}{n}$$

is an increasing sequence with $m-k-1$ elements, and by Eq. (4), Eq. (5) holds. Hence we have

$$1 - \frac{i_{m-1}}{n} = \frac{i_{k+1}}{n} - \frac{i_k}{n}, \quad k = 1, 2, \ldots, m-2$$

that is to say, the sequence $\{i_1, i_2, \ldots, i_{m-1}, n\}$ is a arithmetical progression. Furthermore, $\{0, i_1, i_2, \ldots, i_{m-1}, n\}$ is a arithmetical progression for $i_1 + i_{m-1} = n$.

Let $d = |i_{k+1} - i_k|$, then we have $d \mid n$ and $d \cdot m = n$. Therefore, $m \mid n$. For the arbitrariness of m, Sub$(L(n))$ can be obtained when m takes all possible divisors of n. □

By Theorem 2, it is easy to obtain all non-trivial sub-algebras of a finite chain-type LIA provides $p \neq 1$.

Corollary 1. *For any $n \in \mathbb{Z}$, and $n \geqslant 4$, if $n-1$ is not prime numbers, then $L(n)$ has non-trivial sub-algebra(s).* □

Corollary 2. *$L(2n+1)$, $(n \geqslant 1)$, has non-trivial sub-algebras.* □

Corollary 3. *$L(2)$ only has trivial sub-algebras.* □

Example 1. Consider the $L(7)$ in Fig. 1, its non-trivial sub-algebras are shown on the right side for comparison. Because L has 7 elements, $n = 6 = 1 \times 2 \times 3$, and L has 3 non-trivial sub-algebras, *i.e.*, S_1 (where $p = 2$), S_2(where $p = 3$), and S_3(where $p = 6$).

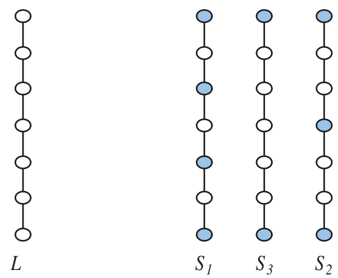


Fig. 1. Example of lattice implication sub-algebra

By Lemma 2, it is easy to show that if $L_1, L_2, \cdots, L_n$ is a set of finite chain-type LIAs, and $S_1, S_2, \cdots, S_n$ are their lattice implication sub-algebras respectively, then $S = S_1 \times S_2 \times \cdots \times S_n$ is a lattice implication sub-algebra of L, where $L = L_1 \times L_2 \times \cdots \times L_n$.

In the following, we give some examples.

Example 2. Suppose the finite chain-type LIAs L_1, L_2, and their non-trivial sub-algebras S_{11}, S_{12}, S_{13}, S_{21}, and S_{22} are shown in Fig. 2. Then two non-trivial lattice implication sub-algebras of $L = L_1 \times L_2$ are shown in Fig. 3, *i.e.*, the sub-algebras $S_{12} \times S_{21}$ and $S_{13} \times S_{21}$, where grayed circles are elements in these sub-algebras.

Here, a question arises naturally that how many non-trivial sub-algebras we can obtain from a finite LIA? It is easy to know that a sub-algebra of a finite LIA may have two forms, *i.e.*, chain-type or non-chain-type. Therefore, we shall investigate for each case.

By Lemma 2, if we suppose that $L = L_1 \times L_2 \times \cdots \times L_n$, then the sub-algebra S generated by direct product of sub-algebras $S_1, S_2, \ldots, S_n$ is not trivial if and only if not all sub-algebras $S_1, S_2, \ldots, S_n$ are trivial. Moreover, such sub-algebra S is non-chain-type. Hence, we should find out all non-trivial sub-algebras for each L_i.

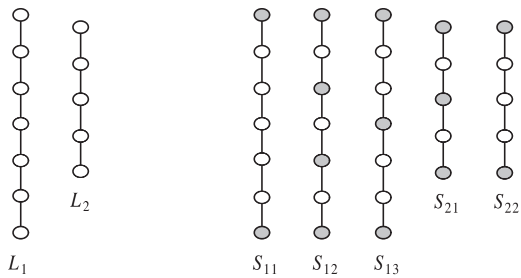


Fig. 2. Two examples of chain-type LIAs

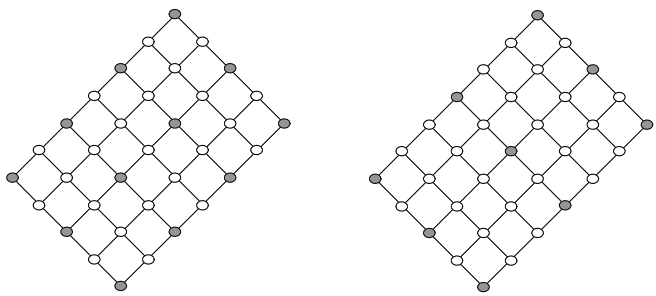


Fig. 3. Two examples of non-trivial sub-algebras generated by direct production

Lemma 4. [6] *Suppose n is a positive integer, and n has standard resolution of prime numbers as follows*

$$n = p_1^{s_1} \cdot p_2^{s_2} \cdots p_k^{s_k},$$

where $p_i (1 \leqslant i \leqslant k)$ are prime numbers and $p_1 < p_2 < \cdots < p_k$. Then the number of all positive divisors of n is $\tau(n)$, where

$$\tau(n) = \prod_{i=1}^{k} (s_i + 1).$$ □

Theorem 3. *If n has standard resolution of prime numbers*

$$n = p_1^{s_1} \cdot p_2^{s_2} \cdots p_k^{s_k},$$

then $L(n+1)$ has

$$\left(\prod_{i=1}^{k} (s_i + 1) \right) - 1$$

non-trivial lattice implication sub-algebra(s).

Proof. By Lemma 4 and Theorem 2, it is easy to know that L has

$$\prod_{i=1}^{k}(s_i+1)$$

sub-algebras. Since L is a sub-algebra of itself, which is the unique trivial sub-algebra. Hence L has

$$\left(\prod_{i=1}^{k}(s_i+1)\right)-1$$

non-trivial sub-algebras.

Example 3. Reconsider the L_1 in Fig. 2. Because $n=6$ has standard resolution of prime numbers

$$n=6=2^1\times 3^1, \tag{6}$$

we have that L_1 has $(1+1)\times(1+1)-1=3$ non-trivial sub-algebras, and they are the very S_{11}, S_{12}, and S_{13}.

Corollary 4. *Let $L=L_1\times L_2\times\cdots\times L_k$ be a direct product of k finite chain-type LIAs, and L_i has n_i+1 elements, $i=1,2,\ldots,k$. If n_i has standard resolution of prime numbers*

$$n_i=p_{i1}^{s_{i1}}\cdot p_{i2}^{s_{i2}}\cdots p_{im_i}^{s_{im_i}},\quad i=1,2,\ldots,k \tag{7}$$

then L has

$$\left(\prod_{i=1}^{k}\left(\prod_{j=1}^{m_i}(s_{ij}+1)\right)\right)-1 \tag{8}$$

non-trivial and non-chain-type sub-algebras. □

It is easy to see that no other non-chain-type sub-algebras can be obtained by other methods. In fact, because a sub-algebra S is an LIA, it can be seen as direct product of some finite chain-type LIAs again if it is not chain-type by Lemma 2. Therefore, we have

$$S\in\prod_{i=1}^{n}\mathrm{Sub}(L_i).$$

Now we shall turn to find out chain-type sub-algebras. Without loss of generality, suppose $L=L_1\times L_2$, and L_1, L_2 are

$$L_1=\left\{0<\frac{1}{m}<\frac{2}{m}<\cdots<\frac{m-1}{m}<1\right\},$$
$$L_2=\left\{0<\frac{1}{n}<\frac{2}{n}<\cdots<\frac{n-1}{n}<1\right\}.$$

For each element in L, we shall denote it by $(i/m, j/n)$ for convenience. Also, we shall suppose that S is a chain-type sub-algebra of L, and S is of form

$$S = \left\{(0,0) < (\frac{i_1}{m}, \frac{j_1}{n}) < (\frac{i_2}{m}, \frac{j_2}{n}) < \cdots < (\frac{i_{k-1}}{m}, \frac{j_{k-1}}{n}) < (1,1)\right\}.$$

By Theorem 2, we have $k \mid n$ and $k \mid m$. That is to say, k is a common divisor of m and n. Hence, we have the following conclusion.

Theorem 4. *Let $L = L_1 \times L_2$ be a direct product of two finite chain-type LIA L_1 and L_2 with $m+1$ and $n+1$ elements respectively. If g is the greatest common divisor of n and m, and g has standard resolution of prime numbers*

$$g = p_1^{s_1} \cdot p_2^{s_2} \cdots p_l^{s_l}, \tag{9}$$

where $p_i (1 \leqslant i \leqslant l))$ are prime numbers, and $p_1 < p_2 < \ldots < p_l$, then L has

$$\prod_{i=1}^{l} (s_i + 1) \tag{10}$$

non-trivial and chain-type sub-algebras. □

In Fig. 4, the finite LIA L is a direct product of two chain-type LIAs with 3 elements respectively. Then its chain-type sub-algebras are shown by grayed elements.

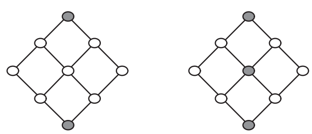


Fig. 4. Examples of chain-type sub-algebra

The conclusion of Theorem 4 can be extended easily.

Theorem 5. *Let $L = L_1 \times L_2 \times \cdots \times L_n$ be a finite LIA, where $L_i (1 \leqslant i \leqslant n)$ is a finite chain-type LIA with $m_i + 1$ elements. If the greatest common divisor of $m_1, m_2, \cdots, m_n$ has standard resolution of prime numbers*

$$g = p_1^{s_1} \cdot p_2^{s_2} \cdots p_k^{s_k},$$

where $p_j (1 \leqslant j \leqslant k)$ is prime number, and $p_1 < p_2 < \cdots < p_k$, then L has

$$\prod_{i=1}^{k} (s_i + 1)$$

non-trivial and chain-type sub-algebras. □

Corollary 5. *Let $L = L_1 \times L_2 \times \cdots \times L_n$ be a finite LIA, where $L_i (1 \leqslant i \leqslant n)$ is a finite chain-type LIA with $m_i + 1$ elements. If the greatest common divisor of m_1, m_2, $\cdots$, m_n is 1, then L has unique chain-type sub-algebra, i.e., $S = (\{O = (0, \ldots, 0), I = (1, \ldots, 1)\}, \vee, \wedge, \prime, \rightarrow)$.* □

Theorem 6. *Let L be a finite LIA, S_1 and $S_2 \in \mathrm{Sub}(L)$. Then $S_1 \cap S_2$, $S_1' \in \mathrm{Sub}(L)$, where*

$$S_1' = \{\alpha' \in L : \quad \alpha \in S_1\},$$
$$S_1 \cap S_2 = \{\alpha \in L : \quad \alpha \in S_1, \alpha \in S_2\}. \square$$

4 Conclusions

This paper discusses the structure of sub-algebra of finite lattice implication algebra. A method for counting and extracting sub-algebras from a finite LIA is proposed. Examples show that we can learn more properties of an LIA by its sub-algebras. As a kind of logical algebra, LIA has close relationship with other logical algebras, how to investigate the relationship among them through their sub-algebras and how to use them in uncertainty reasoning based on lattice-valued logic will be our future works.

Acknowledgement

This work is partially supported by the China NSF with grant number 60474022 and the Special Research Funding to Doctoral Subject of Higher Education Institutes in P.R. China with Grant number 20060613007.

References

1. J. Liu, *Study on Lattice–valued Logic based on lattice implication algebra and its Resolution Principle*, Doctoral Dissertation, Southwest Jiaotong University, Chengdu, 1999
2. J. Liu, D. Ruan, Y. Xu, and Z. Song, A Resolution-Like Strategy Based on a Lattice-Valued Logic, *IEEE Transactions on Fuzzy Systems*, 2003, **11**(4): 560–567.
3. J. Ma, *Lattice–Valued Loigcal System and Automated Reasoning Based on Lattice Implication Algebra*, Doctoral Dissertation, Southwest Jiaotong University, Chengdu, China, 2002.
4. J. Ma and Y. Xu, On structure and complementary elements of lattice implication algebra, *Fuzzy Mathematics and Systems*, to appear.
5. V. Novak, Formal theories in fuzzy logic, 1997, http: //www.researchindex.com
6. C.D. Pan and C.B. Pan, *Basic Numerical Theory*, Beijing, China: Beijing University Press, 1994.
7. G. Wang, *Non–Classical Loigc and Approximate Reasoning*, Beijing, China: Chinese Academic Press, 2000
8. G. Wang, MV-algebras, BL-algebras, R_0-algebras and multiple-valued logic, Fuzzy Systems and Mathematics, 2002, **3**: 1–15

9. Y. Xu, D. Ruan, K. Qin, and J. Liu, *Lattice–Valued Logic – An Alternative Approach to Treat Fuzziness and Incomparability*, Germany: Springer, 2003
10. J. Zhang, *SAT of Logical Formule – Method, Tools and Application*, Beijing, China: Chinese Academic Press, 2000.
11. E. H. Roh, K. H. Kim and Y. B. Jun, Fuzzifications of $\oplus$-closed sets in lattice implication algebras, *East-West J. of Mathematics*, **4** (2002), No. 1, 63-75.
12. Y. B. Jun, E. H. Roh and Y. Xu, *LI*-ideals in lattice implication algebras, *Bull. Korean Math. Soc.* **35** (1998), no. 1, 13–24.
13. Y. B. Jun, Y. Xu, J. Ma, Redefined fuzzy implicative filters, *Information Sciences*, 2006, (in Press)

Semantics Properties of Compound Evaluating Syntagms

Zheng Pei[1], Baoqing Jiang[2], Liangzhong Yi[1], and Yang Xu[3]

[1] School of Mathematics & Computer Engineering, Xihua University, Chengdu, Sichuan, 610039, China
[2] Institute of Data and Knowledge Engineering, Henan University, Kaifeng, Henan 475001, China
[3] Department of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 610031, China
{ZhengPei,LiangzhongYi}pqyz@263.net,
jbq@henu.edu.cn,
xuyang@home.swjtu.edu.cn

Abstract. After generating simple evaluating syntagms which are constructed by linguistic hedges and atomic evaluating syntagms, a voting mechanism is adopted for truth values of evaluating syntagms propositions on universe D. Then, a formal context of voting mechanism about evaluating syntagms is proposed. Based on the formal context, formal concepts of evaluating syntagms are obtained, and these formal concepts are used to semantics of simple evaluating syntagms and compound evaluating syntagms which are generated by simple evaluating syntagms and the connectives $\wedge, \vee, \rightarrow, \neg$.

Keywords: Evaluating Syntagms, Formal Concept Analysis, Voting Mechanism.

1 Introduction

As it is well known, humans employ words in computing and reasoning, arriving at conclusions expressed as words from premises expressed in a natural language or having the form of mental perceptions. The methodology of computing with words (CWW) proposed by Zadeh may be viewed as an attempt to harness the highly expressive power of natural languages by developing ways of CWW or propositions drawn from a natural language [1]-[3], [7]-[15]. In processing linguistic information, the follows are important:

- Evaluating linguistic expressions: A significant part of natural language is formed by a special class of expressions using which we characterize dimensions, sizes, volumes, etc. In general, they characterize a position on an ordered scale (usually on an interval of real numbers), they are called by evaluating linguistic expressions [5], [6]. From fuzzy logic in broader sense point of view, Novák, et al proposes the evaluating linguistic predications, and developes a theory of natural human reasoning [4]-[6].

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 822–831, 2007.
springerlink.com

- A voting mechanism for fuzzy logic: To solve the meaning of fuzzy truth values or membership functions of fuzzy sets, a voting mechanism for fuzzy logic is proposed, and used to the fuzzy predicate calculus [16]-[18].
- Formal concept analysis(FCA): It is a discipline that studies the hierarchical structures induced by a binary relation between a pair of sets. The term concept lattice and formal concept analysis are due to Wille [19]-[21]. As a formalization tool, FCA is used in several areas such as data mining, knowledge discovery, and software engineering, etc [22]-[27].

Based on evaluating linguistic expressions, a voting mechanism for fuzzy logic and FCA, semantics properties of simple evaluating syntagms and compound evaluating syntagms are discussed in the paper.

2 Preliminaries

For evaluating linguistic expressions, atomic evaluating syntagms and simple evaluating syntagms are distinguished, atomic evaluating syntagms usually form pairs of $\langle\langle$nominal adjective$\rangle - \langle$antonym$\rangle$, *e.g*, “young - old”, “weight - light”, “big - small”, “high - low”, etc. Simple evaluating syntagms form

$$\langle \text{linguistic hedge} \rangle \langle \text{atomic evaluating syntagm} \rangle.$$

In [5], linguistic hedges are decomposed into two classes, one is with *narrowing effect* (such as, *very*, *highly*, *more*, etc), the other with *widening effect* (such as, *more or less*, *roughly*, *little*, etc). They are formalized as

$$H = \{h_{-k}, \cdots, h_{-1}, h_0, h_1, \cdots, h_k\}. \tag{1}$$

Where, $\{h_1, \cdots, h_k\}$ are with *narrowing effect*, $\{h_{-k}, \cdots, h_{-1}\}$ with *widening effect*, $\{h_0\}$ is called *central*, and $\forall h_i \in H$, h_i is interpreted by an unary operator,

$$h_i : [0, 1] \longrightarrow [0, 1].$$

The follows rule is used to linguistic hedges

$$R_{lh} : \frac{\alpha/A}{h_i(\alpha)/h_i(A)},$$

where, $\alpha \in [0, 1]$ is truth value of A. A many-sorted language is used to identify predicates [4]-[6].

In [18], a voting mechanism for fuzzy logic is proposed to extend notion of binary valuation for fuzzy concepts and fuzzy predicate. Formally, let L be a countable language of the propositional calculus consisting of a countable set of propositional variables PVL together with the connectives $\wedge, \vee$ and $\neg$. Let SL denote the sentences of L.

Definition 1. *[18] A fuzzy valuation of* L *is a function* $F : SL \times [0, 1] \longrightarrow \{t, f\}$ *such that* $\forall \theta \in SL$, $\forall 0 \le y < y' \le 1$, $F(\theta, y) = f \Longrightarrow F(\theta, y') = f$ *and satisfies the following:* $\forall \theta, \phi \in SL, y \in [0, 1]$

1. $F(\theta \wedge \phi, y) = t \Longleftrightarrow F(\theta, y) = t$ *and* $F(\phi, y) = t$,
2. $F(\theta \vee \phi, y) = t \Longleftrightarrow F(\theta, y) = t$ *or* $F(\phi, y) = t$,
3. $F(\neg\theta, y) = t \Longleftrightarrow F(\theta, 1 - y) = f$.

In the Definition, $y \in [0, 1]$ is viewed as the scepticism level of the voting agent. Based on the voting model, many interesting and important subjects of fuzzy logic have been discussed.

Formal concept analysis(FCA) is widely used in data mining and knowledge discovering, etc [19]-[27].

Definition 2. *A formal context is an ordered triple* $T = (G, M, I)$, *where* G, M *are nonempty sets and* $I : G \times M \longrightarrow \{0, 1\}$ *is a binary relation. The elements of* G *are called objects and the elements of* M *attributes.* $I(g, m) = 1$ *means that object* g *has attribute* m.

I of a formal context can be naturally represented by an two-valued table. The follows two set-valued functions are used to define the formal concept of $T = (G, M, I)$

$$\uparrow : 2^G \to 2^M,\ X^{\uparrow} = \{m \in M; \forall g \in X, I(g, m) = 1\}, \tag{2}$$
$$\downarrow : 2^M \to 2^G, Y^{\downarrow} = \{g \in G; \forall m \in Y, I(g, m) = 1\}. \tag{3}$$

Definition 3. *A formal concept of a context* $T = (G, M, I)$ *is a pair* $(A, B) \in 2^G \times 2^M$ *such that* $A^{\uparrow} = B$ *and* $B^{\downarrow} = A$. *The set* A *is called its extent, the set* B *its intent.*

3 A Formal Context of Voting Mechanism About Evaluating Syntagms

Let voting agents be a finite set $V = \{v_i | i = 1, \cdots, n\}$, linguistic hedges a finite set $H = \{h_{-k}, \cdots, h_{-1}, h_0, h_1, \cdots, h_k\}$, atomic evaluating syntagms $\{l_1, l_2\}$, and its universe a finite set $D = \{d_j | j = 1, \cdots, m\}$. Correspondingly, simple evaluating syntagms are $LA = H \times \{l_1, l_2\} = \{h_{-k}l_1, \cdots, h_k l_2\}$. In this paper, Gaines voting model [16], which can be formalized as follows, is considered

$$F : V \times \mathcal{A} \longrightarrow \{true, false\},\ (v_i, A_k) \longrightarrow true(or false). \tag{4}$$

In which, $\mathcal{A}$ is the set of sentences from some language of the propositional calculus. $F(v_i, A_k) = true(or false)$ means that voting agent v_i assigns "true" (or "false") to proposition $A_k \in \mathcal{A}$. Specially, propositions considered in the paper have the follows form

$$A_{(d_j, h_{k'} l_r)} : d_j \text{ is } h_{k'} l_r,$$

in which, $d_j \in D$, $k' \in \{-k, \cdots, 0, \cdots, k\}$ and $r \in \{1, 2\}$. Due to D and LA are finite, $\mathcal{A} = D \times LA$ is finite. Formally, (4) is equal to the follows formal context

$$F : V \times \mathcal{A} \longrightarrow \{true, false\} \Longleftrightarrow (V \times D, LA, I). \tag{5}$$

Example 1. [9] Suppose the variable $SCORE$ with universe $D = \{1, 2, 3, 4, 5, 6\}$ gives the outcome of a single throw of a particular dice. Let $LA = \{low, medium, high\}$ (here, linguistic hedges H are not considered, and $\{low, medium, high\}$ is called basic linguistic trichotomy in [5]) and voting agents $V = \{v_1, v_2, v_3\}$, then propositions are

$$\begin{array}{lll} A_{(1,low)} : 1 \text{ is low}; & A_{(1,medium)} : 1 \text{ is medium}; & A_{(1,high)} : 1 \text{ is high}; \\ A_{(2,low)} : 2 \text{ is low}; & A_{(2,medium)} : 2 \text{ is medium}; & A_{(2,high)} : 2 \text{ is high}; \\ A_{(3,low)} : 3 \text{ is low}; & A_{(3,medium)} : 3 \text{ is medium}; & A_{(3,high)} : 3 \text{ is high}; \\ A_{(4,low)} : 4 \text{ is low}; & A_{(4,medium)} : 4 \text{ is medium}; & A_{(4,high)} : 4 \text{ is high}; \\ A_{(5,low)} : 5 \text{ is low}; & A_{(5,medium)} : 5 \text{ is medium}; & A_{(5,high)} : 5 \text{ is high}; \\ A_{(6,low)} : 6 \text{ is low}; & A_{(6,medium)} : 6 \text{ is medium}; & A_{(6,high)} : 6 \text{ is high}. \end{array} \tag{6}$$

Every voting agent of V assigns "true" or "false" to propositions of (6) (see Table.1).

Table 1. the formal context of voting model

	$(v_1, 1)$	$(v_2, 1)$	$(v_3, 1)$	$(v_1, 2)$	$(v_2, 2)$	$(v_3, 2)$	$(v_1, 3)$	$(v_2, 3)$	$(v_3, 3)$
low	t	t	t	t	t	t	f	f	t
medium	f	f	f	t	f	f	t	t	t
high	f	f	f	f	f	f	f	f	f
	$(v_1, 4)$	$(v_2, 4)$	$(v_3, 4)$	$(v_1, 5)$	$(v_2, 5)$	$(v_3, 5)$	$(v_1, 6)$	$(v_2, 6)$	$(v_3, 6)$
low	f	f	f	f	f	f	f	f	f
medium	t	t	t	f	t	f	f	f	f
high	t	f	f	t	t	t	t	t	t

In Table.1, attributes are simple evaluating syntagms "low, medium, high". Objects are $(v_1, 1), etc$, t is true and f is false. *e.g.*, "$(v_1, 1)$ and low is t" means that voting agent v_1 assigns "true" to proposition "1 is low".

The advantage of $(V \times D, LA, I)$ to express evaluating syntagms voting model is that FCA can be used to discuss fuzzy linguistic, and provide a new tool for computing with words. These will be see in the rest of paper.

4 Semantics of Evaluating Syntagms

Based on (2) and (3), in $(V \times D, LA, I)$,

$$A^{\uparrow} = \{hl \in LA | \forall (v_i, d_j) \in A, I((v_i, d_j), hl) = t\}, \tag{7}$$

$$B^{\downarrow} = \{(v_i, d_j) \in V \times D | \forall hl \in B, I((v_i, d_j), hl) = t\}. \tag{8}$$

According to Definition 3, based on (7) and (8), all formal concepts can be generated. Here, every formal concept has special meaning, *i.e.*, its intent is evaluating syntagms, and its extent includes voting agent and element of universe

D of evaluating syntagms. Hence, every formal concept can be understood as semantics model of evaluating syntagms in voting agents V. All formal concepts of $(V \times D, LA, I)$ is denoted as

$$CF_{LA} = \{(A, B) | A \subseteq V \times D, B \subseteq LA, A^{\uparrow} = B, B^{\downarrow} = A\}. \tag{9}$$

Example 2. Continues Example 1. All formal concepts of Table 1 are

$(V \times D, \emptyset)$, $(\emptyset, LA)$,
$(\{low\}^{\downarrow}, \{low\}) = (\{(v_1, 1), (v_2, 1), (v_3, 1), (v_1, 2), (v_2, 2), (v_3, 2)\}, \{low\})$,
$(\{(v_2, 2), (v_3, 3)\}, \{low, medium\})$, $(\{(v_1, 4), (v_2, 5)\}, \{medium, high\})$,
$(\{medium\}^{\downarrow}, \{medium\}) = (\{(v_1, 2), (v_1, 3), (v_2, 3), (v_3, 3), (v_1, 4), (v_2, 4),$
$(v_3, 4), (v_2, 5)\}, \{medium\})$, $(\{high\}^{\downarrow}, \{high\}) = (\{(v_1, 4), (v_1, 5), (v_2, 5),$
$(v_3, 5), (v_1, 6), (v_2, 6), (v_3, 6)\}, \{high\})$.

In which, according to inclusion of extents of formal concept, $(V \times D, \emptyset)$ and $(\emptyset, LA)$ are the top element and the bottom element, respectively.

In order to explain semantics of evaluating syntagms, the follows projection operators of formal concept (A, B) are needed.

Definition 4. *In* $(V \times D, LA, I)$. $\forall d_j \in D$,

$$P_{d_j}(A, B) = \{v_i | (A, B) \in CF_{LA}, (v_i, d_j) \in A\}. \tag{10}$$

is called projection of d_j *on voting agents* V *based on formal concept* (A, B).

Example 3. Continues Example 2. For every $d_j \in D$, nonempty $P_{d_j}(A, B)$ are

$P_1(\{low\}^{\downarrow}, \{low\}) = P_2(\{low\}^{\downarrow}, \{low\}) = \{v_1, v_2, v_3\}, \cdots,$
$P_2(\{(v_1, 2), (v_3, 3)\}, \{low, medium\}) = \{v_1\}, \cdots,$
$P_3(\{medium\}^{\downarrow}, \{medium\}) = P_4(\{medium\}^{\downarrow}, \{medium\}) = \{v_1, v_2, v_3\}, \cdots,$
$P_4(\{(v_1, 4), (v_2, 5)\}, \{medium, high\}) = \{v_1\}, \cdots,$
$P_5(\{high\}^{\downarrow}, \{high\}) = P_6(\{high\}^{\downarrow}, \{high\}) = \{v_1, v_2, v_3\}, \cdots,$

When $P_{d_j}(A, B)$ is considered on all formal concept (A, B), then the degree of applicability of evaluating syntagms to d_j, which can be understood as as membership degree of d_j about the evaluating syntagms, can be obtained. "The degree of applicability" is called by "appropriateness degree" in [9] due to it more accurately reflects the underlying semantics. If prior distribution on V is considered, then based on (10), for every $hl \in LA$ and $d_j \in D$, appropriateness degree of hl can be obtained as follows

$$\mu_{hl}(d_j) = Pr\left(\bigcup_{(A,B) \in CF_{LA}}^{hl \in B} P_{d_j}(A, B)\right), \tag{11}$$

in which, probability $Pr(*)$ is calculated on the basis of some underlying prior distribution on V. Under a uniform distribution on V, (11) is

$$\mu_{hl}(d_j) = \frac{|\bigcup_{(A,B)\in CF_{LA}}^{hl\in B} P_{d_j}(A,B)|}{|V|}. \tag{12}$$

Example 4. Continues Example 3. Suppose a uniform distribution on V, then

$$\mu_{low}(1) = \mu_{low}(2) = \frac{|V|}{|V|} = 1, \quad \mu_{low}(3) = \frac{1}{3}; \ \mu_{medium}(2) = \frac{1}{3}, \mu_{medium}(3) =$$

$$\mu_{medium}(4) = 1 \quad , \mu_{medium}(5) = \frac{1}{3}; \quad \mu_{high}(4) = \frac{1}{3}, \mu_{high}(5) = \mu_{high}(6) = 1.$$

Similarly, if considering prior distribution on $V \times D$, a probability distribution (or mass assignment) for the random set on LA cab be defined by $\forall S \subseteq LA$ and $d_j \in D$,

$$m_{d_j}(S) = Pr(\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = S\}), \tag{13}$$

$$m(S) = \sum_{d_j\in D} m_{d_j}(S), \tag{14}$$

in which, probability $Pr(*)$ is calculated by prior distribution on $V \times D$. In $(V \times D, LA, I)$, $\forall(v_j, d_j) \in V \times D$, $S = (v_j, d_j)^{\uparrow}$ is unique, hence, the follows is easily proved.

$$\sum_{B\in\{B|(A,B)\in CF_{LA}\}} m(B) = 1 \tag{15}$$

Example 5. Continues Example 2. Suppose a uniform distribution on $V \times D$, then

$$m(\{low\}) = \sum_{d_j\in D} \frac{|\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = \{low\}\}|}{|V \times D|} = \frac{5}{18};$$

$$m(\{low, medium\}) = \sum_{d_j\in D} \frac{|\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = \{low, medium\}\}|}{|V \times D|} = \frac{1}{9};$$

$$m(\{medium\}) = \sum_{d_j\in D} \frac{|\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = \{medium\}\}|}{|V \times D|} = \frac{2}{9};$$

$$m(\{medium, high\}) = \sum_{d_j\in D} \frac{|\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = \{medium, high\}\}|}{|V \times D|} = \frac{1}{9};$$

$$m(\{high\}) = \sum_{d_j\in D} \frac{|\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = \{high\}\}|}{|V \times D|} = \frac{5}{18};$$

Remark 1. The conclusions of Example 4 and 5 is same in [9]. The difference is that calculating appropriateness degree μ_* and mass assignment $m(*)$ are based on FCA in this paper. On the other hand, prior distribution on $V \times D$ is considered in this paper, for practice problems, the prior distribution on $V \times D$ may be combinatorial probability of the prior distribution on V and D.

5 Semantics of Compound Evaluating Syntagms

Based on $LA = H \times \{l_1, l_2\} = \{h_{-k}l_1, \cdots, h_k l_2\}$ and the connectives $\wedge, \vee, \rightarrow, \neg$, compound evaluating syntagms can be generated recursively by application of the connectives:

Definition 5. *Let $LA = H \times \{l_1, l_2\} = \{h_{-k}l_1, \cdots, h_k l_2\}$. Compound evaluating syntagms, CLA, is defined recursively as follows*

1. *$\forall hl \in LA$, $hl \in CLA$;*
2. *If $\theta, \vartheta \in CLA$, then $\theta \wedge \vartheta, \theta \vee \vartheta, \theta \rightarrow \vartheta, \neg\theta \in CLA$.*

Definition 6. *$\forall \theta \in CLA$, $\lambda(\theta)$ is defined recursively as follows: (1) $\forall hl \in LA$, $\lambda(hl) = \{B|(A, B) \in CF_{LA}, hl \in B\}$; (2) $\lambda(\theta \wedge \vartheta) = \lambda(\theta) \cap \lambda(\vartheta)$; (3) $\lambda(\theta \vee \vartheta) = \lambda(\theta) \cup \lambda(\vartheta)$; (4) $\lambda(\theta \rightarrow \vartheta) = \overline{\lambda(\theta)} \cup \lambda(\vartheta)$; (5) $\lambda(\neg\theta) = \overline{\lambda(\theta)}$. In which, $\overline{\lambda(\theta)} = \{B|(A, B) \in CF_{LA}\} - \lambda(\theta)$.*

Example 6. Continues Example 2.

$$\lambda(\{low\}) = \{\{low\}, \{low, medium\}\}, \quad \lambda(\{medium\}) = \{\{low, medium\}, \{medium\}\}, \quad \lambda(\{high\}) = \{\{medium, high\}, \{high\}\}, \quad \lambda(\{low \rightarrow high\}) = \overline{\lambda(\{low\})} \cup \lambda(\{high\}) = \{\{medium\}, \{medium, high\}, \{high\}\}.$$

Suppose that v_i and d_j are independent random variables, then

$$\begin{aligned} m_{d_j}(S) &= Pr(\{(v_i, d_j)|(v_i, d_j)^{\uparrow} = S\}) \\ &= Pr(\{v_i|(v_i, d_j)^{\uparrow} = S\}) \times Pr(\{d_j|(v_i, d_j)^{\uparrow} = S\}). \end{aligned} \quad (16)$$

In this case, appropriateness degree of compound evaluating syntagms can be obtained as follows

$$\mu_{\theta}(d_j) = \sum_{B \in \lambda(\theta)} \frac{m_{d_j}(B)}{Pr(\{d_j\})} = \sum_{B \in \lambda(\theta)} Pr(\{v_i|(v_i, d_j)^{\uparrow} = B\}), \quad (17)$$

Proposition 1. *$\forall \theta \in CLA$ and $\forall d_j \in D$, $\mu_{\neg\theta}(d_j) = 1 - \mu_{\theta}(d_j)$.*

Proof. Due to $\sum_{(A,B) \in CF_{LA}} Pr(\{v_i|(v_i, d_j)^{\uparrow} = B\}) = Pr(V) = 1$ holds in FCA for every fixed $d_j \in D$,

$$\begin{aligned} \mu_{\neg\theta}(d_j) &= \sum_{B \in \overline{\lambda(\theta)}} Pr(\{v_i|(v_i, d_j)^{\uparrow} = B\}) \\ &= \sum_{(A,B) \in CF_{LA}} Pr(\{v_i|(v_i, d_j)^{\uparrow} = B\}) - \sum_{B \in \lambda(\theta)} Pr(\{v_i|(v_i, d_j)^{\uparrow} = B\}) \\ &= 1 - \mu_{\theta}(d_j). \end{aligned}$$

Proposition 2. *If $Pr(S)(S \subseteq V)$ is a consonant mass assignment, then $\forall hl, h'l' \in LA$ and $\forall d_j \in D$,*

$$\mu_{hl \wedge h'l'}(d_j) = min\{\mu_{hl}(d_j), \mu_{h'l'}(d_j)\}, \ \mu_{hl \vee h'l'}(d_j) = max\{\mu_{hl}(d_j), \mu_{h'l'}(d_j)\}.$$

Proof. According to (11), $\mu_{hl}(d_j) = Pr(\bigcup_{(A,B)\in CF_{LA}}^{hl\in B} P_{d_j}(A,B))$ and $\mu_{h'l'}(d_j) = Pr(\bigcup_{(A,B)\in CF_{LA}}^{h'l'\in B} P_{d_j}(A,B))$. For fixed $d_j \in D$, w.l.o.g, suppose $\mu_{h'l'}(d_j) \leq \mu_{hl}(d_j)$, then

$$Pr(\bigcup_{(A,B)\in CF_{LA}}^{h'l'\in B} P_{d_j}(A,B)) \leq Pr(\bigcup_{(A,B)\in CF_{LA}}^{hl\in B} P_{d_j}(A,B)).$$

By $Pr(S)(S \subseteq V)$ is a consonant mass assignment, the follows holds

$$\bigcup_{(A,B)\in CF_{LA}}^{h'l'\in B} P_{d_j}(A,B) \supseteq \bigcup_{(A,B)\in CF_{LA}}^{hl\in B} P_{d_j}(A,B),$$

this means that $h'l' \in B \Longleftrightarrow \{hl, h'l'\} \in B$. Hence,

$$\begin{aligned}
\mu_{hl\wedge h'l'}(d_j) &= \sum_{B\in\lambda(hl\wedge h'l')} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \sum_{B\in\lambda(hl)\cap\lambda(h'l')} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \sum_{(A,B)\in CF_{LA}}^{\{hl,h'l'\}\in B} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \sum_{(A,B)\in CF_{LA}}^{\{h'l'\}\in B} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \mu_{h'l'}(d_j) = min\{\mu_{hl}(d_j), \mu_{h'l'}(d_j)\}.
\end{aligned}$$

Similarly, $\mu_{hl\vee h'l'}(d_j) = max\{\mu_{hl}(d_j), \mu_{h'l'}(d_j)\}$ can be proved.

For $\lambda(hl \wedge (\neg h'l')) = \lambda(hl) \cap \lambda(\neg h'l') = \lambda(hl) \cap \overline{\lambda(h'l')}$, if $\mu_{hl}(d_j) \leq \mu_{h'l'}(d_j)$, then $hl \in B \Longleftrightarrow \{hl, h'l'\} \in B$, this means $\mu_{hl\wedge\neg h'l'}(d_j) = 0$ when $\mu_{hl}(d_j) \leq \mu_{h'l'}(d_j)$. On the other hand, if $\mu_{h'l'}(d_j) < \mu_{hl}(d_j)$, then

$$\begin{aligned}
\mu_{hl\wedge(\neg h'l')}(d_j) &= \sum_{B\in\lambda(hl)\cap\overline{\lambda(h'l')}} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \sum_{B\in\lambda(hl)-\lambda(h'l')} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \sum_{B\in\lambda(hl)} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) - \sum_{B\in\lambda(h'l')} Pr(\{v_i|(v_i,d_j)^{\uparrow} = B\}) \\
&= \mu_{hl}(d_j) - \mu_{h'l'}(d_j).
\end{aligned}$$

Hence, $\forall d_j \in D$, $\mu_{hl\wedge(\neg h'l')}(d_j) = max\{0, \mu_{hl}(d_j) - \mu_{h'l'}(d_j)\}$.

Proposition 3. *If* $Pr(S)(S \subseteq V)$ *is a consonant mass assignment, then* $\forall hl, h'l' \in LA$ *and* $\forall d_j \in D$,

$$\mu_{hl \to h'l'}(d_j) = min\{1, 1 - \mu_{hl}(d_j) + \mu_{h'l'}(d_j)\}.$$

Proof. According to above,

$$\begin{aligned}
\mu_{hl \to h'l'}(d_j) &= \sum_{B \in \overline{\lambda(hl)} \cup \lambda(h'l')} Pr(\{v_i | (v_i, d_j)^{\uparrow} = B\}) \\
&= \sum_{B \in \overline{\lambda(hl) \cap \overline{\lambda(h'l')}}} Pr(\{v_i | (v_i, d_j)^{\uparrow} = B\}) \\
&= 1 - \sum_{B \in \lambda(hl) \cap \overline{\lambda(h'l')}} Pr(\{v_i | (v_i, d_j)^{\uparrow} = B\}) \\
&= 1 - max\{0, \mu_{hl}(d_j) - \mu_{h'l'}(d_j)\} \\
&= min\{1, 1 - \mu_{hl}(d_j) + \mu_{h'l'}(d_j)\}.
\end{aligned}$$

6 Conclusion

In this paper, by using a voting mechanism and FCA, semantics properties of evaluating syntagms are discussed. The conclusion of this paper are same as in [9]. The difference is our analysis based on FCA. The advantages of using FCA will be that the properties of FCA, *e.g.*, hierarchy of formal concepts, *etc*, can be using in evaluating syntagms or CWW.

Acknowledgments

This work is supported by the excellent young foundation of sichuan province (Grant no. 06ZQ026-037), the National Natural Science Foundation of China (Grant no.60474022), the Special Research Funding to Doctoral Subject of Higher Education Institutes in China (Grant no. 20060613007), Henan Innovation Project for University Prominent Research Talents (Grant No. 2007KYCX018) and a Project Supported by Scientific Reserch Fund of SiChuan Provincial Education Department(grant no. 2005A121, 2006A084).

References

1. Zadeh, L. A.: Fuzzy logic = computing with words. Fuzzy Systems. 4 (1996) 103–111
2. Wang, P. P.: Computing with words. John Wiley and Sons. (2001)
3. Zadeh, L. A.: Toward a theory of fuzzy information granulation and its centrality in houman reasoning and fuzzy logic. Fuzzy Sets and Systems. 90 (1997) 103–111
4. Novák, V., Perfilieva, I., Močkoř, J.: Mathematical principles of fuzzy logic. Kluwer Academic Publishers. (1999)

5. Novák, V.: Antonyms and linguistic quantifiers in fuzzy logic. Fuzzy Sets and Systems. 124 (2001) 335–351
6. Dvořák, A., Novák, V.: Fromal theories and linguistic descriptions. Fuzzy Sets and Systems. 143 (2004) 169–188
7. Tüurkşen, I. B.: Type 2 representation and reasoning for CWW. Fuzzy Sets and Systems. 127 (2002) 17–36
8. Lawry, J.: A methodology for computing with words. International Journal of Approximate Reasoning. 28 (2001) 51–89
9. Lawry, J.: A framework for linguistic modelling. Artificial Intelligence. 155 (2004) 1–39
10. Tang, Y., Zhang, J.: Linguistic modelling based on semantic similarity relation among linguistic labels. Fuzzy Sets and Systems. 157 (2006) 1662–1673
11. Bordogna, G., Chiesa, S., Geneletti, D.: Linguistic modelling of imperfect spatial information as a basis for simplifying spatial analysis. Information Sciences. 176 (2006) 366–389
12. Cooman, G.: A behavioural model for vague probability assessments. Fuzzy Sets and Systems. 154 (2005) 305–358
13. Ho, N. C., Khang, T. D., Huynh, V. N.: An algebraic approach to linguistic hedges in Zadehs fuzzy logic. Fuzzy Sets and Systems. 129 (2002) 229–254
14. Pei, Z., Xu, Y.: Lattice implication algebra model of linguistic variable Truth and its inference. Applied Computational Intelligence, World Scientific. (2004) 93–98
15. Herrera, F., Lopez E., Rodriguez, M. A.: A linguistic decision model for promotion mix management solved with genetic algorithms. Fuzzy Sets and Systems. 131 (2002) 47–61
16. Gaines, B. R.: Fuzzy and probability uncertainty logics. Journal of Information and Control. 38 (1978) 154–169
17. Gaines, B. R.: Foundations of fuzzy reasoning international. Journal of Man-Machine Studies. 8 (1976) 623–668
18. Lawry, J.: A voting mechanism for fuzzy logic. International Journal of Approximate Reasoning. 19 (1998) 315–333
19. Wille, R.: Restructuring the lattice theory: an approach based on hierarchies of concepts. Rival, I.(Ed.): Ordered Sets, Reidel, Dordrecht, Boston. (1982) 445–470
20. Wille, R.: Lattices in data analysis: how to draw them with a computer. Algorithms and order, Kluwer Acad. Publ., Dordrecht. (1989) 33–58
21. Wille, R.: Concepe lattices and conceptual knowledge systems. Comput. Math. Apll. 23(6-9) (1992) 493–515
22. Jin, J., Qin, K., Pei, Z.: Reduction-Based Approaches Towards Constructing Galois (Concept) Lattices. In G. Wang et al. (Eds.): RSKT 2006, LNAI 4062. (2006) 107–113
23. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with TITANIC. Data & Knowledge Engineering. 42 (2002) 189–222
24. Berry, A., Sigayret, A.: Representing a concept lattice by a graph. Discrete Applied Mathematics. 144 (2004) 27–42
25. Berry, A., SanJuan, E., Sigayret, A.: Generalized domination in closure systems. Discrete Applied Mathematics. 154 (2006) 1064–1084
26. Diday, E., Emilion, R. : Maximal and stochastic Galois lattices. Discrete Applied Mathematics. 127 (2003) 271–284
27. Kim, M., Compton, P.: Evolutionary document management and retrieval for specialized domains on the web. Int. J. Human-Computer Studies. 60 (2004) 201–241

Characteristic Morphisms and Models of Fuzzy Logic in a Category of Sets with Similarities*

Jiří Močkoř

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic
Jiri.Mockor@osu.cz

Abstract. Let Ω be a complete residuated lattice. By $\mathbf{SetF}(\Omega)$ we denote a category of sets with similarity relations (A, δ) with values in Ω. We investigate an interpretation of a first order predicate fuzzy logic in a model based on objects of this category $\mathbf{SetF}(\Omega)$. A notion of a fuzzy set in this category is introduced and interpretation of formulas as fuzzy sets are defined. Characteristic morphisms are defined for such fuzzy sets and relationship between interpretation by fuzzy sets and interpretation defined by characteristic morphisms are investigated.

1 Introduction

Let us recall what does an interpretation (or model) of many sorted first order predicate fuzzy logic mean ([9]). Let J be a language of fuzzy logic which consists (as classically) of a set $\mathcal{S}$ of sorts, a set of predicate symbols $P \in \mathcal{P}$, each of a sort $\iota_1 \times \cdots \times \iota_n$ for some $\iota_k \in \mathcal{S}$ and a set of functional symbols $f \in \mathcal{F}$, each of a sort $\iota_1 \times \cdots \times \iota_n \to \iota$. Moreover J contains also a set Ω of logical constants. Then a model of a language J is

$$\mathcal{D} = (\{A_\iota : \iota \in \mathcal{S}\}, \{P_\mathcal{D} : P \in \mathcal{P}\}, \{f_\mathcal{D} : f \in \mathcal{F}\}),$$

where

(a) A_ι is a set,
(b) $P_\mathcal{D} \subsetapprox A_{\iota_1} \times \cdots \times A_{\iota_n}$ is a fuzzy set with values in Ω,
(c) $f_\mathcal{D} : A_{\iota_1} \times \cdots \times A_{\iota_n} \to A_\iota$ is a map.

For a variable x let ι_x is a sort of x and we set $\mathcal{D}(X) = \prod_{x \in X} A_{\iota_x}$. This set then serves as a domains for interpretation of formulas and terms. In fact, by well known procedure which is based on inductive principle we can obtain that for a term t of a sort ι and for a formula ψ in J with free variables in X,

(a) an interpretation of t in $\mathcal{D}$ is a map $\|t\|_\mathcal{D} : \mathcal{D}(X) \to A_\iota$,
(b) an interpretation of ψ in $\mathcal{D}$ is a fuzzy set $\|\psi\|_\mathcal{D} \subsetapprox \mathcal{D}(X)$ with values in Ω, i.e. $\|\psi\|_\mathcal{D} : \mathcal{D}(X) \to \Omega$.

* The paper has been supported by the project MSM 6198898701 of the MŠMT ČR and 1M6798555601.

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 832–840, 2007.
springerlink.com

It is clear that an interpretation is based significantly on a set Ω. In case that $\Omega = (L, \wedge, \vee, \otimes, \rightarrow)$ is a complete residuated lattice (which is the most important case for fuzzy set theory) then this interpretation reflects naturally relations between logical and lattice connectives. For example, if $\psi \equiv \phi \Rightarrow \sigma$ then for any $\mathbf{a} \in \mathcal{D}(X)$ we have

$$\|\psi\|_{\mathcal{D}}(\mathbf{a}) = \|\phi\|_{\mathcal{D}}(\mathbf{a}) \rightarrow \|\sigma\|_{\mathcal{D}}(\mathbf{a})$$

and analogously

$$\|(\exists x)\psi\|_{\mathcal{D}}(\mathbf{a}) = \bigvee_{\mathbf{b} \in \mathcal{D}(X)} \|\psi[x/\mathbf{b}]\|_{\mathcal{D}}(\mathbf{a}).$$

Using this interpretation we can define various valuations of formulas and introduce a notion of semantics of a language J. Recall that if $\mathcal{D}$ is a model of J with Ω a residuated lattice then we can define a truth value $[\psi]_{\mathcal{D}}$ of a formula ψ in $\mathcal{D}$ such that

$$[\psi]_{\mathcal{D}} = \bigvee_{\mathbf{a} \in \mathcal{D}(X)} \|\psi\|_{\mathcal{D}}(\mathbf{a}).$$

Then for example we can say that $\mathcal{D}$ is a model of ψ (in symbol $\mathcal{D} \models \psi$) if $[\psi]_{\mathcal{D}} = 1$. A lot of other notions well known from fuzzy set theory can be then derived by using these models.

The principal aim of the paper is firstly to define an analogical model of a language J by using different categories of fuzzy sets. Namely we introduce a model of J based on an interpretation of formulas as fuzzy sets in some category of sets with similarity relations. For this purpose we have to define firstly a notion of a fuzzy set in this category. Further, we show that this category has also properties similar to properties of a topos. For example, for special subobjects (called *complete*) there are analogies of characteristic morphisms well known from a topos. It will be shown that among complete subobjects are also fuzzy sets and it follows that formulas can be also interpreted by using operations over characteristic morphisms. We finally show that both approaches to interpretation of formulas are principally identical.

2 Category of Sets with Similarity Relations

Let Ω be a complete residuated algebra. We will be dealing with a category $\mathbf{SetF}(\Omega)$ which consists of objects (A, δ) (called Ω-sets), where A is a set and δ is a similarity relation, i.e. a map $\delta : A \times A \rightarrow \Omega$ such that

(a) $\delta(x, x) = 1 = 1_{\Omega}$,
(b) $\delta(x, y) = \delta(y, x)$,
(c) $\delta(x, y) \otimes \delta(y, z) \leq \delta(x, z)$.

A morphism $f : (A, \delta) \rightarrow (B, \gamma)$ in $\mathbf{SetF}(\Omega)$ is a map $f : A \rightarrow B$ such that $\gamma(f(x), f(y)) \geq \delta(x, y)$ for all $x, y \in A$. In case that Ω is a complete Heyting algebra this category is well known and in this case it is a topos. In a general case it can be proved that a category $\mathbf{SetF}(\Omega)$ is complete. For example, if (A_i, δ_i) are Ω-sets for $i \in I$, their product in $\mathbf{SetF}(\Omega)$ is an Ω-set (A, δ), where $A = \prod_{i \in I} A_i$ and $\delta(\mathbf{a}, \mathbf{b}) =$

$\bigwedge_{i\in I}\delta_i(a_i,b_i)$. Let us mention that a terminal object in this category is $\mathbf{To}=(L,\wedge)$ with the unique morphism $!:(A,\delta)\to\mathbf{To}$ such that $!(a)=\delta(a,a)=1$. Moreover, a morphism f in this category is a monomorphism if and only if f is injective.

We introduce fuzzy sets objects in Ω-sets by the following more general definition.

Definition 1. *Let* $\mathbf{K}$ *be a category with* Ω*-sets as objects. Then a fuzzy set* s *in an object* (A,δ) *in a category* $\mathbf{K}$ *(in symbol* $s \subseteq_{\mathbf{K}} (A,\delta)$*) is a morphism*

$$s:(A,\delta)\to(\Omega,\leftrightarrow)$$

in $\mathbf{K}$*, where* $\alpha\leftrightarrow\beta=(\alpha\to\beta)\wedge(\beta\to\alpha)$.

For example, a fuzzy set $s \subseteq_{\mathbf{SetF}(\Omega)} (A,\delta)$ is a map $s:A\to\Omega$ such that $s(x)\otimes\delta(x,y)\le s(y)$ for all $x,y\in A$ (These objects are also called *extensional objects* (see [3])).

Theorem 1. *Let* $\mathcal{F}(A,\delta)=\{s: s \subseteq_{\mathbf{SetF}(\Omega)} (A,\delta)\}$. *Then* $\mathcal{F}$ *defines a contravariant functor* $\mathcal{F}:\mathbf{SetF}(\Omega)\to\mathbf{Set}$

In the category $\mathbf{SetF}(\Omega)$ a subobject classifier does not exist, in general. On the other hand, Höhle [4] proves that there exists an object which has very similar property and which can classify some special subobjects. This subobject is of the form

$$\Omega^*=(\{(\alpha,\beta)\in L\times L \mid \alpha\ge\beta\},\mu),$$
$$\mu((\alpha_1,\beta_1),(\alpha_2,\beta_2))=\alpha_1\otimes(\beta_1\to\beta_2)\wedge\alpha_2\otimes(\beta_2\to\beta_1).$$

It should be observed that μ does not satisfy the condition (a) from the definition of similarity relation. In fact, we have $\mu((\alpha,\beta),(\alpha,\beta))=\alpha$. From now the category $\mathbf{SetF}(\Omega)$ will be extended by the object (Ω^*,μ) and in that case for example a product of objects $(A_i,\delta_i), i\in I$ in that category will be $(\{(a_i)_i\in\prod_{i\in I}A_i:\delta_i(a_i,a_i)=\delta_j(a_j,a_j), i,j\in I\},\delta)$ and a morphism $!:(\Omega^*,\mu)\to\mathbf{To}$ is such that $!(\alpha,\beta)=\alpha$.

In [8] we have proved the following theorem which will be of importance for our purposes.

Theorem 2. *There exists a natural equivalence*

$$\zeta:\mathcal{F}(-)\cong Hom_{\mathbf{SetF}(\Omega)}(-,(\Omega^*,\mu)).$$

Recall how the natural equivalence is defined. For a fuzzy set $s \subseteq_{\mathbf{SetF}(\Omega)} (A,\delta)$ a characteristic morphism $\chi_s=\zeta_{(A,\delta)}(s):(A,\delta)\to\Omega^*$ is defined such that $\chi_s(a)=(1_\Omega,s(a))\in\Omega^*$ for any $a\in A$ and conversely, for a morphism $\chi:(A,\delta)\to\Omega^*$ a fuzzy set $s_\chi \subseteq_{\mathbf{SetF}(\Omega)} (A,\delta)$ is defined such that $s_\chi=pr_2\circ\chi$, where $pr_2:\Omega^*\to\Omega$ is a second projection map.

Let (A,δ) be an Ω-set. Then a set $S\subseteq A$ is called *complete* (in (A,δ)) if

$$S=\{a\in A:\bigvee_{x\in S}\delta(a,x)=1_\Omega\}.$$

Let $\text{Sub}(A,\delta)$ be the set of all subobjects of (A,δ) which are of a form (S,δ) where $S \subseteq A$ and let $\text{Sub}^c(A,\delta)$ be the set of all complete subobjects, i.e. subobjects (S,δ) such that S is complete in (A,δ). It can be proved simply that Sub and Sub^c are contravariant functors $\mathbf{SetF}(\Omega) \to \mathbf{Set}$. Moreover, for any Ω-set (A,δ) there exists a map $\psi_{(A,\delta)} : \text{Sub}(A,\delta) \to \mathcal{F}(A,\delta)$ such that $\psi_{(A,\delta)}(S,\delta)(a) = \bigvee_{x\in S} \delta(a,x)$, where $(S,\delta) \in \text{Sub}(A,\delta)$ and $a \in A$.

Then in [8] we have proved the following theorem:

Theorem 3. *Let (A,δ) be an Ω-set.*

(a) For any $(S,\delta) \in Sub(A,\delta)$ the following diagram commutes.

$$\begin{array}{ccc} (S,\delta) & \longrightarrow & (A,\delta) \\ {\scriptstyle !}\big\downarrow & & \big\downarrow {\scriptstyle \rho_{(S,\delta)}} \\ \mathbf{To} & \xrightarrow{\ \top\ } & (\Omega^*,\mu), \end{array}$$

where $\rho_{(S,\delta)}(a) = \zeta_{(A,\delta)}.\psi_{(A,\delta)}(S,\delta)(a) = (1_\Omega, \bigvee_{x\in S} \delta(a,x))$ for any $a \in A$.

(b) The above diagram is a pullback diagram if and only if S is a complete set in (A,δ).

From this theorem it follows that Ω-set Ω^* can be considered a *subobject classifier* for subobjects of (A,δ) which are of a type (S,δ), where S is a complete subset of (A,δ). This classification property will be used for defining an interpretation of formulas in this category.

An interpretation of formulas in objects of a category $\mathbf{SetF}(\Omega)$ we will start in a more classical way. The following simple but important lemma is of importance for definition of an interpretation of formulas.

Lemma 1. *Let (A,δ) be an Ω-set and let $s : A \to \Omega$ be a map. We define a map $\overline{s} : A \to \Omega$ such that $\overline{s}(a) = \bigvee_{x\in A} \delta(a,x) \otimes s(x)$ for all $a \in A$. Then*

(a) $\overline{s} \mathrel{\underset{\sim}{\subset}}_{\mathbf{SetF}(\Omega)} (A,\delta)$,

(b) If $s \mathrel{\underset{\sim}{\subset}}_{\mathbf{SetF}(\Omega)} (A,\delta)$ then $\overline{s} = s$,

(c) $\overline{s} = \bigwedge\{t : t \mathrel{\underset{\sim}{\subset}}_{\mathbf{SetF}(\Omega)} (A,\delta), t \geq s\}$.

3 Model of Fuzzy Logic in a Category $\mathrm{SetF}(\Omega)$

We start with a following definition of a model of J in a category $\mathbf{SetF}(\Omega)$.

Definition 2. *A model of a language J in a category $\mathbf{SetF}(\Omega)$ is*

$$\mathcal{D} = (\{(A_\iota,\delta_\iota) : \iota \in \mathcal{S}\}, \{P_\mathcal{D} : P \in \mathcal{P}\}, \{f_\mathcal{D} : f \in \mathcal{F}\}),$$

where

(a) (A_ι,δ_ι) is a Ω-set from a category $\mathbf{SetF}(\Omega)$,

(b) $P_\mathcal{D} \mathrel{\underset{\sim}{\subset}}_{\mathbf{SetF}(\Omega)} (A_{\iota_1},\delta_{\iota_1}) \times \cdots \times (A_{\iota_n},\delta_{\iota_n})$,

(c) $f_\mathcal{D} : (A_{\iota_1},\delta_{\iota_1}) \times \cdots \times (A_{\iota_n},\delta_{\iota_n}) \to (A_\iota,\delta_\iota)$ is a morphism in a category $\mathbf{SetF}(\Omega)$.

Now we define an interpretation of terms and formulas in a model $\mathcal{D}$. Let $t = f(t_1, \dots, t_n)$ be a term, where t_i are terms of sorts ι_i and f is a functional symbol of a sort $\iota_1 \times \dots \times \iota_n \to \iota$. Let X be a set of variables which contains all variables of t. Then by induction principle the interpretation $\|t_i\|_{\mathcal{D}} = \|t_i\|_{\mathcal{D},X}$ is already defined as a morphism $\|t_i\|_{\mathcal{D}} : (\mathcal{D}(X), \delta_X) \to (A_{\iota_i}, \delta_{\iota_i})$ in a category $\mathbf{SetF}(\Omega)$, where $(\mathcal{D}(X), \delta_X) = \prod_{x \in X}(A_{\iota_x}, \delta_{\iota_x})$ and we can consider a product morphism $\prod_i \|t_i\|_{\mathcal{D}} : (\mathcal{D}(X), \delta_X) \to \prod_i(A_{\iota_i}, \delta_{\iota_i})$. Then we set $\|t\|_{\mathcal{D}} = f_{\mathcal{D}} \circ \prod_i \|t_i\|_{\mathcal{D}}$.

The interpretation of formulas will be done in the following definition.

Definition 3. *Let ψ be a formula and let X be a set of variables containing all free variables of ψ.*

(a) Let $\psi \equiv (t_1 = t_2)$, where t_i are terms of a sort ι. Then for $\mathbf{a} \in \mathcal{D}(X)$ we set $\|\psi\|_{\mathcal{D}}(\mathbf{a}) = \bigvee_{\mathbf{x} \in \mathcal{D}(X)} \delta_\iota(\|t_1\|_{\mathcal{D}}(\mathbf{x}), \|t_2\|_{\mathcal{D}}(\mathbf{x})) \otimes \delta_X(\mathbf{a}, \mathbf{x})$.
(b) Let $\psi \equiv \sigma \Rightarrow \tau$. Then we set

$$\|\psi\|_{\mathcal{D}}(\mathbf{a}) = \bigvee_{\mathbf{x} \in \mathcal{D}(X)} (\|\sigma\|_{\mathcal{D}}(\mathbf{x}) \to \|\tau\|_{\mathcal{D}}(\mathbf{x})) \otimes \delta_X(\mathbf{x}, \mathbf{a}).$$

(c) Let $\psi \equiv \neg\sigma$. Then we set $\|\psi\|_{\mathcal{D}}(\mathbf{a}) = \|\sigma\|_{\mathcal{D}}(\mathbf{a}) \to 0$.
(d) Let Ψ be a set of formulas with free variables in a set X. Then we set $\| \bigvee \Psi\|_{\mathcal{D}}(\mathbf{a}) = \bigvee_{\psi \in \Psi} \|\psi\|_{\mathcal{D}}(\mathbf{a})$.
(e) Let Ψ be a set of formulas with free variables in a set X. Then we set $\| \bigwedge \Psi\|_{\mathcal{D}}(\mathbf{a}) = \bigwedge_{\psi \in \Psi} \|\psi\|_{\mathcal{D}}(\mathbf{a})$.
(f) Let $\psi \equiv (\exists x)\sigma$ where x is of a sort ι. Then we set

$$\|\psi\|_{\mathcal{D}}(\mathbf{a}) = \bigvee_{\mathbf{b} \in A_\iota} \|\sigma[\mathbf{b}/x]\|_{\mathcal{D}}(\mathbf{a}).$$

(g) Let $\psi \equiv (\forall x)\sigma$ where x is of a sort ι. Then we set

$$\|\psi\|_{\mathcal{D}}(\mathbf{a}) = \bigwedge_{\mathbf{b} \in A_\iota} \|\sigma[\mathbf{b}/x]\|_{\mathcal{D}}(\mathbf{a}).$$

This definition is correct as the following theorem states.

Theorem 4. *Let ψ be a formula with free variables contained in a set of variables X. Then $\|\psi\|_{\mathcal{D}} \subseteq_{\sim \mathbf{SetF}(\Omega)} \mathcal{D}(X)$.*

For illustration we show a proof of theorem for some formulas. It is clear that the proof for (a) and (b) follows from Lemma 2.1

(c) Let $\psi \equiv \neg\sigma$. For any $\mathbf{a}, \mathbf{b} \in \mathcal{D}(X)$ we have $\delta_X(\mathbf{a}, \mathbf{b}) \otimes \|\sigma\|_{\mathcal{D}}(\mathbf{b}) \leq \|\sigma\|_{\mathcal{D}}(\mathbf{a})$ and it follows that

$$\|\neg\sigma\|_{\mathcal{D}}(\mathbf{a}) \leq \delta_X(\mathbf{a}, \mathbf{b}) \otimes \|\sigma\|_{\mathcal{D}}(\mathbf{b}) \to 0 = \delta_X(\mathbf{a}, \mathbf{b}) \to (\|\sigma\|_{\mathcal{D}}(\mathbf{b}) \to 0).$$

Hence, $\delta_X(\mathbf{a}, \mathbf{b}) \otimes \|\neg\sigma\||_{\mathcal{D}}(\mathbf{a}) \leq \|\neg\sigma\|_{\mathcal{D}}(\mathbf{b})$.

(e) We have

$$\|\bigwedge \Psi\|_{\mathcal{D}}(\mathbf{a}) \otimes \delta_X(\mathbf{a},\mathbf{b}) = (\bigwedge_{\psi\in\Psi} \|\psi\|_{\mathcal{D}}(\mathbf{a})) \otimes \delta_X(\mathbf{a},\mathbf{b})$$

$$\leq \bigwedge_{\psi\in\Psi} (\|\psi\|_{\mathcal{D}}(\mathbf{a}) \otimes \delta_X(\mathbf{a},\mathbf{b})) \leq \bigwedge_{\psi\in\Psi} \|\psi\|_{\mathcal{D}}(\mathbf{b}) = \|\bigwedge \Psi\|_{\mathcal{D}}(\mathbf{b}).$$

Now let $\mathcal{D}$ be a model of J in a category $\mathbf{SetF}(\Omega)$. A truth valuation of formulas from J can be then defined as follows.

Definition 4. *Let ψ be a formula with a set of free variables in a set X and let $\mathcal{D}$ be a model of J in a category $\mathbf{SetF}(\Omega)$. Then we set*

$$[\psi]_{\mathcal{D}} = \bigvee_{\mathbf{a}\in\mathcal{D}(X)} \|\psi\|_{\mathcal{D}}(\mathbf{a}).$$

It can be proved that this definition is correct, i.e. it does not depend on a set X of free variables containing all free variables of ψ.

Moreover, we say that a formula ψ is true in a model $\mathcal{D}$ of J in a category $\mathbf{SetF}(\Omega)$ (in symbol $\mathcal{D} \models \psi$) if $[\psi]_{\mathcal{D}} = 1$. The following theorem is an example of formulas which are true in all models $\mathcal{D}$ in a category $\mathbf{SetF}(\Omega)$.

Theorem 5. *Let $\mathcal{D}$ be a model of J in a category $\mathbf{SetF}(\Omega)$.*

(a) $\mathcal{D} \models (\psi \Rightarrow (\sigma \Rightarrow \psi))$.
(b) $\mathcal{D} \models ((\psi \Rightarrow \sigma) \Rightarrow (\neg\sigma \Rightarrow \neg\psi))$.

Proof: (a) We have

$$\|\psi \Rightarrow (\sigma \Rightarrow \psi)\|(\mathbf{a}) = \bigvee_{\mathbf{x}\in\mathcal{D}(X)} (\|\psi\|(\mathbf{x}) \to \|\sigma \Rightarrow \psi\|(\mathbf{x})) \otimes \delta_X(\mathbf{x},\mathbf{a}) \geq$$

$$\geq \bigvee_{\mathbf{x},\mathbf{y}\in\mathcal{D}(X)} (\|\psi\|(\mathbf{x}) \to (\|\sigma\|(\mathbf{y}) \to \|\psi\|(\mathbf{y})) \otimes \delta_X(\mathbf{x},\mathbf{y})) \otimes \delta_X(\mathbf{a},\mathbf{x}) \geq$$

$$\geq \bigvee_{\mathbf{x},\mathbf{y}\in\mathcal{D}(X)} (\|\psi\|(\mathbf{x}) \to (\|\sigma\|(\mathbf{y}) \to \|\psi\|(\mathbf{y}))) \otimes \delta_X(\mathbf{x},\mathbf{y}) \otimes \delta_X(\mathbf{a},\mathbf{x}) \geq$$

$$\geq \bigvee_{\mathbf{x}\in\mathcal{D}(X)} (\|\psi\|(\mathbf{x}) \to (\|\sigma\|(\mathbf{x}) \to \|\psi\|(\mathbf{x}))) \otimes \delta_X(\mathbf{a},\mathbf{x}) =$$

$$\bigvee_{\mathbf{x}\in\mathcal{D}(X)} \delta_X(\mathbf{a},\mathbf{x}) = 1.$$

(b) Since $\alpha \to \beta \leq (\beta \to 0) \to (\alpha \to 0)$ holds in any residuated lattice, we have

$$\bigvee_{\mathbf{y}\in\mathcal{D}(X)} (\|\psi\|(\mathbf{y}) \to \|\sigma\|(\mathbf{y})) \otimes \delta_X(\mathbf{x},\mathbf{y}) \leq \bigvee_{\mathbf{y}\in\mathcal{D}(X)} ((\|\sigma\|(\mathbf{y}) \to 0) \to$$

$$\to (\|\psi\|(\mathbf{y}) \to 0)) \otimes \delta_X(\mathbf{y},\mathbf{x}).$$

It then follows that

$$
\begin{aligned}
&\|(\psi \Rightarrow \sigma) \Rightarrow (\neg\sigma \Rightarrow \neg\psi)\|(\mathbf{a}) = \\
&= \bigvee_{\mathbf{x}\in\mathcal{D}(X)} (\|\psi \Rightarrow \sigma\|(\mathbf{x}) \rightarrow \|\neg\sigma \Rightarrow \neg\psi\|(\mathbf{x})) \otimes \delta_X(\mathbf{a},\mathbf{x}) = \\
&= \bigvee_{\mathbf{x}\in\mathcal{D}(X)} ((\bigvee_{\mathbf{y}\in\mathcal{D}(X)} (\|\psi\|(\mathbf{y}) \rightarrow \|\sigma\|(\mathbf{y})) \otimes \delta_X(\mathbf{x},\mathbf{y})) \rightarrow (\bigvee_{\mathbf{y}\in\mathcal{D}(X)} ((\|\sigma\|(\mathbf{y}) \rightarrow 0) \rightarrow \\
&\rightarrow (\|\psi\|(\mathbf{y}) \rightarrow 0)) \otimes \delta_X(\mathbf{x},\mathbf{y}))) \otimes \delta_X(\mathbf{a},\mathbf{x}) = \bigvee_{\mathbf{x}\in\mathcal{D}(X)} \delta_X(\mathbf{a},\mathbf{x}) = 1.
\end{aligned}
$$

4 Characteristic Morphisms and Formula Interpretations

Another possiblity of formula interpretation in a category $\mathbf{SetF}(\Omega)$ is based on a notion of a characteristic morphism $|\psi| : (\mathcal{D}(X), \delta_X) \rightarrow \Omega^*$ of a formula ψ with a set of free variables contained in a set X of variables. According to Theorem 2.2, $\zeta^{-1}_{(A,\delta)}(|\psi|)$ is then a fuzzy set in (A, δ) and we want to investigate a relationship between that fuzzy set and a fuzzy set $\|\psi\|$. A construction of $|\psi|$ is based on logical connectives and it is done classically by inductive principle.If for example a binary logical connective $\bigtriangledown$ appears in a formula ψ, then some interpretation of $\bigtriangledown$ has to be defined first, which is a morphism

$$(\Omega^*, \mu) \times (\Omega^*, \mu) \xrightarrow{[\bigtriangledown]} (\Omega^*, \mu).$$

Definition of $[\bigtriangledown]$ is derived as an analogy of classical logical connectives in a Boolean logic. For example if $\bigtriangledown = \wedge$ or $\vee$ or $\neg$, respectively, then $[\bigtriangledown]$ is defined such that

$$
\begin{aligned}
[\wedge]((a,b),(a,c)) &= (a, b \wedge c), \\
[\vee]((a,b),(a,c)) &= (a, b \vee c) \\
[\neg](a,b) &= (a, a \otimes (b \rightarrow 0),
\end{aligned}
$$

respectively. To interprete formulas by characteristic morphism we summarize firstly some relationship between monomorphisms and fuzzy sets in a category $\mathbf{SetF}(\Omega)$. Let $f : (X, \sigma) \rightarrowtail (A, \delta)$ be a monomorphism in a category $\mathbf{SetF}(\Omega)$. That monomorphism then defines an object $(\sigma(X), \delta)$ from the set $\mathrm{Sub}(A, \delta)$. Then according to 2.3, there exists a morphism $\rho_{(\sigma(X),\delta)}$ from (A, δ) to Ω^* which is (according to 2.2) a characteristic morphism of some fuzzy set $\triangle_f : (A, \delta) \rightarrow \Omega$. A construction of $\triangle_f$ is presented on the following diagram.

$$
\begin{array}{ccccc}
(X,\sigma) & \xrightarrow{f} & (A,\delta) & & \\
\mapsto\downarrow & & \| & & \\
(\sigma(X),\delta) & \xrightarrow{\subseteq} & (A,\delta) & \xrightarrow{\zeta^{-1}_{(A,\delta)}(\rho_{(\sigma(X),\delta)})=\triangle_f} & (\Omega,\leftrightarrow) \\
 & & \downarrow \rho_{(\sigma(X),\delta)} & & \\
 & & (\Omega^*,\mu) & &
\end{array}
$$

Let us consider a monomorphism $1_{(A,\delta)} \times 1_{(A,\delta)} : (A,\delta) \to (A,\delta) \times (A,\delta)$. According to the above construction a fuzzy set $\triangle_{(A,\delta)} := \triangle_{1_{(A,\delta)} \times 1_{(A,\delta)}} \subsetsim (A,\delta) \times (A,\delta)$ is defined such that $\triangle_{1_{(A,\delta)} \times 1_{(A,\delta)}}(a,b) = \bigvee_{x \in A} \delta(a,x) \wedge \delta(b,x)$. This fuzzy set will be used for interpretation of a fragment of fuzzy logic by characteristic morphisms as follows.

Definition 5. *Let ψ be a formula which contains only $\wedge, \vee, \neg$ and $=$ as logical symbols and let free variables of ψ be contained in a set X of variables. Then the interpretation $|\psi|_{\mathcal{D}}$ is defined as a morphism $|\psi|_{\mathcal{D}} : \mathcal{D}(X) \to (\Omega^*, \mu)$ by the induction as follows.*

(a) Let $\psi \equiv t_1 = t_2$. The morphisms $\|t_i\|_X : \mathcal{D}(X) \to (A,\delta)$ are already defined by inductive principle, and hence $|\psi|_{\mathcal{D}}$ is defined as the composition of the following morphisms:

$$|\psi|_{\mathcal{D}} : \mathcal{D}(X) \xrightarrow{\|t_1\|_X \times \|t_2\|_X} (A,\delta) \times (A,\delta) \xrightarrow{\triangle_{(A,\delta)}} (\Omega^*, \mu).$$

(b) Let $\psi \equiv \psi_1 \wedge \psi_2$. Then for every i there are interpretations $|\psi_i|_{\mathcal{D}} : \mathcal{D}(X) \to (\Omega^*, \mu)$ and we put

$$|\psi|_{\mathcal{D}} = [\wedge] \circ (|\psi_1|_{\mathcal{D}} \times |\psi_2|_{\mathcal{D}}).$$

(c) Let $\psi \equiv \psi_1 \vee \psi_2$.Then for every i there are interpretations $|\psi_i|_{\mathcal{D}} : \mathcal{D}(X) \to (\Omega^*, \mu)$ and we put

$$|\psi|_{\mathcal{D}} = [\vee] \circ (|\psi_1|_{\mathcal{D}} \times |\psi_2|_{\mathcal{D}}).$$

(d) Let $\psi = \neg\phi$. Then there exists a morphism $|\phi|_{\mathcal{D}} : \mathcal{D}(X) \to (\Omega^*, \mu)$ and we set

$$|\psi|_{\mathcal{D}} : \mathcal{D}(X) \xrightarrow{|\phi|_{\mathcal{D}}} (\Omega^*, \mu) \xrightarrow{[\neg]} (\Omega^*, \mu).$$

The following theorem describes a relationship between two types of interpretations, i.e. between $\|\psi\|_{\mathcal{D}}$ and $|\psi|_{\mathcal{D}}$.

Theorem 6. *Let ψ be a formula which contains only $\wedge, \vee, \neg$ and $=$ as logical symbols and let free variables of ψ be contained in a set X of variables. Then*

$$\zeta_{\mathcal{D}(X)}(\|\psi\|_{\mathcal{D}}) = |\psi|_{\mathcal{D}}.$$

References

1. Goguen, J.A., L-fuzzy sets. *J.Math. Anal. Appli. 18*(1967), 145-174.
2. Higgs, D., *A category approach to Boolean-valued set theory.* University of Waterloo (1973).
3. Höhle, U., Presheaves over GL-monoids. *Non-Classical Logic and Their Applications to Fuzzy Subsets*, Kluwer Academic Publ., Dordrecht, New York(1995), 127-157.
4. Höhle, U., M-Valued sets and sheaves over integral, commutative cl-monoids. *Applications of Category Theory to Fuzzy Subsets*, Kluwer Academic Publ., Dordrecht, Boston (1992), 33-72.
5. Höhle, U., Classification of Subsheaves over GL-algebras. *Proceedings of Logic Colloquium 98 Prague*, Springer Verlag (1999).
6. Höhle, U., Commutative, residuated l-monoids. *Non-Classical Logic and Their Applications to Fuzzy Subsets*, Kluwer Academic Publ. Dordrecht, New York (1995), 53–106.

7. Höhle, U., Monoidal closed categories, weak topoi and generalized logics. *Fuzzy Sets and Systems* **42**(1991), 15–35.
8. Močkoř, J., Complete subobjects of fuzzy sets over MV-algebras. *Czech.Math.J.* **129(54)**(2004), 379–392.
9. Novák,V., Perfilijeva, I., Močkoř,J., *Mathematical principles of fuzzy logic* Kluwer Academic Publishers,Boston,Dordrecht, London (1999).

Fixed Points and Solvability of Systems of Fuzzy Relation Equations

Irina Perfilieva

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic
Irina.Perfilieva@osu.cz

Abstract. The problem of solvability of a system of fuzzy relation equations with $\sup -*$-composition is considered in a finite semilinear space over a residuated lattice. In this setting the problem of solvability given above is similar to the problem of solvability of a system of linear equations in the form $A\mathbf{x} = \mathbf{b}$. We put emphasis on a right-hand side vector $\mathbf{b}$ and consider the problem of solvability as a problem of characterization of all vectors $\mathbf{b}$ for which the original system of (fuzzy relation) equations is solvable. We prove that a system of equations with $\sup -*$-composition is solvable if and only if $\mathbf{b}$ is a fixed point of the shrivel operator (introduced in this paper). Moreover, a set of all fixed points is a semi-linear subspace of an original space. Some other results are presented as well.

Keywords: Residuated lattice, solvability of a system of equations, semilinear space, fixed point.

1 Introduction

Linear behavior of fuzzy systems has been recently discovered in [4]. Two cases may occur: a behavior of a system is characterized by fuzzy IF-THEN rules (expert knowledge and similar) or it is characterized by a set of input-output pairs of fuzzy sets (monitoring, collecting knowledge, etc.). In the second case, the problem of solving a respective system of fuzzy relation equations arises [5, 7, 8, 11]. However, it has not yet been observed (besides, probably in [2]) that the mentioned problem is similar to the problem of solving systems of linear equations. In this contribution, we show that a system of fuzzy relation equations can be considered as a system of linear-like equations in a semilinear space over a residuated lattice. We will concentrate on systems with $\sup -*$-composition because they are more popular in practical applications.

The specificity of a system of fuzzy relation equations with $\sup -*$-composition is the following: if it is solvable then it has the uniquely expressed greatest solution. Therefore, we put emphasis on a solvability. In this contribution we change an angle under which this problem is usually considered (see, e.g., [1, 3, 5, 7, 10]). We concentrate on a characterization of possible right-hand sides which make a system solvable with a given left-hand side. We prove that the right-hand side must be a fixed point of

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 841–849, 2007.
springerlink.com

a special shrivel operator introduced in this paper (roughly speaking, this operator is given by the composition of the matrix of the system and its inverse).

The paper is organized as follows: in Section 2 we give an introduction to semi-linear spaces, in Section 3 we give formulation of a problem and in the subsequent Section 4 we prove fixed point theorems. Then we prove a number of results about isomorphisms between factor spaces and finally consider a structure of L^n determined by the fixed points.

2 Semi-linear Spaces

Let $\mathcal{L} = \langle L, \vee, \wedge, *, \rightarrow, 0, 1\rangle$ be a residuated lattice. Then its $\vee$-reduct

$$\mathcal{L}_\vee = \langle L, \vee, *, 0, 1\rangle$$

is a commutative semiring [4].

A *semilinear space* is taken as a semimodule over a $\vee$-reduct of a residuated lattice supplied with an additional external operation.

Definition 1. *Let $\mathcal{L}_\vee$ be the $\vee$-reduct of a residuated lattice $\mathcal{L}$. Let $\langle A, \vee, \mathbf{0}\rangle$ be a $\vee$-semilattice with the least element.*

We say that $\mathcal{A} = \langle A, \vee, \bar{*}, \mathbf{0}\rangle$ is an idempotent semi-linear space over a residuated lattice (shortly, a semi-linear space) if an external multiplication $\bar{*} : L \times A \rightarrow A$ is defined so that

(a) $\langle A, \vee, \mathbf{0}\rangle$ is a semimodule over $\mathcal{L}_\vee$ (a semimodule multiplication is $\bar{*}$),
(b) for each $\lambda \in L$ a mapping $h_\lambda : A \rightarrow A$, defined by $h_\lambda(\mathbf{a}) = \lambda \bar{*} \mathbf{a}$, has a residual, i.e. an isotone mapping $g_\lambda : A \rightarrow A$ such that

$$(g_\lambda \circ h_\lambda)(\mathbf{a}) \geq \mathbf{a}, \tag{1}$$

$$(h_\lambda \circ g_\lambda)(\mathbf{a}) \leq \mathbf{a}. \tag{2}$$

It follows from the definition that a carrier A of a semi-linear space is a partially ordered set where

$$\mathbf{a} \leq \mathbf{b} \iff \mathbf{a} \vee \mathbf{b} = \mathbf{b}.$$

We say that a semi-linear space $\mathcal{A}$ is *lattice ordered* if its carrier is a lattice with respect to the order given above. The elements of a semi-linear space are called vectors and denoted by bold characters, and elements of L are called scalars and denoted by Greek characters.

Let us define another external operation $\bar{\rightarrow}$ on A:

$$\lambda \bar{\rightarrow} \mathbf{a} = g_\lambda(\mathbf{a}).$$

It is true that for any $\mathbf{a} \in A$

$$\lambda \bar{\rightarrow} \mathbf{a} = \max\{\mathbf{b} \in A \mid \lambda \bar{*} \mathbf{b} \leq \mathbf{a}\}$$

if and only if the right-hand side exists. Therefore, if $\mathcal{A}$ has the greatest element $\mathbf{1}$ then for any $\mathbf{a} \in A$

$$0 \bar{\rightarrow} \mathbf{a} = \mathbf{1}.$$

Lemma 1. *Let $\mathcal{L}$ be a residuated lattice and $\mathcal{A} = \langle A, \vee, \bar{*}, \mathbf{0}\rangle$ a lattice based semi-linear space over $\mathcal{L}$. Then the operation $\bar{\rightarrow}$ is distributive over $\wedge$, i.e. for all $\mathbf{a}, \mathbf{b} \in A$ and $\lambda \in L$*

$$\lambda \bar{\rightarrow} (\mathbf{a} \wedge \mathbf{b}) = (\lambda \bar{\rightarrow} \mathbf{a}) \wedge (\lambda \bar{\rightarrow} \mathbf{b}).$$

The lemma below exposes a powerful property which external operations have - *adjunction*.

Lemma 2. *Let $\mathcal{A}$ be a lattice based semi-linear space over $\mathcal{L}$. Then external operations $\bar{*}$ and $\bar{\rightarrow}$ constitute an adjoint pair, i.e. for any $\mathbf{a}, \mathbf{b} \in A$ and $\mu \in L$ the adjunction property*

$$\mu \bar{*} \mathbf{b} \leq \mathbf{a} \iff \mathbf{b} \leq \mu \bar{\rightarrow} \mathbf{a} \tag{3}$$

holds true.

By (3), we may establish many other useful properties of semi-linear spaces in the same way as it is done in [6] for residuated lattices. For example,

$$(\mu \vee \lambda) \bar{\rightarrow} \mathbf{a} = (\mu \bar{\rightarrow} \mathbf{a}) \wedge (\lambda \bar{\rightarrow} \mathbf{a})$$

holds true for for any $\mathbf{a} \in A$ and $\mu, \lambda \in L$.

The following are examples of a semi-linear space which we will use in the sequel.

Example 1. 1. A reduct $\langle L, \vee, *, 0\rangle$ of a residuated lattice $\mathcal{L}$ is a lattice ordered semi-linear space over itself. The operation $\bar{\rightarrow}$ is equal to $\rightarrow$.

2. Let $\mathcal{L} = \langle L, \vee, \wedge, *, \rightarrow, 0, 1\rangle$ be a residuated lattice on L. The set of all n-dimensional vectors L^n, $n \geq 1$, such that

$$(a_1, \ldots, a_n) \leq (b_1, \ldots, b_n) \iff a_1 \leq b_1, \ldots, a_n \leq b_n$$

is a lattice ordered semi-linear space over $\mathcal{L}$ where for arbitrary $\lambda \in L$

$$\lambda \bar{*} (a_1, \ldots, a_n) = (\lambda * a_1, \ldots, \lambda * a_n),$$
$$\lambda \bar{\rightarrow} (a_1, \ldots, a_n) = (\lambda \rightarrow a_1, \ldots, \lambda \rightarrow a_n).$$

The least element in L^n is the vector $\mathbf{0} = (0, \ldots, 0)$ and the greatest element in L^n is the vector $\mathbf{1} = (1, \ldots, 1)$. The lattice ordered semi-linear space L^n will be referred to as a *semi-linear vector space*.

3. Let $\mathcal{L}$ be a residuated lattice on L, $X \neq \emptyset$ and L^X a set of all L-valued functions on X such that

$$f \leq g \iff f(x) \leq g(x), \quad x \in X.$$

Put

$$(\lambda \bar{*} f)(x) = \lambda * f(x), \quad x \in X,$$
$$(\lambda \bar{\rightarrow} f)(x) = \lambda \rightarrow f(x), \quad x \in X.$$

The least element $\mathbf{0}$ in L^X is the function identically equal to zero and the greatest element $\mathbf{1}$ in L^X is the function identically equal to one. The lattice ordered semi-linear space L^X will be referred to as a *semi-linear functional space*.

Definition 2. *Let $\langle A_1, \vee, \bar{*}, \mathbf{0}\rangle$ and $\langle A_2, \vee, \bar{*}, \mathbf{0}\rangle$ be two semi-linear spaces over $\mathcal{L}$. A homomorphism H from A_1 to A_2 is a mapping $H : A_1 \mapsto A_2$ such that for all $\mathbf{a}, \mathbf{b} \in A_1$, $\lambda \in L$*

$$H(\mathbf{a} \vee \mathbf{b}) = H(\mathbf{a}) \vee H(\mathbf{b}),$$
$$H(\lambda \bar{*} \mathbf{a}) = \lambda \bar{*} H(\mathbf{a}),$$
$$H(\mathbf{0}) = \mathbf{0},$$

Example 2. Let $A_1 = L^m$, $m \geq 1$, and $A_2 = L^n$, $n \geq 1$, be semi-linear vector spaces over $\mathcal{L}$ (see Example 1, case 2). Let R be an $(n \times m)$-matrix with elements r_{ij} from L. We define the homomorphism $H_R : L^m \mapsto L^n$ so that $H_R(\mathbf{a}) = (H_R(\mathbf{a})_1, \ldots, H_R(\mathbf{a})_n)$ where

$$H_R(\mathbf{a})_i = \bigvee_{j=1}^{m} (r_{ij} * a_j), \quad i = 1, \ldots, n. \tag{4}$$

It is easy to see that $H_R(\mathbf{0}) = \mathbf{0}$ and for all $\mathbf{a}, \mathbf{b} \in A_1$, $\lambda, \mu \in L$

$$H_R(\lambda \bar{*} \mathbf{a} \vee \mu \bar{*} \mathbf{b}) = \lambda \bar{*} H_R(\mathbf{a}) \vee \mu \bar{*} H_R(\mathbf{b}).$$

Definition 3. *Let $\langle A_1, \vee, \bar{*}, \mathbf{0}\rangle$ and $\langle A_2, \vee, \bar{*}, \mathbf{0}\rangle$ be two semi-linear spaces over $\mathcal{L}$ and $H : A_1 \mapsto A_2$ a homomorphism. A residual (of H) homomorphism G from A_2 to A_1 is a mapping $G : A_2 \mapsto A_1$ such that for all $\mathbf{a} \in A_1$, $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b} \in A_2$*

$$\mathbf{b}_1 \leq \mathbf{b}_2 \Rightarrow G(\mathbf{b}_1) \leq G(\mathbf{b}_2)$$

and

$$(G \circ H)(\mathbf{a}) \geq \mathbf{a}, \tag{5}$$
$$(H \circ G)(\mathbf{b}) \leq \mathbf{b}. \tag{6}$$

Moreover, if A_1 and A_2 have the greatest elements then $G(\mathbf{1}) = \mathbf{1}$.

It is not difficult to prove that if a homomorphism $H : L^m \mapsto L^n$ is determined by a $(n \times m)$-matrix R such that $H = H_R$, then its residual homomorphism $G_R : L^n \mapsto L^m$ is unique and given as follows:

$$G_R(\mathbf{b}) = \bigwedge_{i=1}^{n} (r_{ij} \rightarrow b_i), \quad j = 1, \ldots, m. \tag{7}$$

Moreover, for all $\mathbf{b}, \mathbf{c} \in A_2$, $\lambda, \mu \in L$

$$G_R((\lambda \bar{\rightarrow} \mathbf{b}) \wedge (\mu \bar{\rightarrow} \mathbf{c})) = (\lambda \bar{\rightarrow} G_R(\mathbf{b})) \wedge (\mu \bar{\rightarrow} G_R(\mathbf{b})).$$

3 Formulation of a Problem

In what follows, we fix a complete residuated lattice with a support L and consider L^m and L^n, $m, n \geq 1$, as semilinear spaces over $\mathcal{L}$.

Let $A = (a_{ij})$ be a $(n \times m)-$ matrix and $\mathbf{b} = (b_1, \dots, b_n)$ vector, both have components from L. The following system of equations is considered with respect to an unknown vector $\mathbf{x} = (x_1, \dots, x_m)$:

$$\bigvee_{j=1}^{m} (a_{ij} * x_j) = b_i, \quad i = 1, \dots n, \tag{8}$$

or in a vector form

$$A \circ \mathbf{x} = \mathbf{b}.$$

It has been mentioned in Introduction that analogs of system of equations (8) have been considered in the literature devoted to fuzzy sets and systems, see e.g. [1, 3, 5, 7, 10]. From these sources we took the following result.

Proposition 1. *System (8) is solvable if and only if the vector* $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_m)$ *such that*

$$\hat{x}_j = \bigwedge_{i=1}^{n} (a_{ij} \to b_i), \quad j = 1, \dots, m \tag{9}$$

is its greatest solution.

Moreover, if $\mathbf{x}^1$ and $\mathbf{x}^2$ are solutions to (8) then $\mathbf{x}^1 \vee \mathbf{x}^2$ where the operation $\vee$ is taken componentwise is a solution too. Therefore, the set of solutions of (8) form a $\vee$-semilattice with the "unit" element.

Proposition 1 turns the problem of finding a solution to (8) to the investigation whether (8) is solvable. The latter can be considered in two formulations:

(a) given a $(n \times m)-$ matrix A and vector $\mathbf{b} = (b_1, \dots, b_n)$, investigate whether (8) is solvable,
(b) given a $(n \times m)-$ matrix A, characterize all vectors $\mathbf{b} = (b_1, \dots, b_n)$ such that (8) is solvable.

We will investigate the problem of solvability of (8) in the formulation (b). Let us remark that both formulations are similar to the classical ones encountering in linear algebra.

3.1 Fixed Points of the Shrivel Operator

Let $A = (a_{ij})$ be a $(n \times m)-$ matrix and $\mathbf{b} = (b_1, \dots, b_n)$ vector, both have components from L. The following mapping from L^n to L^n is determined by

$$(A \circ A^{\to})(\mathbf{b})_i = \bigvee_{j=1}^{m} (a_{ij} * \bigwedge_{l=1}^{n} (a_{lj} \to b_l)). \tag{10}$$

We say that (10) defines the *shrivel operator* $(A \circ A^{\rightarrow}) : L^n \mapsto L^n$ which maps $(b_1, \ldots, b_n) \in L^n$ onto $((A \circ A^{\rightarrow})(\mathbf{b})_1, \ldots, (A \circ A^{\rightarrow})(\mathbf{b})_n) \in L^n$. It is easy to see that operator $(A \circ A^{\rightarrow})$ is a composition $H_A \circ G_A$ (cf. (6)) of the residual homomorphism G_A given by A and the homomorphism H_A given by A too.

The following proposition easily follows from Proposition 1.

Proposition 2. *System (8) is solvable if and only if*

$$(A \circ A^{\rightarrow})(\mathbf{b}) = \mathbf{b}$$

or that $\mathbf{b}$ *is a fixed point of operator* $(A \circ A^{\rightarrow})$.

Denotation. $\mathcal{F}(A \circ A^{\rightarrow})$ is the set of fixed points of $(A \circ A^{\rightarrow})$.

Let us remark that this type of transformation is not new, it has being investigated in the literature by different names: in (KolMas,blyth) it is called as a composition of a mapping and its residual. Therefore, we may use some known facts from that literature reformulating them according to our terminology.

Proposition 3. *Let* $A = (a_{ij})$ *be a* $(n \times m)-$ *matrix with components from L,* $(A \circ A^{\rightarrow}) : L^n \mapsto L^n$ *the corresponding shrivel operator. Then the following holds true:*

a) for each $\mathbf{b} \in L^n$, $(A \circ A^{\rightarrow})(\mathbf{b}) \leq \mathbf{b}$,
b) for each $\mathbf{b} \in L^n$, $(A \circ A^{\rightarrow})(\mathbf{b})$ *is a fixed point of of* $(A \circ A^{\rightarrow})$,
c) for each $\mathbf{b}_1, \mathbf{b}_2 \in L^n$, $\mathbf{b}_1 \leq \mathbf{b}_2$ *implies* $(A \circ A^{\rightarrow})(\mathbf{b}_1) \leq (A \circ A^{\rightarrow})(\mathbf{b}_2)$,
d) $\mathbf{b} \in L^n$ *is a fixed point of* $(A \circ A^{\rightarrow})$ *if and only if there exists* $\mathbf{x} \in L^m$ *such that* $A \circ \mathbf{x} = \mathbf{b}$ *(in other words,* $\mathbf{b}$ *is representable by a linear combination of vector-columns of matrix A),*
e) if $\mathbf{b}_1, \mathbf{b}_2 \in L^n$ *are fixed points of* $(A \circ A^{\rightarrow})$ *then* $\mathbf{b}_1 \vee \mathbf{b}_2$ *is a fixed point too.*

4 Fixed Points of $(A \circ A^{\rightarrow})$

In the following theorem we will characterize a set of fixed points of $(A \circ A^{\rightarrow})$ as a subspace of L^n.

Theorem 1. *Let* $A = (a_{ij})$ *be a* $(n \times m)-$ *matrix with components from L,* $(A \circ A^{\rightarrow}) : L^n \mapsto L^n$ *the corresponding shrivel operator. Then* $\mathcal{F}(A \circ A^{\rightarrow})$ *is a subspace of* L^n.

Proof. At first, we will prove that $\mathcal{F}(A \circ A^{\rightarrow})$ is non-empty and contains the zero vector $\mathbf{0} \in L^n$. Indeed, for each $i = 1, \ldots n$

$$(A \circ A^{\rightarrow})(\mathbf{0}) = \bigvee_{j=1}^{m} (a_{ij} * \bigwedge_{l=1}^{n} (a_{lj} \rightarrow 0)) \leq \bigvee_{j=1}^{m} (a_{ij} * (a_{ij} \rightarrow 0)) = 0.$$

In order to prove that $\mathcal{F}(A \circ A^{\rightarrow})$ is a subspace of L^n we verify that if $\mathbf{b}, \mathbf{b}_1, \mathbf{b}_2 \in L^n$ are fixed points of $(A \circ A^{\rightarrow})$ then $\mathbf{b}_1 \vee \mathbf{b}_2$ and $\lambda * \mathbf{b}$ are fixed points too. Due to Proposition 3, case d), it is sufficient to verify the second claim. Indeed,

$$(A \circ A^{\rightarrow})(\lambda * \mathbf{b})_i = \bigvee_{j=1}^{m}(a_{ij} * \bigwedge_{l=1}^{n}(a_{lj} \rightarrow \lambda * b_l)) \geq \bigvee_{j=1}^{m}(a_{ij} * \lambda * \bigwedge_{l=1}^{n}(a_{lj} \rightarrow b_l)) =$$
$$\lambda * \bigvee_{j=1}^{m}(a_{ij} * \bigwedge_{l=1}^{n}(a_{lj} \rightarrow b_l)) = \lambda * (A \circ A^{\rightarrow})(\mathbf{b})_i$$

where we used the BL-inequality: $(a \rightarrow b) * (c \rightarrow d) \leq (a * c) \rightarrow (b * d)$. Together with Proposition 3, case a) we have

$$(A \circ A^{\rightarrow})(\lambda * \mathbf{b})_i = \lambda * (A \circ A^{\rightarrow})(\mathbf{b})_i, \quad i = 1, \ldots, n.$$

Corollary 1. *$\mathcal{F}(A \circ A^{\rightarrow})$ is a linear envelope of vector-columns of matrix A, i.e.*

$$\mathcal{F}(A \circ A^{\rightarrow}) = L(\mathbf{a}_1, \ldots, \mathbf{a}_m)$$

where for each $j = 1 \ldots, m$, $\mathbf{a}_j = (a_{1j}, \ldots, a_{nj})$.

Proof. This fact easily follows from Proposition 3, case d) and the theorem above.

Corollary 2. *Assume that a set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ contains a basis of L^n. Then each vector from L^n is a fixed point of $(A \circ A^{\rightarrow})$, i.e. belongs to $\mathcal{F}(A \circ A^{\rightarrow})$.*

Proof. If a set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ contains a basis of L^n then $L(\mathbf{a}_1, \ldots, \mathbf{a}_m) = L^n$.

Corollary 3. *Assume that $m = n$ and a set of vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ contains a basis of L^n. Then the transformation of L^n given by $A^{\rightarrow}$ is inverse to the transformation of L^n given by A, i.e. for all $\mathbf{b} \in L^n$, $(A \circ A^{\rightarrow})(\mathbf{b}) = \mathbf{b}$.*

5 Theorem of Isomorphisms

Let $\mathbf{b}_0$ be a fixed point of $(A \circ A^{\rightarrow})$, i.e. $\mathbf{b}_0 \in L(\mathbf{a}_1, \ldots, \mathbf{a}_m)$. We denote

$$L^{\uparrow}(\mathbf{b}_0) = \{\mathbf{b} \in L^n \mid (A \circ A^{\rightarrow})(\mathbf{b}) = \mathbf{b}_0\} \tag{11}$$
$$L^{\downarrow}(\mathbf{b}_0) = \{\mathbf{x} \in L^m \mid (A \circ \mathbf{x} = \mathbf{b}_0\} \tag{12}$$

Let us remark that by the choice of $\mathbf{b}_0$, $L^{\uparrow}(\mathbf{b}_0)$ is not empty, and by Proposition 3, case d), $L^{\downarrow}(\mathbf{b}_0)$ is not empty too.

Let $L^n/L^{\uparrow}$, $L^m/L^{\downarrow}$ be respective factor spaces of L^n and L^m (considered over $\mathcal{L}_{\vee} = \langle L, \vee, *, 1\rangle$) where we can define

$$L^{\uparrow}(\mathbf{b}_0) \vee L^{\uparrow}(\mathbf{c}_0) = L^{\uparrow}(\mathbf{b}_0 \vee \mathbf{c}_0)$$
$$\lambda * L^{\uparrow}(\mathbf{b}_0) = L^{\uparrow}(\lambda * \mathbf{b}_0)$$

and similar for $L^m/L^{\downarrow}$.

Theorem 2. *The following isomorphisms between the above introduced factor spaces can be established:*

(i) $L^n/L^{\uparrow} \cong L(\mathbf{a}_1, \ldots, \mathbf{a}_m)$,
(ii) $L^m/L^{\downarrow} \cong L(\mathbf{a}_1, \ldots, \mathbf{a}_m)$,
(iii) $L^n/L^{\uparrow} \cong L^m/L^{\downarrow}$.

6 Structure of L^n

Let $A = (a_{ij})$ be a $(n \times m)-$ matrix with components from L, $(A \circ A^{\rightarrow}) : L^n \mapsto L^n$ the corresponding shrivel operator. Then

$$L^n = \bigcup_{\mathbf{b}_0 \in \mathcal{F}(A \circ A^{\rightarrow})} L^{\uparrow}(\mathbf{b}_0) = \bigcup_{\mathbf{b}_0 \in L(\mathbf{a}_1,\ldots,\mathbf{a}_m)} L^{\uparrow}(\mathbf{b}_0)$$

and

$$L^{\uparrow}(\mathbf{b}_0) = \{\mathbf{b} \in L^n \mid (A \circ A^{\rightarrow})\mathbf{b} = \mathbf{b}_0\}.$$

Let us describe a structure of $L^{\uparrow}(\mathbf{b}_0)$. From the above,

$$\mathbf{b} \in L^{\uparrow}(\mathbf{b}_0) \iff \bigvee_{j=1}^{m} (a_{ij} * \underbrace{\bigwedge_{l=1}^{n} (a_{lj} \rightarrow b_l)}_{\hat{x}_j}) = b_i^0, \quad i = 1, \ldots, n$$

so that $(\hat{x}_1, \ldots, \hat{x}_m) \in L^{\downarrow}(\mathbf{b}_0)$.

Moreover, each vector $\mathbf{x} \in L^{\downarrow}(\mathbf{b}_0)$ determines

$$L^{\uparrow}_{\mathbf{x}}(\mathbf{b}_0) = \{\mathbf{b} \in L^n \mid \bigwedge_{l=1}^{n} (a_{lj} \rightarrow b_l) = x_j, \quad j = 1, \ldots, m\}.$$

From [9] we know that $L^{\uparrow}_{\mathbf{x}}(\mathbf{b}_0)$ is a $\wedge$-semi-lattice with the least element $\mathbf{b}_0$. Therefore, we may prove the following theorem which characterizes the structure of L^n with respect to the shrivel operator $(A \circ A^{\rightarrow})$.

Theorem 3. *Let $A = (a_{ij})$ be a $(n \times m)-$ matrix with components from L, $(A \circ A^{\rightarrow}) : L^n \mapsto L^n$ the corresponding shrivel operator. Then*

(i) $L^n = \bigcup_{\mathbf{b}_0 \in L(\mathbf{a}_1,\ldots,\mathbf{a}_m)} L^{\uparrow}(\mathbf{b}_0)$,
(ii) $L^{\uparrow}(\mathbf{b}_0) = \bigcup_{\mathbf{x} \in L^{\downarrow}(\mathbf{b}_0)} L^{\uparrow}_{\mathbf{x}}(\mathbf{b}_0)$,
(iii) $L^{\uparrow}_{\mathbf{x}}(\mathbf{b}_0)$ *is a $\wedge$-semi-lattice with the least element* $\mathbf{b}_0$,

7 Conclusion

The problem of solvability of a system of fuzzy relation equations with $\sup -*$-composition has been considered in a finite semilinear space L^n over a residuated lattice L. We put emphasis on a right-hand side vector $\mathbf{b}$ and considered the problem of solvability as a problem of characterization of all vectors $\mathbf{b}$ such that the respective system of equations is solvable. We proved the following results:

– a system of equations with $\sup -*$-composition is solvable if and only if $\mathbf{b}$ is a fixed point of the respective shrivel operator,
– a set of all fixed points is a semi-linear subspace of an original space,
– if a set of vectors-columns of a matrix of a system contains a basis of L^n then each vector from L^n is a fixed point of the shrivel operator,

– L^n can be represented as a union of pairwise disjoint sets of vectors which shriveled up to a respective fixed point.

Acknowledgments. The research was partially supported by projects MSM 6198898701 and 1M0572 of the MŠMT ČR.

References

1. Bandler, W. and L. Kohout: 1980, Semantics of implication operators and fuzzy relational products. Int. J. Man-Machine Studies 12, 89116.
2. Cuninghame-Green, R. A. and K. Cechlarova: 1995, Residuation in fuzzy algebra and some applications. Fuzzy Sets and Systems 71, 227239.
3. De Baets, B.: 2000, Analytic solution methods for fuzzy relation equations. In: D. Dubois and H. Prade (eds.): The Handbooks of Fuzzy Sets Series, Vol.1. Dordrecht: Kluwer, pp. 291340.
4. Di Nola, A., A. Lettieri, I. Perfilieva, and V. Novak: 2007, Algebraic Analysis of Fuzzy Systems. Fuzzy Sets and Systems 158, 122.
5. DiNola, A., S. Sessa, W. Pedrycz, and E. Sanchez: 1989, Fuzzy Relation Equations and Their Applications to Knowledge Engineering. Boston: Kluwer.
6. Hohle, U.: 1995, Commutative residuated l-monoids. In: U. Hohle and E. P. Klement (eds.): Non-Classical Logics and Their Applications to Fuzzy Subsets. A Handbook of the Mathematical Foundations of Fuzzy Set Theory. Dordrecht: Kluwer, pp. 53106.
7. Miyakoshi, M. and M. Shimbo: 1985, Solutions of composite fuzzy relation equations with triangular norms. Fuzzy Sets and Systems 16, 5363.
8. Pedrycz, W.: 1990, Inverse problem in fuzzy relational equations. Fuzzy Sets and Systems 36, 277291.
9. Perfilieva, I. and L. Noskova: 2007, System of fuzzy relation equations with inf ! composition in semi-linear spaces: maximal solutions. Fuzzy Sets and Systems p. submitted.
10. Perfilieva, I. and A. Tonis: 2000, Compatibility of Systems of Fuzzy Relation Equations. Int. J. of General Systems 29, 511528.
11. Sanchez, E.: 1976, Resolution of composite fuzzy relation equations. Information and Control 30, 3848.

Towards a Proof Theory for Basic Logic

Thomas Vetterlein

European Centre for Soft Computing
Edificio Científico-Tecnológico, C/ Gonzalo Gutiérrez Quirós s/n
33600 Mieres, Spain
Thomas.Vetterlein@softcomputing.es
www.softcomputing.es

Abstract. Proof systems based on rules with the property that all formulas contained in the assumptions are contained as subformulas in the conclusion as well, are particularly suitable for automated proof search. Systems of this kind were found for several well-known fuzzy logics. However, for BL (Basic Logic), the logic of continuous t-norms and their residua, the situation is less satisfactory. We consider two proof systems for BL which fulfill the desired property in quite contrasting ways.

1 Introduction

Fuzzy logics are usually presented in the Hilbertian style; propositions are proved from a rather comprehensive set of axioms by means of a usually small number of rules, among which there is modus ponens and often not more. The logics which we have in mind are those whose propositions are interpreted by values from the real unit interval and whose language contains a conjunction interpreted by a left-continuous t-norm, an implication interpreted by the corresponding residuum, and a falsity constant interpreted by the real value 0. As a rich source of information about these logic, we recommend P. Hájek's monograph [8].

For automated proof search, Hilbert-style systems are clearly inappropriate. The problem is that modus ponens does not have the subformula property; the formula which disappears when using this rule cannot be reconstructed from the conclusion. In recent years, proof systems for various fuzzy logics were presented consisting exclusively of rules which do have the subformula property [2, 3, 4, 6, 10, 11]. They are called analytic, as they offer the possibility to decompose a proposition whose provability is to be checked step by step into its atomic constituents. In these systems, the question of provability of a proposition is reducible to the much easier tractable question about the validity of statements which do not involve any logical connective.

Analytic proof systems were, for instance, found for Łukasiewicz logic [10], product logic [11], Gödel logic [4], and the logic MTL [2]. The tool on which these calculi are based are generalizations of Gentzen's sequents. In particular, hypersequents are used, which are multisets of sequents. Hypersequents were further generalized to r-hypersequents in [6], and proof systems based on

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 850–860, 2007.
springerlink.com

r-hypersequents were found for all the three standard extensions of BL, in a way that the logical rules coincide [6].

Each of the mentioned calculi features a set logical rules of the minimal possible size and moreover an easily apprehensible set of structural rules. It is an open question if a similarly elegant proof system exists also for Basic Logic, the logic of continuous t-norms and their respective residua. In this paper, we discuss recent work on this problem.

There are two proof systems for BL which, at least formally, come relatively close to the desired type. The first one is the calculus RHBL from [5], which we will shortly review (Section 3). The second one is our calculus rHMŁ [15]. We will give an introduction to rHMŁ and compile its basic properties (Section 4) and in particular explain its proof search capabilities (Section 5).

2 Basic Logic - The Usual Hilbert-Style Formulation

Basic (Fuzzy) Logic, or BL for short, was introduced by P. Hájek [8]. We summarize the basic facts.

The propositional version of BL uses the language $\odot, \rightarrow, 0$. An evaluation of BL is a structure-preserving map from the algebra of propositions to an algebra $([0,1]; \odot, \rightarrow, 0)$, where $[0,1]$ is the real unit interval, $\odot$ is a continuous t-norm, and $\odot, \rightarrow$ form an adjoint pair. The valid propositions are those being assigned 1 by all evaluations.

The notion of a proof in BL is as follows. Using the axiom schemes

$$
\begin{aligned}
&\text{(A1)}\ \ [(\alpha \rightarrow \beta) \odot (\beta \rightarrow \gamma)] \rightarrow (\alpha \rightarrow \gamma),\\
&\text{(A2)}\ \ \alpha \odot \beta \rightarrow \alpha,\\
&\text{(A3)}\ \ \alpha \odot \beta \rightarrow \beta \odot \alpha,\\
&\text{(A4)}\ \ [\alpha \odot (\alpha \rightarrow \beta)] \rightarrow [\beta \odot (\beta \rightarrow \alpha)],\\
&\text{(A5a)}\ \ [(\alpha \odot \beta) \rightarrow \gamma] \rightarrow [\alpha \rightarrow (\beta \rightarrow \gamma)],\\
&\text{(A5b)}\ \ [\alpha \rightarrow (\beta \rightarrow \gamma)] \rightarrow [(\alpha \odot \beta) \rightarrow \gamma],\\
&\text{(A6)}\ \ 0 \rightarrow \alpha,\\
&\text{(A7)}\ \ [((\alpha \rightarrow \beta) \rightarrow \gamma) \odot ((\beta \rightarrow \alpha) \rightarrow \gamma)] \rightarrow \gamma,
\end{aligned}
$$

we prove propositions by means of the modus ponens, i.e. the rule

$$\frac{\alpha \quad \alpha \rightarrow \beta}{\beta}.$$

We have (weak) standard completeness: A proposition is provable in BL exactly if it is valid in BL.

Our intention is to replace this Hilbert-style proof system by an alternative one. As a preparatory step, we will in the remainder of this section comment on the semantics on which BL is based, and subsequently we will recall known approaches for proof systems of fuzzy logics.

Validity in BL refers canonically to the set of all continuous t-norms. As exhibited e.g. in [13], it is possible to restrict this set without enlarging the set of valid propositions. Namely, the variety of BL-algebra, which is generated by all continuous t-norm algebras, is already generated by all those t-norm algebras which are based on finite ordinal sums of Łukasiewicz t-norms. Indeed, for any non-valid proposition φ, there is an evaluation v into a t-norm algebra such that $v(\varphi) < 1$; obviously, the range of v may be assumed the ordinal sum of finitely many Łukasiewicz and product algebras; moreover, by cutting off unused parts of the negative cones constituting the product algebras, we may replace the image of v by a t-norm algebra composed from Łukasiewicz algebras only.

So let for $k \geq 1$

$$\begin{aligned} S_k \; &= \{(n,r) \in \mathbb{Z} \times \mathbb{R} \colon \;\; n = 0, \;\; -1 \leq r \leq 0 \\ &\qquad \text{or } \; -(k-1) \leq n \leq -1, \;\; -1 \leq r < 0\}; \end{aligned}$$

endow S_k with the lexicographical order; define

$$(m,r) \odot (n,s) \; = \begin{cases} (m, (r+s) \vee -1) & \text{if } m = n, \\ (m,r) \wedge (n,s) & \text{else,} \end{cases} \tag{1}$$

$$(m,r) \rightarrow (n,s) \; = \begin{cases} (n,s) & \text{if } m > n, \\ (m, s-r) & \text{if } m = n \text{ and } r > s, \\ (0,0) & \text{if } (m,r) \leq (n,s) \end{cases} \tag{2}$$

for any $(m,r), (n,s) \in S_k$; and let $e = (0,0)$ and $z_k = (-(k-1), -1)$. (Here, $r \wedge s$ and $r \vee s$ denote the smaller and the larger of two reals r and s, respectively.) Then, in accordance with our previous remarks, a proposition is valid in BL if and only if it is assigned e by all evaluations into the algebra $(S_k; \leq, \odot, \rightarrow, z_k, e)$.

Rather than basing the validity of BL on the countably many algebras $S_1, S_2, \ldots$, we may also use a single one. A possibility is to take

$$S_\infty \; = \{(n,r) \in \mathbb{N} \times \mathbb{R} \colon \, -1 \leq r < 0\} \cup \{\infty\}, \tag{3}$$

where ∞ is a new element added on top. The operations $\odot$ and $\rightarrow$ may be defined similarly to (1) and (2), respectively, but some special care is needed for the element ∞. The zero element is $(-1, 0)$, the one element is ∞.

Note that the necessity in this case to add a single isolated element is caused by the presence of the constant 0. For Hoop Logic, the logic based on basic hoops [7], the situation would be different; we could simply take the union of the S_k, $k < \omega$, and use the unmodified definitions (1) and (2).

Next, let us address the topic of alternative proof systems for fuzzy logics. We do not refer especially to BL here. The concept which we are going to explain is an advancement of Gentzen's sequent calculus and has been elaborated in numerous papers in recent years. The so-called hypersequents are due to Avron [1] and Pottinger [14]; in [3], sequents-of-relations were introduced; a common generalization both these concepts are the r-hypersequents, introduced in [6]. It is the latter notion with which we will deal here.

In fuzzy logics, an r-sequent $\Gamma \lesseqqgtr \Delta$ consists of two finite multisets of propositions $\Gamma = \gamma_1, ..., \gamma_m$ and $\Delta = \delta_1, ..., \delta_n$ and moreover a relation symbol $\lesseqqgtr$, which is either $\leq$ or $<$. The validity of r-sequents is defined individually for each logic; in the general case, we may say the following. For a given evaluation v of the logic's propositions into a set of truth values S, a sequent is called valid under v if $\bigotimes_{i=1}^{m} v(\gamma_i) \lesseqqgtr \bigotimes_{i=1}^{n} v(\delta_i)$, where $\otimes$ is a binary isotone associative function operating on some upper bounded totally ordered set $T \supseteq S$, and $\lesseqqgtr$ refers to the order or strict order of T, respectively. For instance, in case $S = S_1 = [-1, 0]$, $\otimes$ can be the addition of reals and $T = \mathbb{R}^-$; in general, however, $\otimes$ and T can be determined rather arbitrarily.

Furthermore, an r-hypersequent $\Gamma_1 \lesseqqgtr \Delta_1 | \ldots | \Gamma_n \lesseqqgtr \Delta_n$ is a multiset of r-sequents. A hypersequent $\mathcal{H}$ is called valid under some evaluation v if one of the sequents contained in $\mathcal{H}$ is valid under v. $\mathcal{H}$ is called valid if $\mathcal{H}$ is valid under all evaluations.

Now, a rule is a pair of a finite set of r-hypersequents – the assumptions –, and a single r-hypersequent – the conclusion. A calculus is a collection of rules. Proofs are finite trees of instances of rules, such that every assumption of a rule is the conclusion of an immediately preceding rule. A proposition φ is defined to be provable if the r-hypersequent $\emptyset \leq \varphi$ is the conclusion at root of some proof.

3 The Bova-Montagna Calculus RHBL

It seems that for the logic BL, the concept of an r-hypersequent-based proof system as explained in the previous section and applied successfully to a large number of fuzzy logics, must be modified in some way if analyticity is the property we aim at. One possible way to generalize the involved notions was recently proposed by S. Bova and F. Montagna [5].

Let us roughly summarize this approach. First of all, validity in BL is considered with respect to the single algebra whose base set is S_∞ given by (3). The main idea in [5] is to generalize the concept of an r-hypersequent. Namely, in view of the special structure of S_∞, further binary relations apart from $\leq$ and $<$ can easily be defined. The calculus RHBL is based on r-hypersequents, where the relation symbol may be one of $\ll$ or $\preccurlyeq$ or $\preccurlyeq_z$, where $z \in \mathbb{Z}$. For instance, $\ll$ applies to pairs of multisets of length ≤ 1 and refers to strict inequality of the first components of elements from S_∞.

The calculus RHBL consists of eight logical rules. The rules fulfill a property even more desirable than the subformula property: they are invertible. There is a rule for each connective and each side, separately for $\ll$-hypersequents and for the general case. The idea when introducing e.g. the formula $\alpha \odot \beta$, is that there are different proof branches according to the different possibilities for the mutual relationship of α and β. For instance, in the premises of the rule introducing $\alpha \odot \beta$, five different possibilities are distinguished how α is related to β.

By means of the calculus RHBL, it is possible to decompose a proposition step by step into atomic r-hypersequents. Proving the latter's' validity is not possible within RHBL; valid r-hypersequents are axioms. However, a method is described in [5] how the validity of an atomic r-hypersequent is checked effectively by means of a linear program.

4 The Calculus rHMŁ for a Conservative Extension of BL

A proof system for BL based on a different idea was proposed in [15]. This system actually does not refer to BL directly, but to MŁ, the so-called Logic of Multiples of the Łukasiewicz t-Norm.

The language of MŁ contains like BL the connectives $\odot, \to$ and the constant 0; in addition, however, there is a unary connective $\triangledown$. Propositions of MŁ are interpreted in t-norm algebras which are ordinal sums of Łukasiewicz algebras; and $\triangledown$ is interpreted by the function mapping a truth value t to the greatest idempotent below t. So in particular, a proposition of MŁ not containing the new connective $\triangledown$ is valid if and only if it is valid in BL; that is, MŁ is a conservative extension of BL. So a calculus suitable for MŁ may be used to check provability in BL equally well as provability in MŁ.

An extension of BL similar to our MŁ was proposed earlier by Hájek [9] and others. However, in contrast e.g. to the logic defined in [9], MŁ is not based on all continuous t-norm algebras, but only on ordinal sums of Łukasiewicz algebras; we use $\triangledown$ to rule out those ordinal sums in which the product algebra appears.

Hájek's ideas were further elaborated by Montagna in [12], where a so-called storage operator was introduced for any appropriate class of MTL-algebras. In fact, our connective $\triangledown$ is Montagna's storage operator applied to ordinal multiples of Łukasiewicz algebras.

Although we think that MŁ is interesting in itself, we admit that the primary reason to consider this logic is that we may define an analytic proof system for MŁ apparently much easier than for BL. However, concerning the question if it makes sense to consider a logic like MŁ, we should mention Montagna's paper a second time; the introduction of [12] contains interesting hints concerning the interpretation of a storage operator, so in particular of our $\triangledown$.

The calculus for MŁ which we are going to discuss here, is called rHMŁ. We list its key features for easy reference:

(i) rHMŁ is based on r-hypersequents as introduced in [6], that is, with relation symbols $\leq$ and $<$ only.

(ii) The interpretation of r-hypersequents generalizes the one of rHŁ in [6].

(iii) There is a rule for each binary connective and each side, separately for the case that $\triangledown$ is the outermost connective of the introduced formula or not. For instance, a proposition $\alpha \odot \beta$ is composed from α and β, whereas $\triangledown(\alpha \odot \beta)$ is composed from $\triangledown\alpha$ and $\triangledown\beta$.

(iv) The rules introducing the binary logical connectives are invertible.

(v) The structural rules treat r-hypersequents which contain literals only, where a literal is an atom or an atom to which the connective $\triangledown$ is applied.

(vi) rHMŁ has elementary axioms, that is, the axioms are r-sequents with at most one literal in the antecedent and succedent.

(vii) Among the structural rules, there is a rule distinguishing the cases $\alpha \leq \beta$ and $\beta < \alpha$, where α, β are literals. We are referring to the rule (Cut$\leq$/$>$) below, which may be called an analytic cut.

We will now explain the logic MŁ and its calculus in detail.

As in the case of BL, it is more convenient to evaluate propositions of MŁ in the algebras S_k rather than in t-norm algebras. So a proposition φ of MŁ is valid if for every evaluation v into $(S_k; \odot, \rightarrow, \triangledown, z_k, e)$, we have $v(\varphi) = e$, where $\triangledown\colon S_k \rightarrow S_k$ is defined by

$$\triangledown(n, r) \;=\; \begin{cases} (n, -1) & \text{if } r < 0; \\ (0, 0) & \text{if } (n, r) = (0, 0) \end{cases}$$

for $(n, r) \in S_k$.

The validity of r-hypersequents will be based on algebras different from the S_k; we use the same "trick" as in the case of the calculus rHŁ for Łukasiewicz logic defined in [6]. Let for every $k \geq 1$

$$\begin{aligned} T_k \;&= \{(n, r) \in \mathbb{Z} \times \mathbb{R}\colon\;\; n = 0, \;\; r \leq 0 \\ &\qquad \text{or } \; -(k-1) \leq n \leq -1, \;\; r < 0\}; \end{aligned}$$

endow T_k with the lexicographical order; for $(m, r), (n, s) \in T_k$, define

$$(m, r) \cdot (n, s) \;=\; \begin{cases} (m, r + s) & \text{if } m = n, \\ (m, r) \wedge (n, s) & \text{else,} \end{cases}$$

$$(m, r) \rightarrow (n, s) \;=\; \begin{cases} (n, s) & \text{if } m > n; \\ (m, s - r) & \text{if } m = n \text{ and } r > s, \\ (0, 0) & \text{if } (m, r) \leq (n, s); \end{cases}$$

and let $e = (0, 0)$. Then an r-sequent $\alpha_1, \ldots, \alpha_m \leqq \beta_1, \ldots, \beta_n$, where $m, n \geq 0$, is valid under the evaluation v with range S_k if

$$v(\alpha_1) \cdot \ldots \cdot v(\alpha_m) \;\leqq\; v(\beta_1) \cdot \ldots \cdot v(\beta_n);$$

here, S_k is considered a subset of T_k and the product $\cdot$ refers to T_k; moreover, the product of the empty set is understood to be e. – The validity of r-hypersequents is defined accordingly.

We next define the calculus itself. In each rule, the three dots at the beginning of each r-hypersequent replace an arbitrary finite multiset of r-sequents, for each rule uniformly. Furthermore, the symbol $\leqq$ is to be replaced by $\leq$ or $<$, for each rule uniformly.

In addition, an r-hypersequent will be called quasiatomic if every proposition contained in it is of the form α or $\triangledown\alpha$ for an atom α. Finally, for a multiset Γ and an atom α, $\Gamma \setminus \alpha$ denotes the multiset Γ with all occurrences of α and $\triangledown\alpha$ removed.

Definition 1. *The logical rules of the calculus rHMŁ are the following:*

$$(\odot\text{l})\ \frac{\dots \mid \Gamma, \alpha, \beta \leqq \Delta \qquad \dots \mid \Gamma, \nabla\alpha \leqq \Delta \mid \Gamma, \nabla\beta \leqq \Delta}{\dots \mid \Gamma, \alpha \odot \beta \leqq \Delta}$$

$$(\nabla\odot\text{l})\ \frac{\dots \mid \Gamma, \nabla\alpha \leqq \Delta \mid \Gamma, \nabla\beta \leqq \Delta}{\dots \mid \Gamma, \nabla(\alpha \odot \beta) \leqq \Delta}$$

$$(\odot\text{r})\ \frac{\dots \mid \Gamma \leqq \Delta, \alpha, \beta \mid \Gamma \leqq \Delta, \nabla\alpha \qquad \dots \mid \Gamma \leqq \Delta, \alpha, \beta \mid \Gamma \leqq \Delta, \nabla\beta}{\dots \mid \Gamma \leqq \Delta, \alpha \odot \beta}$$

$$(\nabla\odot\text{r})\ \frac{\dots \mid \Gamma \leqq \Delta, \nabla\alpha \qquad \dots \mid \Gamma \leqq \Delta, \nabla\beta}{\dots \mid \Gamma \leqq \Delta, \nabla(\alpha \odot \beta)}$$

$$(\to\text{l})\ \frac{\dots \mid \Gamma \leqq \Delta \mid \Gamma, \beta \leqq \Delta, \alpha \qquad \dots \mid \Gamma \leqq \Delta \mid \beta < \alpha}{\dots \mid \Gamma, \alpha \to \beta \leqq \Delta}$$

$$(\nabla\to\text{l})\ \frac{\dots \mid \Gamma \leqq \Delta \mid \beta < \alpha \qquad \dots \mid \Gamma, \nabla\beta \leqq \Delta \mid \alpha \leq \beta}{\dots \mid \Gamma, \nabla(\alpha \to \beta) \leqq \Delta}$$

$$(\to\text{r})\ \frac{\dots \mid \Gamma \leqq \Delta \qquad \dots \mid \Gamma, \alpha \leqq \Delta, \beta \mid \alpha \leq \beta}{\dots \mid \Gamma \leqq \Delta, \alpha \to \beta}$$

$$(\nabla\to\text{r})\ \frac{\dots \mid \Gamma \leqq \Delta \mid \beta < \alpha \qquad \dots \mid \Gamma \leqq \Delta, \nabla\beta \mid \alpha \leq \beta}{\dots \mid \Gamma \leqq \Delta, \nabla(\alpha \to \beta)}$$

$$(\nabla\nabla\text{l})\ \frac{\dots \mid \Gamma, \nabla\alpha \leqq \Delta}{\dots \mid \Gamma, \nabla\nabla\alpha \leqq \Delta} \qquad (\nabla\nabla\text{r})\ \frac{\dots \mid \Gamma \leqq \Delta, \nabla\alpha}{\dots \mid \Gamma \leqq \Delta, \nabla\nabla\alpha}$$

The following structural rules of rHMŁ refer to quasiatomic r-hypersequents. Any expression $\nabla\alpha$ *in a rule's conclusion, where* $\alpha = \nabla\beta$ *for some atom* β*, is meant to be* $\nabla\beta$*.*

$$(\text{A1})\ \emptyset \leq \emptyset \qquad (\text{A2})\ \alpha \leq \alpha \qquad (\text{A3})\ 0 \leq \alpha \qquad (\text{A4})\ 0 < \emptyset$$

$$(\text{EW})\ \frac{\dots}{\dots \mid \Gamma \leqq \Delta} \qquad (\text{EC})\ \frac{\dots \mid \Gamma \leqq \Delta \mid \Gamma \leqq \Delta}{\dots \mid \Gamma \leqq \Delta}$$

$$(\text{Cut}{\leq}/{>})\ \frac{\dots \mid \Gamma \leq \Delta \qquad \dots \mid \Delta < \Gamma}{\dots},$$

where Δ and Γ contain at most one literal,
and any variable in $\Delta \cup \Gamma$ appears in the side r-hypersequent

$$(\text{O})\ \frac{\dots \mid \Gamma \setminus \alpha \leqq \Delta \setminus \alpha}{\dots \mid \Gamma \leqq \Delta \mid \nabla\alpha \leq \nabla\beta},$$

where α and β are atoms such that α or $\nabla\alpha$ is in $\Gamma \cup \Delta$, and β or $\nabla\beta$ is in $\Gamma \cup \Delta$

$$(\nabla\text{l})\ \frac{\ldots \mid \Gamma, \alpha \lesseqgtr \Delta}{\ldots \mid \Gamma, \nabla\alpha \lesseqgtr \Delta} \qquad (\nabla\text{lr})\ \frac{\ldots \mid \beta \lesseqgtr \alpha}{\ldots \mid \nabla\beta \leq \nabla\alpha} \qquad (\nabla\text{r})\ \frac{\ldots \mid \emptyset \leq \alpha}{\ldots \mid \emptyset \leq \nabla\alpha}$$

$$(\text{wl})\ \frac{\ldots \mid \Gamma \lesseqgtr \Delta}{\ldots \mid \Gamma, \alpha \lesseqgtr \Delta} \qquad (\text{w0l})\ \frac{\ldots \mid \Gamma \leq \Delta}{\ldots \mid \Gamma, 0 < \Delta}$$

$$(\text{w}\nabla\text{l})\ \frac{\ldots \mid \alpha_1, \ldots, \alpha_n \leq \Delta}{\ldots \mid \alpha_1, \ldots, \alpha_n, \nabla\beta < \Delta \mid \nabla\alpha_1 < \nabla\beta \mid \ldots \mid \nabla\alpha_n < \nabla\beta \mid \emptyset \leq \nabla\beta},$$

where $n \geq 0$; in case $n = 0$ the r-sequents "$\ldots < \nabla\beta$" are omitted

$$(\text{M})\ \frac{\ldots \mid \Gamma_1 \leq \Delta_1 \qquad \ldots \mid \Gamma_2 \leq \Delta_2}{\ldots \mid \Gamma_1, \Gamma_2 \leq \Delta_1, \Delta_2} \qquad (\text{S})\ \frac{\ldots \mid \Gamma_1, \Gamma_2 \lesseqgtr \Delta_1, \Delta_2}{\ldots \mid \Gamma_1 \lesseqgtr \Delta_1 \mid \Gamma_2 \lesseqgtr \Delta_2}$$

$$(\text{S}<)\ \frac{\ldots \mid \Gamma, \alpha_1, \ldots, \alpha_n \leq \Delta_1, \Delta_2}{\ldots \mid \Gamma \leq \Delta_1 \mid \alpha_1, \ldots, \alpha_n < \Delta_2 \mid \nabla\alpha_1 < \nabla\beta \mid \ldots \mid \nabla\alpha_n < \nabla\beta},$$

where (i) $n \geq 1$, and (ii) $\beta \in \Gamma \cup \Delta_1$

It is tedious, but not really difficult to check that rHMŁ is sound and that the logical rules are all invertible, w.r.t. the above defined validity. In particular, all propositions provable in rHMŁ are valid in MŁ.

To see that rHMŁ is actually complete, we rely on the fact that, by backwards application of the logical rules, we may decompose a proposition step by step until we arrive at a (possibly quite large) number of quasiatomic r-hypersequents. It is, furthermore, not so trivial to see that the quasiatomic r-hypersequents are provable by means of the structural rules of rHMŁ. For this proof, we refer to [15]. Note that among the structural rules – in contrast to RHBL not among the logical rules –, there is the rule (Cut≤/>) which distinguishes the relative order of two atoms. Taken all mentioned facts together, we get weak standard completeness for rHMŁ:

Theorem 1. *The calculus rHMŁ is sound and complete for MŁ: A proposition α is valid in MŁ if and only if α is provable in rHMŁ.*

5 Proof Search with the Calculus rHMŁ

By backwards application of the proof rules, we may use the calculus rHMŁ to check if a proposition of MŁ is valid or not; in particular, we may check the validity of a proposition of BL. We will outline the method.

First Fact. (i) All logical rules of rHMŁ are invertible. (ii) Applying the logical rules successively upwards terminates after finitely many steps with quasiatomic r-hypersequents.

This procedure is in particular applicable to the r-hypersequent $\emptyset \leq \alpha$ for any given proposition α. It follows that the question if α is valid in MŁ is reducible to the question if certain quasiatomic r-hypersequents are valid.

In what follows, we call an r-sequent $\Gamma \lesseqq \Delta$ basic if Γ is either empty or contains one proposition of the form $\nabla\alpha$ for an atom α, and similarly for Δ.

Second Fact. The following rules are admissible in rHMŁ and furthermore invertible:

(ExtCut≤/>)

$$\frac{\begin{array}{l} \ldots \mid \Gamma \setminus \alpha \lesseqq \Delta \setminus \alpha \mid \nabla\alpha \leq \nabla\beta \\ \qquad\qquad \ldots \mid \Gamma \setminus \beta \lesseqq \Delta \setminus \beta \mid \nabla\beta \leq \nabla\alpha \\ \qquad\qquad\qquad\qquad \ldots \mid \Gamma \lesseqq \Delta \mid \nabla\alpha < \nabla\beta \mid \nabla\beta < \nabla\alpha \end{array}}{\ldots \mid \Gamma \lesseqq \Delta},$$

where Γ and Δ are quasiatomic, and
both atoms α and β are subformulas in $\Gamma \cup \Delta$;

$$\text{(ExtCut}\leq/>0\text{)} \quad \frac{\ldots \mid \Gamma \setminus \alpha \lesseqq \Delta \setminus \alpha \mid \nabla\alpha < \emptyset \qquad \ldots \mid \Gamma \lesseqq \Delta \mid \emptyset \leq \nabla\alpha}{\ldots \mid \Gamma \lesseqq \Delta},$$

where Γ and Δ are quasiatomic
and the atom α is a subformula in $\Gamma \cup \Delta$.

This means that a proof search for a quasiatomic r-hypersequent $\mathcal{H}$ goes as follows. If $\mathcal{H}$ contains a non-basic r-sequent $\Gamma \lesseqq \Delta$ such that the distinct atoms α and β appear in it and such that $\mathcal{H}$ does not contain both $\nabla\alpha < \nabla\beta$ and $\nabla\beta < \nabla\alpha$, apply (ExtCut≤/>) backwards. Alternatively, if $\mathcal{H}$ contains a non-basic r-sequent $\Gamma \lesseqq \Delta$ such that the variable α appears in $\Gamma \cup \Delta$ and such that $\mathcal{H}$ does not contain $\emptyset \leq \nabla\alpha$, apply (ExtCut≤/>0). Proceed then in the same way as long as this is possible.

Consider next successively all the r-hypersequents at the leaves. Let $\mathcal{L}$ be one of them; and let $\mathcal{L}_b$ be the r-hypersequent arising from $\mathcal{L}$ by deleting all r-sequents which are not basic. Note that the validity of $\mathcal{L}_b$ can be translated to a statement about bounded totally ordered sets and thus be checked effectively. We proceed in dependence of the result:

(i) Let $\mathcal{L}_b$ be valid. Then we may prove $\mathcal{L}_b$ by means of the structural rules; see [15, Lemma 3.7]. Subsequently, $\mathcal{L}$ is proved from $\mathcal{L}_b$ by external weakening, (EW).

(ii) Let $\mathcal{L}_b$ be not valid. Then we may write $\mathcal{L} = \mathcal{L}_1 | \ldots | \mathcal{L}_k | \mathcal{R}$ such that (i) $\mathcal{L}_i$ and $\mathcal{L}_j$, $i \neq j$, do not have any atom in common; (ii) for any i and any distinct atoms α and β in $\mathcal{L}_i$, we have that $\nabla\alpha < \nabla\beta$ and $\nabla\beta < \nabla\alpha$ is in $\mathcal{L}_i$; similarly, for any variable α, we have that $\emptyset \leq \nabla\alpha$ is in $\mathcal{L}_i$; (iii) we may discard $\mathcal{R}$ without affecting the validity of $\mathcal{L}$.

We conclude that $\mathcal{L}$ is valid iff at least one of the $\mathcal{L}_i$ is valid. For a given i, however, $\mathcal{L}_i$ is valid if $\mathcal{L}'_i$ is valid in rHŁ [6], where $\mathcal{L}'_i$ arises from $\mathcal{L}_i$ by replacing all r-sequents $\emptyset \leq \nabla\alpha$ by $\emptyset \leq \alpha$ and then all remaining expressions $\nabla\alpha$ by 0.

Third Fact. The validity of atomic r-hypersequents of rHŁ can be checked effectively.

Indeed, this problem is equivalent to the problem if an associated system of linear inequalities is inconsistent, and can be solved by linear programming methods. This is a result of [6].

Adding more details would clearly go beyond the scope of this paper, but the material on which the present argumentation is based, can be found in [15].

6 Conclusion

It is a peculiar fact that for one of the best known fuzzy logics, the logic BL of continuous t-norms, the conception of an analytic proof system based on r-hypersequents causes serious problems. We considered two alternatives to Hilbert-style systems available at the moment – Bova's and Montagna's system RHBL, which adapts the notion of an r-hypersequent to a specific model of the propositions of BL, and our system rHML, which is a calculus for a conservative extension of BL. In both cases, the validity of a proposition w.r.t. BL can be effectively decided, the decision procedure being exponential. Further simplifications of the methods are desirable.

References

1. Avron, A.: A constructive analysis of RM, J. Symbolic Logic **52** (1987), 939–951
2. Baaz, M., Ciabattoni, A., Montagna, F.: Analytic calculi for monoidal t-norm based logic, Fundam. Inform. **59** (2004), 315–332.
3. Baaz, M., Fermüller, C. G.: Analytic calculi for projective logics, in: Murray, N. V. (ed.), "Automated reasoning with analytic tableaux and related methods", Proceedings of TABLEAUX 1999 (Saratoga Springs 1999), Springer-Verlag, Berlin 1999; 36–50.
4. Baaz, M., Zach, R.: Hypersequents and the proof theory of intuitionistic fuzzy logic, in: Clote, P. G., et al. (eds.), "Computer science logic", 14th conference of the EACSL (Fischbachau 2000), Springer-Verlag, Berlin 2000; 187–201.
5. Bova, S., Montagna, F.: Proof Search in Hajek's Basic Logic, available at arxiv.org/abs/cs/0605094.
6. Ciabattoni, A., Fermüller, C. G., Metcalfe, G.: Uniform rules and dialogue games for fuzzy logics, in: Baader, F., et al. (eds.), "Logic for programming, artificial intelligence, and reasoning", Proceedings of the 11th conference LPAR (Montevideo 2004), Springer-Verlag, Berlin 2005; 496–510.
7. Esteva, F., Godo, L., Hájek, P., Montagna, F.: Hoops and fuzzy logic, J. Log. Comput. **13** (2003), 531–555.
8. Hájek, P.: "Metamathematics of Fuzzy Logic", Kluwer Acad. Publ., Dordrecht 1998.
9. Hájek, P.: Some hedges for continuous t-norm logics, Neural Networks **12** (2002), 159–164.
10. Metcalfe, G., Olivetti, N., Gabbay, D.: Sequent and hypersequent calculi for Łukasiewicz and Abelian logics, *ACM Transactions on Computational Logic* **6** (2005), 578–613.

11. Metcalfe, G., Olivetti, N., Gabbay, D.: Analytic calculi for product logics, Arch. Math. Logic **43** (2004), 859–889.
12. Montagna, F.: Storage Operators and Multiplicative Quantifiers in Many-valued Logics, J. Logic Computation **14** (2004), 299 - 322.
13. Montagna, F.: Generating the variety of BL-algebras, Soft Comp. **9** (2005), 869–874.
14. Pottinger, G.: Uniform cut-free formulations of T, S4, S5 (abstract), J. Symb. Logic **48** (1983), 900.
15. Vetterlein, T.: Logics of ordinal multiples of standard t-norms, preprint; available at `ls1-www.cs.uni-dortmund.de/~vetterl/Artikel/LogicsOrdinalMultiples.pdf`.

MV-Algebras with the Cantor–Bernstein Property

Antonio Di Nola[1] and Mirko Navara[2]

[1] Department of Mathematics and Informatics, University of Salerno, via S. Allende, 84081 Baronissi, Salerno (Italy)
adinola@unisa.it

[2] Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Technická 2, 166 27 Praha (Czech Republic)
navara@cmp.felk.cvut.cz
http://cmp.felk.cvut.cz/∼navara

Abstract. We study the structures which satisfy a generalization of the Cantor–Bernstein theorem. This work is inspired by related results concerning quantum structures (orthomodular lattices). It has been proved that σ-complete MV-algebras satisfy a version of the Cantor–Bernstein theorem which assumes that the bounds of isomorphic intervals are boolean. This result has been extended to more general structures, e.g., effect algebras and pseudo-BCK-algebras.

There is another direction of research which has been paid less attention. We ask which algebras satisfy the Cantor–Bernstein theorem in the same form as for σ-complete boolean algebras (due to Sikorski and Tarski) without any additional assumption. In the case of orthomodular lattices, it has been proved that this class is rather large. E.g., every orthomodular lattice can be embedded as a subalgebra or expressed as an epimorphic image of a member of this class. On the other hand, also the complement of this class is large in the same sense. We study the analogous question for MV-algebras and we find out interesting examples of MV-algebras which possess or do not possess this property. This contributes to the mathematical foundations by showing the scope of validity of the Cantor–Bernstein theorem in its original form.

Keywords: Cantor–Bernstein theorem, MV-algebra, boolean element of an MV-algebra, partition of unity, direct product decomposition, σ-complete MV-algebra, multiplicative MV-algebra.

1 Introduction

Following the Cantor–Bernstein theorem for sets X, Y, if there are injective mappings $X \to Y$ and $Y \to X$, then there is a one-one mapping of X onto Y. The theorem was proved by Dedekind in 1887, conjectured by Cantor in 1895, and again proved by Bernstein in 1898 [23, p. 85]. Sikorski [29] and Tarski [30] proved the following generalization of the (Dedekind)-Cantor-Bernstein theorem: For any two σ-complete boolean algebras A and B and elements $a \in A$ and $b \in B$, if B is isomorphic to the interval $[0, a]_A$ and A is isomorphic to $[0, b]_B$, then A and B are isomorphic. To obtain the classical Cantor–Bernstein theorem, it suffices to assume that A and B are the powersets of X and Y, respectively, with the natural set-theoretic boolean operations.

Our work is inspired by generalizations of the Cantor–Bernstein theorem which appeared recently. It has been generalized to orthomodular lattices (mathematical structures obtained naturally as event structures of quantum systems) and to MV-algebras

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 861–868, 2007.
springerlink.com

(which form the basis for the semantics of the Łukasiewicz logic). Numerous further generalizations to more general structures (incl. effect algebras as a common generalization of orthomodular lattices and MV-algebras) followed—see the bibliography at the end of this paper. As these algebras do not satisfy the Cantor–Bernstein theorem for arbitrary σ-homomorphisms, additional assumptions are necessary.

Another line of research has been initiated in [5]. Here the question is which algebras satisfy the Cantor–Bernstein theorem in its form for boolean algebras, without any additional condition. This question has been studied in [5, 6] for orthomodular lattices. The main conclusion is that there are many such algebras.

Our aim in this paper is to find analogies in MV-algebras. We ask which MV-algebras satisfy the Cantor–Bernstein theorem without any additional conditions. We collect several observations showing that also this class may be interesting, although an exact analogy to the results obtained for orthomodular lattices in [5, 6] cannot be expected.

2 Inspiration: Orthomodular Lattices

Let us first briefly summarize the main results obtained in [5, 6] for orthomodular lattices (OMLs). They are presented only for inspiration, thus the reader not acquainted with orthomodular lattices may skip this section. For the notions of σ-homomorphism etc. in this context we refer to [1, 19, 26]. In the following sections, we shall look for analogies of these results for MV-algebras.

We shall only deal with σ-OMLs, i.e., with those OMLs which are closed under the formation of countable suprema and infima. We shall use the elementary fact that an interval in a σ-OML constitutes, with the operations naturally inherited from the host OML, a σ-OML (see [26]).

For an element a in an OML L, we say that a is *central* if $x = (x \wedge a) \vee (x \wedge \neg a)$ for all $x \in L$. Then L is isomorphic to the direct product of intervals $[0, a]_L \times [0, \neg a]_L$. The *center*, $C(L)$, of L is the set of all central elements. It is a boolean sub-σ-algebra of L. (The centers in OMLs corresponds to the *boolean skeletons* in MV-algebras.)

Definition 1. *[5] Let L be a σ-complete OML. Then L is said to satisfy the* Cantor–Bernstein property *if it has the following property: If, for some $a, b \in L$, $a \leq b$, the interval $[a, b]_L$ is isomorphic to the entire L, then L is isomorphic to each interval $[c, d]_L$ with $c \leq a$ and $d \geq b$ ($c, d \in L$). (In the original terminology of [5, 6], L is called* interval homogeneous.*)*

The relation of this notion to the Cantor–Bernstein theorem can be expressed as follows:

Proposition 1. *[5] A σ-OML M satisfies the Cantor–Bernstein property iff it satisfies the following condition: If there is an is a σ-OML L isomorphic to an interval $[0, b]_M$ and M is isomorphic to an interval $[0, a]_L$, then L is isomorphic to M.*

It is shown in [5] that there are OMLs which do not satisfy the Cantor–Bernstein property:

Theorem 1. *The class of σ-OMLs satisfying the Cantor–Bernstein property is not closed under the formation of products.*

The richness of the class of σ-OMLs satisfying the Cantor–Bernstein property is demonstrated by the following result:

Theorem 2. *Every σ-OML is a σ-epimorphic image of a σ-OML satisfying the Cantor–Bernstein property.*

On the other hand, the class of σ-OMLs *not* satisfying the Cantor–Bernstein property is also large:

Theorem 3. *Every σ-OML is a σ-epimorphic image of a σ-OML* not *satisfying the Cantor–Bernstein property.*

Similar conclusions have been obtained for subalgebras:

Theorem 4. *Every σ-OML is a sub-σ-OML of a σ-OML satisfying the Cantor–Bernstein property.*

Theorem 5. *Every σ-OML is a sub-σ-OML of a σ-OML* not *satisfying the Cantor–Bernstein property.*

3 MV-Algebras

We refer to [2], [3], and [25] for basic information on MV-algebras.

An MV-algebra A is *σ-complete* iff every sequence of elements of A, has supremum in A with respect to the underlying order of A.

Let us recall that an element a in an MV-algebra A is called *boolean* iff $a \oplus a = a$. We let $\mathbf{B}(A)$ denote the set of all boolean elements of A (the *boolean skeleton*). It is not hard to see that the operations of A make $\mathbf{B}(A)$ into a boolean algebra. If A is a σ-complete MV-algebra, then $\mathbf{B}(A)$ is a σ-complete boolean algebra, and the σ-infinitary operations of $\mathbf{B}(A)$ agree with the restrictions of the corresponding operations of A.

A *homomorphism* between two MV-algebras is a map that sends zero to zero, and preserves the operations $\oplus$ and $\neg$. A one-to-one surjective homomorphism is called an *isomorphism*.

For an MV-algebra $A = (A, 0, \oplus, \neg)$ and for any element $a \in A$, we define the *interval* $[0, a]_A$ by

$$[0, a]_A = \{x \in A \mid 0 \leq x \leq a\}.$$

It can be considered an MV-algebra $([0, a], \oplus_a, \neg_a, 0)$ if the operations $\neg_a \colon [0, a]_A \to [0, a]_A$ and $\oplus_a \colon [0, a]_A \times [0, a]_A \to [0, a]_A$ are defined by

$$\neg_a x = \neg\left(\neg a \oplus x\right)\ , \tag{1}$$

$$x \oplus_a y = (x \oplus y) \wedge a\,. \tag{2}$$

A generalization to intervals with non-zero lower bounds is easy (see also [7]).

If a is not a boolean element of A, then $[0, a]_A$ need not be a homomorphic image of A. Let $n \in \mathbb{N}$ and let $S_n = \{0, 1/n, 2/n, \ldots, 1\}$ be the MV-chain with $n + 1$ elements (with the Łukasiewicz operations). Then

$$[0,1/n]_{S_n} = \{0,1/n\}$$

is not a homomorphic image of S_n, because S_n has no other proper ideals than $\{0\}$.

On the other hand, the existence of a homomorphism of A onto $[0,a]_A$ need not imply that a is a boolean element of A. As a matter of fact, in the *standard MV-algebra* $[0,1]$ (with the Łukasiewicz operations), multiplication by $1/2$ is a homomorphism of $[0,1]$ onto the interval MV-algebra $[0,1/2]$, but the element $1/2$ is not boolean in $[0,1]$.

Corollary 1. *For each $a \in \mathbf{B}(A)$, the mapping $x \mapsto (x \wedge a, x \wedge \neg a)$ is an isomorphism of A onto the product MV-algebra $[0,a]_A \times [0,\neg a]_A$.*

4 MV-Algebraic Cantor–Bernstein Theorem

The following version of the Cantor-Bernstein theorem has been proved in [4]:

Theorem 6. *Let A and B be σ-complete MV-algebras. Let $a \in \mathbf{B}(A)$, $b \in \mathbf{B}(B)$, and assume α to be an isomorphism of A onto the interval algebra $[0,b]_B$, and β an isomorphism of B onto the interval algebra $[0,a]_A$. Then A and B are isomorphic.*

There is also a different form of Cantor-Bernstein theorem for MV-algebras proved by Jakubík [13].

Many generalizations have followed (see the bibliography). However, all of them require some additional conditions. Here the added assumptions are that the MV-algebra is σ-complete and that the bounds of intervals, a and b, are boolean. In the sequel we ask in which MV-algebras these conditions are unnecessary.

Definition 2. *For an MV-algebra M, we define the* Cantor–Bernstein property *as follows: If, for some $p,q \in M$, $p \le q$, the interval $[p,q]_M$ is isomorphic to the entire M, then M is isomorphic to each interval $[r,s]_M$ with $r \le p$ and $s \ge q$.*

Def. 2 can be rephrased in a slightly simplified form.

Proposition 2. *The Cantor–Bernstein property for MV-algebras is equivalent to the following property: If, for some $a \in M$, the interval $[0,a]_M$ is isomorphic to the entire M, then M is isomorphic to the interval $[0,b]_M$ for each $b \ge a$.*

It may happen that an MV-algebra does not contain a proper isomorphic subinterval; then it satisfies the Cantor–Bernstein property trivially. If M is an MV-algebra and there is an $a \in M \setminus \{1\}$ such that $M \cong [0,a]_M$, then there is an isomorphism $\varphi\colon M \to [0,a]_M$ and $\mathcal{A}(M)$ contains not only $a = \varphi(1)$, but also a strictly decreasing sequence $\varphi^2(1) = \varphi(a)$, $\varphi^3(1), \ldots$. Thus all finite MV-algebras satisfy the Cantor–Bernstein property. Nevertheless, there are also many infinite MV-algebras which satisfy the Cantor–Bernstein property for non-trivial reasons. In the sequel we summarize some observations from [7].

5 MV-Chains as Non-trivial Examples of MV-Algebras with Cantor–Bernstein Property

Here we present non-trivial examples of MV-chains with the Cantor–Bernstein property. We denote by $\mathbb{Q}_2$ the set of all dyadic rationals, i.e., rational numbers whose denominators are integer powers of 2.

Example 1. The standard MV-algebra $[0,1]$ and the MV-chain $\mathbb{Q} \cap [0,1]$ possess the Cantor–Bernstein property. However, the MV-chain $M = \mathbb{Q}_2 \cap [0,1]$ has not the Cantor–Bernstein property. Indeed, $[0, \frac{1}{2}]_M \cong M$, while $[0, \frac{3}{4}]_M$ is isomorphic to the ordinal sum of three copies of M, but not to M. Similar examples may be constructed for other bases than 2.

The latter example can be generalized:

Theorem 7. *[7] Let M be a subalgebra of the standard MV-algebra $[0,1]$. If M satisfies the Cantor–Bernstein property and contains a proper isomorphic interval, then it contains $\mathbb{Q} \cap [0,1]$.*

The condition in the latter theorem is necessary, but not sufficient.

The situation in non-Archimedean MV-algebras is different:

Example 2. The Chang MV-algebra $S_{\infty,1}$ is isomorphic to the ordinal sum of $\mathbb{N}_0$ and $-\mathbb{N}_0$, more exactly, to the interval $[(0,0),(1,0)]_L$ in the lexicographic product $L = \mathrm{Lex}(\mathbb{Z},\mathbb{Z})$. Using the Mundici functor Γ, $S_{\infty,1} \cong \Gamma(\mathrm{Lex}(\mathbb{Z},\mathbb{Z}),(1,0))$. It has the Cantor–Bernstein property. The same holds for all MV-algebras of the form $\Gamma(\mathrm{Lex}(\mathbb{Z},\mathbb{Z},\ldots,\mathbb{Z}),(1,0,\ldots))$.

The above example appears to be exceptional among MV-chains: The MV-algebra $S_{\infty,n} \cong \Gamma(\mathrm{Lex}(\mathbb{Z},\mathbb{Z}),(n,0))$ has the Cantor–Bernstein property iff $n = 1$. This result is generalized in the following theorem:

Theorem 8. *For every $n \in \mathbb{N}$, every MV-algebra of the form $\Gamma(\mathrm{Lex}(\mathbb{Z},\mathbb{Z},\ldots,\mathbb{Z}),(n,0,\ldots))$ has the Cantor–Bernstein property iff $n = 1$.*

6 Direct Products and the Cantor–Bernstein Property

The following example shows that there are many σ-complete MV-algebras which possess the Cantor–Bernstein property and admit proper isomorphic subintervals; moreover, the isomorphism may be non-trivial.

Example 3. Let $X_1,\ldots,X_j$ be a finite sequence of disjoint sets of strictly decreasing infinite cardinalities. Let M be the direct product

$$M = S_1^{X_1} \times \ldots \times S_k^{X_k} = \prod_{i=1}^{k} S_i^{X_i}. \tag{3}$$

We view the elements of M as functions $a\colon \bigcup_{i\le k} X_i \to [0,1]$. Suppose that the interval $[0,a]_M$ is isomorphic to M. This is possible only if $\operatorname{card}(a^{-1}(1)\cap X_1) = \operatorname{card} X_1$, because the factors $S_i^{X_i}$, $i>1$, admit intervals isomorphic to S_1^{ω}, but only for cardinalities $\omega < \operatorname{card} X_1$. Similarly, we obtain

$$\operatorname{card}(a^{-1}(1)\cap X_i) = \operatorname{card} X_i \quad \text{for all} \quad i=1,\dots,k. \tag{4}$$

Condition (4) is necessary and sufficient for $[0,a]_M \cong M$. If it is satisfied for some $a \in M$, than it holds for all $b \in [a,1]_M$, thus M satisfies the Cantor–Bernstein property. The isomorphism between $[0,a]_M$ and M may be non-trivial in the sense that we may admit $a(x) = \frac{i}{k} < 1$ for some $x \in X_k$; then the factor of $[0,a]_M$ corresponding to $x \in X_k$ is isomorphic to $S_i \not\cong S_k$, i.e., to some factor S_i corresponding to $y \in X_i$.

Ex. 3 can be further generalized so that we admit also finite cardinalities.

Example 4. Let $X_1,\dots,X_k$ be a sequence of disjoint sets and $j \le k$ such that $X_1,\dots,X_{j-1}$ have strictly decreasing infinite cardinalities and $X_j,\dots,X_k$ be finite. Let M be the direct product (3). Then condition (4) is necessary and sufficient for $[0,a]_M \cong M$. Thus M satisfies the Cantor–Bernstein property.

In this example, if $[0,a]_M \cong M$, then $a \restriction \bigcup_{i=j}^{k} X_i = 1$ and the isomorphism has to coincide with the identity on the factors $\prod_{i=j}^{k} S_i^{X_i}$.

In the previous examples, it was necessary to have strictly decreasing infinite cardinalities. We show that non-strict inequalities do not suffice:

Example 5. Let X_1, Y, Z be disjoint sets of the countable infinite cardinality and let $X_2 = Y \cup Z$. Let M be the direct product

$$M = S_1^{X_1} \times S_2^{Y} \times S_2^{Z} = S_1^{X_1} \times S_2^{X_2}\,.$$

We define $a \in M$ by

$$a(x) = \begin{cases} 1 & \text{if } x \in Z,\\ \frac{1}{2} & \text{if } x \in Y,\\ 0 & \text{if } x \in X_1. \end{cases}$$

Then $[0,a]_M \cong S_1^{Y} \times S_2^{Z} \cong M$. We define $b \in [a,1]_M$ by

$$b(x) = \begin{cases} 1 & \text{if } x \in X_2 = Y \cup Z,\\ 0 & \text{if } x \in X_1. \end{cases}$$

Then $[0,b]_M \cong S_2^{X_2} \not\cong M$, because $[0,b]_M$ has no factor isomorphic to S_1. Thus M does not satisfy the Cantor–Bernstein property.

Despite the above examples, there seems to be little chance to prove analogies of Ths. 2, 4 for MV-algebras. The reason is that the constructions in orthomodular lattices relied on complex constructions which do not have counterparts in MV-algebras. In particular, a proper class of non-isomorphic OMLs with given properties is constructed following the method of [20, 30, 31]. Using them, a product is constructed from non-isomorphic OMLs. In MV-algebras, the selection of such tools is rather limited [3].

Acknowledgments. This research was supported by the grant 201/07/1136 of the Czech Science Foundation.

References

1. L. Beran: *Orthomodular Lattices. Algebraic Approach.* Academia, Praha and D. Reidel, Dordrecht, 1984.
2. Cignoli R., Mundici D.: An invitation to Chang's MV-algebras. In: *Advances in Algebra and Model Theory*, M. Droste, R. Göbel, Eds., Gordon and Breach Publishing Group, Reading, UK, 1997, 171–197.
3. Cignoli R., D'Ottaviano I.M.L., Mundici D.: *Algebraic Foundations of Many-valued Reasoning.* Trends in Logic, Volume 7. Kluwer Academic Publishers, Dordrecht, 1999.
4. De Simone, A., Mundici, D., Navara, M.: A Cantor–Bernstein theorem for σ-complete MV-algebras. *Czechoslovak Math. J.* **53**(**128**) (2003), No. 2, 437–447.
5. De Simone, A., Navara, M., Pták, P.: On interval homogeneous orthomodular lattices. *Comment. Math. Univ. Carolin.* **42** (2001), No. 1, 23–30.
6. De Simone, A., Navara, M.: On the permanence properties of interval homogeneous orthomodular lattices. *Math. Slovaca* **54** (2004), 13–21.
7. Di Nola, A., Navara, M.: Cantor–Bernstein property for MV-algebras, to appear.
8. Dvurečenskij, A.: Central elements and Cantor–Bernstein theorem for pseudo-effect algebras. J. Austral. Math. Soc. **74** (2003), 121–143.
9. Freytes, H.: An algebraic version of the Cantor–Bernstein–Schröder Theorem. *Czechoslovak J. Math.* **54** (2004), 609–621.
10. Goodearl, K.R.: *Partially Ordered Abelian Groups with Interpolation.* AMS, Providence, RI, 1986.
11. Grätzer, G.: *General Lattice Theory.* 2nd ed., J. Wiley, Basel, 1998.
12. Hanf W.: On some fundamental problems concerning isomorphism of boolean algebras. *Math. Scand.* **5** (1957) 205-217.
13. Jakubík J.: Cantor–Bernstein theorem for MV-algebras. *Czechoslovak Math. J.* **49(124)** (1999) 517–526.
14. Jakubík, J.: On orthogonally σ-complete lattice ordered groups. *Czechoslovak Math. J.* **52**(**127**) (2002), No. 4, 881–888.
15. J. Jakubík: A theorem of Cantor–Bernstein type for orthogonally σ-complete pseudo MV-algebras. *Tatra Mt. Math. Publ.* **22** (2002), 91–103.
16. Jakubík, J.: Convex mappings of archimedean MV-algebras. *Math. Slovaca*, to appear.
17. Jakubík, J.: Cantor–Bernstein theorem for lattices. Submitted.
18. Jenča, G.: A Cantor–Bernstein type theorem for effect algebras. *Algebra Univers.* **48** (2002), 399–411.
19. G. Kalmbach: *Orthomodular Lattices.* Academic Press, London, 1983.
20. M. Kallus, V. Trnková: Symmetries and retracts of quantum logics. *Int. J. Theor. Phys.* **26** (1987), 1–9.
21. Kinoshita S.: A solution to a problem of Sikorski. *Fund. Math.* **40** (1953) 39–41.
22. Kühr, J.: Cantor–Bernstein theorem for pseudo-BCK-algebras, preprint, 2006.
23. Levy A.: *Basic Set Theory.* Springer-Verlag, Berlin. Perspectives in Mathematical Logic, 1979.
24. J.D. Monk, R. Bonnet: *Handbook of Boolean Algebras I.* North Holland Elsevier Science Publisher B.V., 1989.
25. Mundici D.: Interpretation of AF C^*-algebras in Łukasiewicz sentential calculus. *J. Functional Analysis*, **65** (1986) 15–63.

26. P. Pták, S. Pulmannová: *Orthomodular Structures as Quantum Logics.* Kluwer, Dordrecht/Boston/London, 1991.
27. Riečan, B., Mundici, D.: Probability on MV-algebras. In: E. Pap (ed.), *Handbook of Measure Theory*, North-Holland, Amsterdam, 2001.
28. Sikorski R.: *Boolean Algebras.* Springer-Verlag, Ergebnisse Math. Grenzgeb., Berlin, 1960.
29. Sikorski R.: A generalization of a theorem of Banach and Cantor–Bernstein. *Colloquium Math.* **1** (1948) 140–144 and p. 242.
30. Tarski A.: *Cardinal Algebras.* Oxford University Press, New York, 1949.
31. V. Trnková: Automorphisms and symmetries of quantum logics. *Int. J. Theor. Physics* **28** (1989), 1195–1214.

On Łukasiewicz Logic with Truth Constants

Roberto Cignoli[1], Francesc Esteva[2], and Lluís Godo[2]

[1] Instituto Argentino de Matemática, IAM
Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET
Saavedra 15, piso 3, C1083ACA Buenos Aires, Argentine
cignoli@dm.uba.ar

[2] Institut d'Investigació en Intel·ligència Artificial, IIIA
Consejo Superior de Investigaciones Científicas, CSIC
Campus Universitat Autònoma de Barcelona s/n, 08193 Bellaterra, Spain
{esteva,godo}@iiia.csic.es

Abstract. Canonical completeness results for Ł($\mathcal{C}$), the expansion of Łukasiewicz logic Ł with a countable set of truth-constants $\mathcal{C}$, have been recently proved in [5] for the case when the algebra of truth constants $\mathcal{C}$ is a subalgebra of the rational interval $[0, 1] \cap \mathbb{Q}$. The case when $C \not\subseteq [0, 1] \cap \mathbb{Q}$ was left as an open problem. In this paper we solve positively this open problem by showing that Ł($\mathcal{C}$) is strongly canonical complete for finite theories for *any* countable subalgebra $\mathcal{C}$ of the standard Łukasiewicz chain $[0, 1]_{Ł}$.

Keywords: Łukasiewicz logic, truth-constants, canonical standard completeness.

1 Introduction

The study of Łukasiewicz infinite-valued logic when adding truth constants goes back to Pavelka [14] who built a propositional many-valued logical system which turned out to be equivalent to the expansion of Łukasiewicz logic by adding into the language a truth-constant $\overline{r}$ for each *real* $r \in [0, 1]$, together with a number of additional axioms. Although the resulting logic is not strongly complete with respect to the intended semantics defined by the Łukasiewicz t-norm, (like the original Łukasiewicz logic), Pavelka proved that his logic, denoted here PL, is complete in a weaker sense. Namely, he defined the *truth degree* of a formula φ in a theory T as $\| \varphi \|_T = \inf\{e(\varphi) \mid e$ is a PL-evaluation model of $T\}$, and the *provability degree* of φ in T as $| \varphi |_T = \sup\{r \mid T \vdash_{\mathrm{PL}} \overline{r} \to \varphi\}$ and proved that these two degrees coincide. This kind of completeness is usually known as Pavelka-style completeness, and strongly relies on the continuity of Łukasiewicz truth functions (see also [7]). Novák extended Pavelka's approach to Łukasiewicz first order logic [10, 11]. The approach has been fully developed in the frame of the so-called *fuzzy logic with evaluated syntax* by Novák *et al.* [13].

Hájek [8] showed that Pavelka's logic PL could be simplified while keeping Pavelka-style completeness results. Indeed he showed that it is enough to extend the language only by a countable number of truth-constants, one for each *rational* in $[0, 1]$, and by adding to the logic the two following additional axiom schemata, called book-keeping axioms:

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 869–875, 2007.
springerlink.com

$$\overline{r}\&\overline{s} \leftrightarrow \overline{r *_{Ł} s}$$
$$\overline{r} \to \overline{s} \leftrightarrow \overline{r \Rightarrow_{Ł} s}$$

for each $r, s \in [0,1] \cap \mathbb{Q}$, where $*_{Ł}$ and $\Rightarrow_{Ł}$ are the Łukasiewicz t-norm ($x *_{Ł} y = \max(0, x + y - 1)$ and its residuum ($x \Rightarrow_{Ł} y = \min(1, 1 - x + y)$) respectively. He called this new system Rational Pavelka logic, RPL for short. Moreover, he proved that RPL is strongly complete for finite theories in the usual sense (not only Pavelka-style) by showing that rational truth-constants can be defined in suitable theories over Łukasiewicz logic. See also [13, Section 4.3.12] and the recent Novák's paper [12] for the possibility of disposing the irrational truth-constants in the framework of fuzzy logic with evaluated syntax.

In some recent papers [6, 15, 5], a new way to study expansions of logics of a left-continuous t-norm with truth-constants has been developed. If L_* is a logic of (left-continuous) t-norm $*$, and $\mathcal{C}$ is a countable L_*-subalgebra of the standard L_*-algebra $[0,1]_*$, then the logic $L_*(\mathcal{C})$ is defined as follows:

(i) the language of $L_*(\mathcal{C})$ is the one of L_* expanded with a new variable $\overline{r}$ for each $r \in C$

(ii) the axioms of $L_*(\mathcal{C})$ are those of L_* plus the bookeeping axioms

$$\overline{r}\&\overline{s} \leftrightarrow \overline{r * s}$$
$$\overline{r} \to \overline{s} \leftrightarrow \overline{r \Rightarrow_* s}$$

for each $r, s \in C$, where $\Rightarrow_*$ is the residuum of the norm $*$.

The key point is that $L_*(\mathcal{C})$ is an algebraizable logic and thus one can study the varieties of algebras associated to them and completeness results with respect to canonical $L_*(\mathcal{C})$-algebras, i.e. expanded L_*-algebras over [0, 1] where, for each $r \in C$, the truth-constant $\overline{r}$ is interpreted by its own value r. This type of completeness is called *canonical completeness*. Following this approach, the expansion of Gödel (and of some t-norm based logic related to the Nilpotent Minimum t-norm) with rational truth-constants, and the expansion of Product logic with countable sets of truth-constants have been respectively studied in [6] and in [15], while in [5] the authors study the general case of a logic of a continuous t-norm adding truth constants. In fact, in that paper the canonical completeness issue for $Ł(\mathcal{C})$, the expansion of Łukasiewicz logic Ł with a countable set of truth-constants $\mathcal{C}$, is solved for all cases when the algebra of truth constants $\mathcal{C}$ is a subalgebra of the rational interval $[0,1] \cap \mathbb{Q}$ (see [5, Prop. 24]). The case when C contains irrational values, i.e. when $C \not\subseteq [0,1] \cap \mathbb{Q}$, was left as an open problem[1].

In this paper we solve positively this open problem. Namely, we show that $Ł(\mathcal{C})$ is strongly canonical complete for finite theories for *any* countable subalgebra $\mathcal{C}$ of standard Łukasiewicz chain $[0,1]_{Ł}$. The way of proving this completeness result is by showing that any $Ł(\mathcal{C})$-chain is partially embeddable into the canonical $Ł(\mathcal{C})$-chain $[0,1]_{Ł(C)}$[2]. And to show this, we use two facts:

[1] Note that, unlike the fuzzy logic with evaluated syntax approach developed in [13], the notion of proof in the $Ł(\mathcal{C})$ logics is finitary, and hence formulas with irrational truth-constants cannot be replaced by infinitely-many formulas with rational truth-constants for completeness purposes.

[2] It is shown in [5, Proposition 11] that the partial embeddability property is not only sufficient but also necessary to prove finite strong completeness.

(i) the partial embeddability property of Product logic with truth constants proved in [15], and

(ii) a generalization of the observation in [2] showing that the standard MV-algebra $[0, 1]_Ł$ can be embedded in a segment of the standard Product algebra $[0, 1]_\Pi$.

The paper is organized as follows. In the next section we give some basic facts, definitions and known results and in the third section after some antecedents the proof of partial embeddability for Łukasiewicz logic with truth constants is given, and as consequence, the canonical finite strong completeness of the logics $Ł(\mathcal{C})$ is obtained, for any countable $\mathcal{C}$. We conclude with some final remarks about the extension of the completeness results in [5].

As a matter of notation, in this paper we shall use upper case calligraphic letters, e.g. $\mathcal{A}$, to denote algebras while the corresponding roman letters, e.g. A, will be used to denote their universes.

2 ℓ-Groups and Their Relation with MV and Product Algebras

Through this note by an ℓ-group we will understand a lattice-ordered *commutative* group. Given an ℓ-group $\mathcal{G} = (G, +, -, 0)$, G^- will denote its negative cone, i. e., $G^- = \{x \in G : x \leq 0\}$. It is well known (see, for example [3]) that important examples of product logic algebras (Π- algebras from now on) are the negative cones of ℓ-groups with a bottom added. Indeed, given a ℓ-group $\mathcal{G}$, and $\bot \notin G$, define on the set $G^- \cup \{\bot\}$ the binary operations $\odot$ and $\Rightarrow$ as follows:

$$x \odot y = \begin{cases} x + y & \text{if } x, y \in G^-, \\ \bot & \text{otherwise,} \end{cases}$$

and

$$x \Rightarrow y = \begin{cases} 0 \wedge (y - x) & \text{if } x, y \in G^-, \\ 0 & \text{if } x = \bot, \\ \bot & \text{if } x \in G^- \text{ and } y = \bot. \end{cases}$$

Then it is easy to check that $\langle G^- \cup \{\bot\}, \odot, \Rightarrow, \bot\rangle$ is a Π-algebra, that will be denoted by $\mathcal{P}(\mathcal{G})$. Notice that G^- with the restriction of the operations $\odot$ and $\rightarrow$ becomes a *cancellative hoop* as defined in [1].

Moreover if $0 < u \in G$, then it is easy to check that the segment $[-u, 0]$, equipped with the operations

$$x \otimes y = (x + y) \vee -u \quad \text{and} \quad x \rightarrow y = (y - x) \wedge 0 \tag{1}$$

becomes an MV-algebra that we will denote by $\mathcal{MV}(\mathcal{G}, -u)$. (cf. [1, p. 242]).

On the other hand it was observed in [2] that the standard MV-algebra $[0, 1]_Ł = ([0, 1], *_Ł, \Rightarrow_Ł, 0)$ can be embedded in a segment of the standard Π-algebra $[0, 1]_\Pi = ([0, 1], *_\Pi, \Rightarrow_\Pi, 0)$, where $*_\Pi$ is the usual product and $\Rightarrow_\Pi$ its residuum. The authors use this fact to give a faithful interpretation of Łukasiewicz logic into product logic. As a matter of fact, each MV-algebra can be embedded into a segment of a Π-algebra. Indeed, given a MV-algebra $\mathcal{A}$, there is an ℓ-group $\mathcal{G}$ and an order unit u of G such that $\mathcal{A}$ is isomorphic to the MV-algebra $\Gamma(\mathcal{G}, u)$ obtained by defining on the segment $[0, u] = \{x \in G : 0 \leq x \leq u\}$ the operations

$$x \otimes' y = (x + y - u) \vee 0 \text{ and } x \to' y = (u - x + y) \wedge u$$

(see [9, 4]). Now the mapping $x \mapsto x - u$ is an isomorphism of $\Gamma(\mathcal{G}, u)$ onto the MV-algebra $\mathcal{MV}(\mathcal{G}, -u)$ obtained by equipping the segment $[-u, 0]$ of G^- with the operations $\otimes$ and $\to$ as defined in (1). We shall denote by Γ^- de composition of Γ with mapping $x \mapsto x - u$. It is clear that Γ^-, like Γ, establishes a functorial equivalence between the categories of MV-algebras and ℓ-groups with a distinguished order unit. In what follows we shall use both Γ^- and Γ. In particular, the standard MV-algebra $[0, 1]_{Ł} = \Gamma(\mathbb{R}, 1)$ will be isomorphic to $\Gamma^-(\mathbb{R}, 1)$, i. e., to the MV-algebra defined on the segment $[-1, 0]$ equipped with the operations

$$x \otimes y = \max(x + y, -1), \text{ and } x \to y = \min(y - x, 0).$$

Notice that given an MV-algebra $\mathcal{A}$ we have an ℓ-group $\mathcal{G}$ and an order unit u such that $\mathcal{A} \cong \Gamma^-(\mathcal{G}, u)$ and this is in fact an interval of the product algebra $\mathcal{P}(\mathcal{G})$ defined by the negative cone of $\mathcal{G}$ adding a bottom element.

3 Adding Truth Constants

Given a subalgebra $\mathcal{C}$ of the standard MV-algebra $\Gamma(\mathbb{R}, 1)$, one can define (see for example [8, 5]) the logic Ł$(\mathcal{C})$ as the expansion of Łukasiewicz logic Ł by the set $\overline{C} = \{\overline{c} \mid c \in C\}$ of truth constants and the corresponding book-keeping axioms, i.e., for all $r, s \in C$,

$$\overline{r} \& \overline{s} \equiv \overline{\max(r + s - 1, 0)}$$
$$\overline{r} \to \overline{s} \equiv \overline{\min(1 - r + s, 1)}$$

The only inference rule is modus ponens. The notion of proof is as in Łukasiewicz logic. We will use the notation $\vdash_{Ł(C)}$ to refer to proofs in Ł$(\mathcal{C})$.

As in [5] we define the corresponding MV($\mathcal{C}$)-algebras as the structures $\mathcal{A} = (A, \wedge, \vee, \otimes, \Rightarrow, \{\overline{r}^{\mathcal{A}}\}_{r \in C})$, where $\mathbf{A} = (A, \otimes, \Rightarrow, \overline{0}^{\mathcal{A}})$ is an MV-algebra, and satisfying the following book-keeping equations:

$$\overline{r}^{\mathcal{A}} \otimes \overline{s}^{\mathcal{A}} = \overline{\max(r + s - 1, 0)}^{\mathcal{A}}$$
$$\overline{r}^{\mathcal{A}} \Rightarrow \overline{s}^{\mathcal{A}} = \overline{\min(1 - r + s, 1)}^{\mathcal{A}}$$

for any $r, s \in C$.

For each MV($\mathcal{C}$)-algebra $\mathcal{A}$, the set $C^{\mathcal{A}} := \{\overline{r}^{\mathcal{A}} : r \in C\}$ is in fact a subalgebra of $\mathcal{A}$, that we will denote $\mathcal{C}^{\mathcal{A}}$. Since $\mathcal{C}$ is a subalgebra of the standard MV-algebra, any homomorphism from $\mathcal{C}$ into a non-trivial[3] MV-algebra is injective [4, Theorem 3.5.1]. Therefore *the mapping defining the interpretation of the constants* $r \mapsto \overline{r}^{\mathcal{A}}$ *is a bijection from* $\mathcal{C}$ *onto* $C^{\mathcal{A}}$.

The *canonical MV($\mathcal{C}$)-algebra* $[0, 1]_{Ł(C)}$ is the MV($\mathcal{C}$)-algebra obtained expanding the standard MV-algebra $\Gamma(\mathbb{R}, 1)$ with the elements of the subalgebra $\mathcal{C}$ as constants.

Moreover given an MV($\mathcal{C}$)-algebra $\mathcal{A}$, an $\mathcal{A}$-evaluation e is just an $\mathbf{A}$-evaluation which is extended by $e(\overline{r}) = \overline{r}^{\mathcal{A}}$ for all $r \in C$. The notions of $\mathcal{A}$-model, $\mathcal{A}$-tautology and logical consequence $\models_{\mathcal{A}}$ are then as in the case of Łukasiewicz logic.

[3] An MV-algebra $\mathcal{A}$ is trivial when it has only one element, i. e., $0^{\mathcal{A}} = 1^{\mathcal{A}}$.

Finally, given a subalgebra $\mathcal{D}$ of the standard product algebra $[0,1]_\Pi$, one can define (see for example [5]) the logic $\Pi(\mathcal{D})$ as the expansion of the Product logic Π by the set $\overline{D} = \{\overline{c} \mid c \in D\}$ of truth constants and the corresponding book-keeping axioms. $\Pi(\mathcal{D})$-algebras are defined in an analogous way MV($\mathcal{C}$)-algebras are defined as expansions of MV-algebras. The canonical $\Pi(\mathcal{D})$-algebra defined over the standard product algebra interpreting the constant $\overline{r}$ by r itself will be denoted by $[0,1]_{\Pi(D)}$.

4 Partial Embedding

First of all recall a simplified version (suficient for what we need in this paper) of the partial embeddability property for product logic with truth constants given in [15, Theorem 5.3]. Let $\mathcal{C}$ be any countable subalgebra of $[0,1]_\Pi$ and let $\mathcal{A}$ a $\Pi(\mathcal{C})$-chain such that the subalgebra $C^A = \{\overline{r}^{\mathcal{A}} \mid r \in C\}$ is isomorphic[4] to $\mathcal{C}$. Then the partial embeddability property for $\Pi(\mathcal{C})$ reads as follows.

Proposition 1 ([15]). *Let $\mathcal{C}$ be a countable subalgebra of $[0,1]_\Pi$ and let $\mathcal{A} = \langle A, \wedge, \vee, \odot, \rightarrow, \{\overline{r}^{\mathcal{A}} : r \in C\}\rangle$ a $\Pi(\mathcal{C})$-chain such that $\mathcal{C}^{\mathcal{A}}$ is isomorphic to $\mathcal{C}$. For any finite subset $X \subseteq A$ there is a partial embedding from X to $[0,1]_{\Pi(C)}$, i.e. there is a mapping $f : X \to [0,1]$ such that:*

- *if $x, y, x \circ y \in X$, then $f(x \circ y) = f(x) \circ' f(y)$ for $\circ = \odot$ and $\circ' = *_\Pi$, or for $\circ = \rightarrow$ and $\circ' = \Rightarrow_\Pi$;*
- *for any $r \in C$ such that $\overline{r}^{\mathcal{A}} \in X$, $f(\overline{r}^{\mathcal{A}}) = r$.*

It is well known [4] that if an MV-algebra $\mathcal{A}$ is isomorphic to $\Gamma(\mathcal{G}, u)$ for some ℓ-group $\mathcal{G}$ with strong unit u, and $\mathcal{S}$ is a subalgebra of $\mathcal{A}$, then there is a (unique) sub-ℓ-group $\mathcal{E}$ of $\mathcal{G}$ such that $u \in \mathcal{E}$ and $\mathcal{S} \cong \Gamma(\mathcal{E}, u)$.

Returning to our problem, suppose $\mathcal{C}$ is a countable subalgebra of the standard MV-algebra $\Gamma(\mathbb{R}, 1) = [0,1]_{\text{Ł}}$. Consequently, the subalgebra $\mathcal{C}$ is isomorphic to $\Gamma(\mathcal{H}, 1)$ for a *unique* sub-ℓ-group $\mathcal{H}$ of $\mathbb{R}$ such that $1 \in H$. Moreover the product chain $\mathcal{P}(\mathcal{H})$ is a product subalgebra of $\mathcal{P}(\mathbb{R})$. Notice that, since $\mathbb{R}$ is an archimedean group, each element of the negative cone H^- can be written as $-n + r$, with $r \in C$ and $n \in \mathbb{N}$. The mapping

$$f : \mathcal{P}(\mathbb{R}) \to [0,1]_\Pi$$

defined by $f(x) = e^x$ for $x < 0$ and $f(\bot) = 0$ is indeed an isomorphism of product algebras, and therefore, $C^\star := \{e^{-n+r} : r \in C, n \in \mathbb{N}\} \cup \{0\}$ is a subalgebra of $[0,1]_\Pi$ isomorphic to $\mathcal{P}(\mathcal{H})$. Hence we can consider the expanded logic $\Pi(\mathcal{C}^\star)$ and its canonical $\Pi(\mathcal{C}^\star)$-algebra $[0,1]_{\Pi(C^\star)}$.

Therefore, we have seen that for each countable subalgebra $\mathcal{C}$ of the standard MV-algebra $[0,1]_{\text{Ł}}$, we can define a corresponding countable subalgebra $\mathcal{C}^\star$ of the standard Π-algebra $[0,1]_\Pi$. Hence, we can associate to the canonical MV($\mathcal{C}$)-chain the canonical $\Pi(\mathcal{C}^\star)$-chain.

[4] It is shown in [15] that for any $\Pi(\mathcal{C})$-chain $\mathcal{A}$, the subalgebra C^A is either isomorphic to the 2 element Boolean algebra (when $\overline{r}^{\mathcal{A}} = \overline{1}^{\mathcal{A}}$ for all $r > 0$) or to $\mathcal{C}$ (when $\overline{r}^{\mathcal{A}} \neq \overline{s}^{\mathcal{A}}$ for each $r \neq s$).

Now we are ready to prove the partial embeddability property for Łukasiewicz logic with truth constants.

Theorem 1. *For any countable subalgebra* $\mathcal{C}$ *of* $[0,1]_Ł$ *and any finite subset* X *of a L(C)-chain* $\mathcal{A} = \langle A, \wedge, \vee, \otimes, \rightarrow, \{\overline{r}^{\mathcal{A}} : r \in C\}\rangle$ *there is a partial embedding from* X *to* $[0,1]_{Ł(C)}$*, i.e. a mapping* $f : X \rightarrow [0,1]$ *such that:*

- *if* $x, y, x \circ y \in X$*, then* $f(x \circ y) = f(x) \circ' f(y)$ *for* $\circ = \otimes$ *and* $\circ' = *_Ł$*, or for* $\circ' = \rightarrow$ *and* $\circ' = \Rightarrow_Ł$*;*
- *for any* $r \in C$ *such that* $\overline{r}^{A} \in X$*,* $f(\overline{r}^{A}) = r$*.*

Proof: If $\mathcal{A}$ is a MV($\mathcal{C}$)-algebra, then there is an ℓ-group $\mathcal{G}$, a sub-ℓ-group $\mathcal{L}$ and an order unit u of $\mathcal{G}$ such that $\mathcal{A} \cong \Gamma(\mathcal{G}, u) \cong \Gamma^-(\mathcal{G}, u)$ and $C^{\mathcal{A}} \cong \Gamma(\mathcal{L}, u) \cong \Gamma^-(\mathcal{L}, u)$. Since $C^{\mathcal{A}}$ is isomorphic to a subalgebra of the standard MV-algebra, it follows that $\mathcal{L}$ is isomorphic to a sub-ℓ-group $\mathcal{H}$ of $\mathbb{R}$, and since u is an order unit, all the elements of the negative cone L^- can be written as $-nu + \overline{r}^{\mathcal{A}}$, for $n \in \mathbb{N}$ and $r \in C$. Thus we can consider the product algebra $\mathcal{P}(\mathcal{G})$ as a $\Pi(\mathcal{C}^\star)$-algebra, with $\overline{e^{-n+r}}^{\mathcal{P}(\mathcal{G})} = -nu + \overline{r}^{\mathcal{A}}$.

Let X be a finite subset of A. From now on we identify $\mathcal{A}$ and $\Gamma(\mathcal{G}, u)$ (hence taking $\overline{0}^{\mathcal{A}} = 0_G$ and $\overline{1}^{\mathcal{A}} = u$), and without loosing generality we can assume $u \in X$. Let $i : \Gamma(\mathcal{G}, u) \rightarrow \Gamma^-(\mathcal{G}, u)$ be defined by $i(x) = x - u$. By the partial embeddability property of Product logic with constants, the $\Pi(\mathcal{C}^\star)$-chain $\mathcal{P}(\mathcal{G})$ is partially embeddable into the canonical $[0,1]_{\Pi(C^\star)}$. Therefore, considering $i(X)$, as a subset of the $\Pi(\mathcal{C}^\star)$-chain $\mathcal{P}(\mathcal{G})$, there is a partial embedding from $i(X)$ into $[0,1]_{\Pi(C^\star)}$ such that $\overline{r}^{\mathcal{A}} - u = i(\overline{r}^{\mathcal{A}}) \mapsto e^{r-1}$, for each $\overline{r}^{\mathcal{A}} \in X$. In particular, $-u = i(\overline{0}^{\mathcal{A}}) \mapsto e^{-1}$ and $0_G = i(\overline{1}^{\mathcal{A}}) \mapsto e^0 = 1$, thus all the elements of $i(X)$ go to the segment $[e^{-1}, 1]$. Applying natural logarithms, we obtain a partial embedding of $i(X)$ into $\Gamma^-(\mathbb{R}, 1)$ such that $i(\overline{r}^{\mathcal{A}}) \mapsto r - 1$ for each $\overline{r}^{\mathcal{A}} \in X$. Thus, composing i with this embedding and finally with the isomorphism from $\Gamma^-(\mathbb{R}, 1)$ to $\Gamma(\mathbb{R}, 1)$ sending $r - 1 \mapsto r$, we obtain a partial embedding of $X \subset \mathcal{A}$ into the canonical Ł($\mathcal{C}$)-chain $[0,1]_{Ł(C)}$. This ends the proof. □

Following the notation introduced in [15, 5] where canonical completeness means that the logic is complete with respect to intended semantics, i.e. in our case with respect to the (canonical) algebra $[0,1]_{Ł(C)}$, the above partial embeddability property leads to the following canonical completeness result.

Corollary 1. *For any countable subalgebra* $\mathcal{C}$ *of* $[0,1]_Ł$*, the logic* $Ł(\mathcal{C})$ *is canonically finite strong complete.*

Proof: By [5, Theorem 26], Ł($\mathcal{C}$) is finite strong complete with respect to the class of Ł($\mathcal{C}$)-chains over the standard $[0,1]_Ł$. But this class contains only the canonical Ł($\mathcal{C}$)-chain since $\mathcal{C}$ is simple (it has no non-trivial filters). □

5 Final Remarks

The general completeness results in [5] for the expansions of the logics of a continuous t-norm adding truth-constants were restricted by the condition (called (C3) in [5]) that every truth-constant $r \in C$ belonging to a Łukasiewicz component of the t-norm generate a finite MV-chain, or equivalently, r should be a rational in the isomorphic copy of the component on $[0,1]$. Checking the proofs in [5], the result of this paper allows to remove such a restriction while keeping all the completeness results.

Acknowledgements. The authors acknowledge the support of a bilateral cooperation CSIC-CONICET project, ref. 2005AR0092. Esteva and Godo also acknowledge partial support of the Spanish project MULOG, TIN2004-07933-C03-01.

References

1. W. Blok, I. M. A. Ferreirim. On the structure of hoops, *Algebra Universalis* **43** (2000) 233–257.
2. M. Baaz, P. Hájek, J. Krajčk, D. Švejda. Embedding logics into product logic, *Studia Logica* **61** (1998) 35–47.
3. R. Cignoli and A. Torrens. An Algebraic Analysis of Product Logic, *Multiple Valued Logic* **5** (2000) 45–65.
4. R. Cignoli, I. M. D'Ottaviano, D. Mundici. *Algebraic foundations of many-valued reasoning*, Kluwer, Dordrecht-Boston-London, 2000.
5. F. Esteva, J. Gispert, L. Godo, C. Noguera. Adding truth constants to logics of continuous t-norm: axiomatization and completeness results. *Fuzzy Set and Systems* **158**, 597–618, 2007.
6. F. Esteva, L. Godo and C. Noguera. On rational weak nilpotent minimum logics, *Journal of Multiple-Valued Logic and Soft Computing* **12** (2006) 9–32.
7. H. Freytes. Pavelka-style completeness in expansions of Lukasiewicz logic, *Archive for Mathematical Logic*, in press.
8. P. Hájek. *Metamathematics of fuzzy logic*, Trends in Logic, vol. 4. Kluwer, 1998.
9. D. Mundici. Interpretation of AF C^*-algebras in Łukasiewicz sentential calculus, *Journal of Functional Analysis* **65** (1986) 15–63.
10. V. Novák. On the syntactico-semantical completeness of first-order fuzzy logic. Part I. Syntax and Semantics. *Kybernetika* **26** (1990) 47-66.
11. V. Novák. On the syntactico-semantical completeness of first-order fuzzy logic. Part II. Main Results. *Kybernetika* **26** (1990) 134-154.
12. V. Novák. Fuzzy logic with countable evaluated syntax revisited. *Fuzzy Sets and Systems*, in press.
13. V. Novák, I. Perfilieva, J. Močkoř. *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Pub., 1999.
14. J. Pavelka. On Fuzzy Logic I, II, III. *Zeitschrift fur Math. Logik und Grundlagen der Math.* **25** (1979) 45-52, 119-134, 447-464.
15. P. Savický, R. Cignoli, F. Esteva, L. Godo, C. Noguera. On product logic with truth-constants, *Journal of Logic and Computation* **16** (2006) 205–225.

EQ-Algebras in Progress

Vilém Novák

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
30. dubna 22, 701 03 Ostrava 1, Czech Republic
Vilem.Novak@osu.cz

Abstract. EQ-algebra is an algebra with three binary operations (meet, product, fuzzy equality) and a top element that has been introduced in [13] as an algebra of truth values for the fuzzy type theory (a higher-order fuzzy logic). Recall that till now, truth values in fuzzy type theory have been supposed to form either of IMTL, BL, MV or ŁΠ-algebra that are special residuated lattices. However, since fuzzy equality is a derived operation in residuated lattice, it is not so natural for fuzzy type theory as the EQ-algebra. In this paper, we continue the research of EQ-algebras. Namely, we have modified some axioms, show further properties of them and outline the filter theory.

Keywords: Residuated lattice, fuzzy equality, fuzzy logic, fuzzy type theory, higher-order fuzzy logic.

1 Introduction

Every many-valued logic is uniquely determined by the algebraic properties of the structure of its truth values. It is generally accepted that in fuzzy logic, it should be a residuated lattice, possibly fulfilling some additional properties. We may now distinguish various kinds of formal fuzzy logics. Most important among them are Łukasiewicz, BL-, MTL-, IMTL- and ŁΠ-fuzzy logics (see [14] for the discussion about proper fuzzy logic). Recall that all these logics have propositional as well as first-order predicate version and enjoy the completeness property.

A natural question raises, whether we can introduce also a higher-order fuzzy logic as a counterpart to classical higher-order logic (*type theory*; see [1]). The answer is positive and the *fuzzy type theory* (FTT) has indeed been introduced in [12]. The algebra of truth values considered there is the IMTL$_\Delta$-algebra, which is a residuated lattice fulfilling prelinearity and the double negation. The algebra is, moreover, endowed by additional operation of Baaz delta Δ (a special unary operation keeping 1 and sending all the other truth values to 0). In [11], other kinds of FTT have been introduced, namely those where the algebra of truth values is one of Łukasiewicz$_\Delta$, BL$_\Delta$ or ŁΠ-algebra. All these algebras have been extensively studied in [4, 6, 5, 7, 3] and elsewhere.

It is specific for FTT that its basic connective is a fuzzy equality. The reason is that it is developed as a generalization of the elegant classical formal system originated by A. Church and L. Henkin (see [1, 8] and the citations therein). In the algebras considered

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 876–884, 2007.
springerlink.com

in FTT so far, however, the basic connective is implication and equivalence (i.e. a fuzzy equality between truth values) is derived on the basis of it. Hence, a natural question arises, whether we can introduce an algebra of truth values specific for FTT. This has been done first time in [13] where the algebra called EQ-algebra has been introduced and its basic properties have been studied. The work on EQ-algebra is continued in this paper.

Let us also remark that a concept related to EQ-algebras is that of *equivalential algebras* introduced in [10]. It turns out, however, that the latter are very special algebras that are of little interest for fuzzy logics.

The axioms originally introduced in [13] have been slightly modified. It is also demonstrated that a very special case of EQ-algebras are *equivalential algebras* introduced already in 1975 by Kabzinński and Wroński in [10]. These algebras, however, are too narrow from the point of view of structures of truth values suitable for fuzzy logic.

2 Preliminaries

2.1 Residuated Lattices

A residuated lattice is the algebra of type (2, 2, 2, 2, 0, 0)

$$\mathcal{L} = \langle L, \vee, \wedge, \otimes, \Rightarrow, \mathbf{0}, \mathbf{1} \rangle \tag{1}$$

where $\langle L, \vee, \wedge, \mathbf{0}, \mathbf{1} \rangle$ is a lattice with the least element $\mathbf{0}$ and the greatest element $\mathbf{1}$. The operation $\otimes$ is multiplication such that $\langle L, \otimes, \mathbf{1} \rangle$ is a commutative monoid and $\Rightarrow$ is the residuation fulfilling the adjunction property

$$a \otimes b \leq c \quad \text{iff} \quad a \leq b \Rightarrow c, \qquad a, b, c \in L. \tag{2}$$

A BL-algebra is a residuated lattice where $(a \Rightarrow b) \vee (b \Rightarrow a) = \mathbf{1}$ (prelinearity) and $a \otimes (a \Rightarrow b) = a \wedge b$ (divisibility) holds for all $a, b \in L$. A BL-algebra is Gödel if $\otimes = \wedge$. As usual a^n denotes $\underbrace{a \otimes \cdots \otimes a}_{n-\text{times}}$.

A special operation is biresiduation:

$$a \Leftrightarrow b = (a \Rightarrow b) \wedge (b \Rightarrow a). \tag{3}$$

This operation is a natural interpretation of equivalence and it has also been used in the formulation of FTT.

The operation $\Leftrightarrow$ is reflexive, symmetric, and transitive in the following sense:

$$(a \Leftrightarrow b) \otimes (b \Leftrightarrow c) \leq a \Leftrightarrow c, \qquad a, b, c \in L.$$

The following holds in residuated lattices:

Lemma 1. *(a)* $a \Leftrightarrow b = \mathbf{1}$ *iff* $a = b$,
(b) $a \Leftrightarrow b \leq (a \Leftrightarrow c) \Leftrightarrow (b \Leftrightarrow c)$,
(c) $a \leq (a \Leftrightarrow b) \Leftrightarrow b$, *(Tax rule),*

(d) $(a \Leftrightarrow b) \otimes (c \Leftrightarrow d) \leq (a \Leftrightarrow c) \Leftrightarrow (b \Leftrightarrow d)$,
(e) $(a \Rightarrow b) \otimes (b \Leftrightarrow c) \leq a \Rightarrow c$,
(f) $((a \Leftrightarrow a) \Leftrightarrow b) = b$,
(g) $(a \wedge b) \Leftrightarrow a = (a \Rightarrow b)$,
(h) $a \wedge b \leq a \Leftrightarrow b$.

Lemma 2. *Let $\mathcal{L}$ be a residuated lattice. Then*

$$((a \wedge b) \Leftrightarrow c) \otimes (d \Leftrightarrow a) \leq (d \wedge b) \Leftrightarrow c.$$

The residuated lattice

$$\mathcal{L}_{Ł} = \langle [0,1], \wedge, \vee, \otimes, \Rightarrow, 0, 1 \rangle$$

where $a \otimes b = 0 \vee (a + b - 1)$ (Łukasiewicz conjunction) and $a \Rightarrow b = 1 \wedge (1 - a + b)$ (Łukasiewicz implication) is the *standard Łukasiewicz MV-algebra*. The biresiduation in it is the operation $a \Leftrightarrow b = 1 - |a - b|$.

2.2 Fuzzy Equality and Ordering

Let $\mathcal{L}$ be a residuated lattice and U a set (a universe). A fuzzy set $A \underset{\sim}{\subset} U$ is identified with a function $A : U \longrightarrow L$ called *membership function*.

The notions of fuzzy equality and ordering are generalizations of the classical notions when interpreting the classical properties of reflexivity, symmetry, transitivity and antireflexivity in algebraic terms as follows.

A fuzzy relation $E : U \times U \longrightarrow L$ is called a *fuzzy equality* on U if it is *reflexive*, i.e.

$$E(u,u) = \mathbf{1}, \qquad u \in U,$$

symmetric, i.e.

$$E(u,v) = E(v,u), \qquad u, v \in U$$

and *transitive*, i.e.

$$E(u,v) \otimes E(v,w) \leq E(u,w), \qquad u, v, w \in U.$$

A fuzzy relation $R : U \times U \longrightarrow L$ is a *fuzzy ordering*, if it is reflexive, transitive, and *antisymmetric*, i.e.

$$R(u,v) \otimes R(v,u) \leq E(u,v).$$

3 EQ-Algebras

3.1 Definition and Fundamental Properties

Definition 1. *EQ-algebra is the algebra*

$$\mathcal{L} = \langle L, \wedge, \otimes, \sim, \mathbf{1} \rangle \tag{4}$$

of type (2, 2, 2, 0) where for all $a, b, c \in L$:

(E1) $\langle L, \wedge, \mathbf{1}\rangle$ *is a commutative idempotent monoid (i.e.* $\wedge$*-semilattice with top element* $\mathbf{1}$*). We put* $a \leq b$ *iff* $a \wedge b = a$*, as usual.*
(E2) $\langle L, \otimes, \mathbf{1}\rangle$ *is a (commutative) monoid and* $\otimes$ *is isotone w.r.t.* $\leq$*.*
(E3) $a \sim a = \mathbf{1}$*,*
(E4) $((a \wedge b) \sim c) \otimes (d \sim a) \leq c \sim (d \wedge b)$*,*
(E5) $(a \wedge b \wedge c) \sim a \leq (a \wedge b) \sim a$*,*
(E6) $(a \wedge b) \sim a \leq (a \wedge b \wedge c) \sim (a \wedge c)$*,*
(E7) $a \wedge b \leq a \sim b$*.*

The operation "$\wedge$*" is called meet (infimum), "*$\otimes$*" is called product and "*$\sim$*" is a fuzzy equality.*

Axiom (E3) is axiom of reflexivity, (E4) is substitution axiom, (E5)–(E6) are monotonicity axioms and (E7) is axiom of boundedness. Note that the substitution axiom can be seen also as a special form of the extensionality (see [7] and elsewhere).

Clearly, $\leq$ is the classical partial order. We will also put

$$a \to b = (a \wedge b) \sim a, \tag{5}$$

and

$$\tilde{a} = a \sim \mathbf{1}, \tag{6}$$

$a, b \in L$. The derived operation (5) will be called implication. Note that axiom (E5), in fact, expresses isotonicity of implication w.r.t. the second variable and (E6) antitonicity of $\to$ w.r.t. the first variable. Thus, we may rewrite (E5),(E6) into

$$a \to (b \wedge c) \leq a \to b, \qquad \text{(E5')}$$
$$a \to b \leq (a \wedge c) \to b, \qquad \text{(E6')}$$

Lemma 3. *The following properties hold in EQ-algebras:*

(a) $a \sim b = b \sim a$*,* *(symmetry)*
(b) $(a \sim b) \otimes (b \sim c) \leq (a \sim c)$*,* *(transitivity)*
(c) $(a \to b) \otimes (b \to c) \leq a \to c$*.* *(transitivity of implication)*

In the sequel, we will freely use symmetry and transitivity of "$\sim$" without special reference to the above lemma.

Lemma 4. *Let* $a \leq b$*. Then*

(a) $a \to b = \mathbf{1}$*.*
(b) $a \sim b = b \to a$*.*
(c) $\tilde{a} \leq \tilde{b}$*.*
(d) $c \to a \leq c \to b$*,* $\quad b \to c \leq a \to c$*.*

Lemma 5. *The following holds for all* $a, b, c \in L$ *in every EQ-algebra.*

(a) $a \otimes b \leq a$*,* $a \otimes b \leq a \wedge b$*,* $c \otimes (a \wedge b) \leq (c \otimes a) \wedge (c \otimes b)$*.*
(b) $a \sim b \leq b \to a$*,* $\quad a \to a = \mathbf{1}$*.*

(c) $(a \to b) \otimes (b \to a) \leq (a \sim b)$.
(d) $a = b$ *implies* $a \sim b = \mathbf{1}$.
(e) $a \leq \tilde{a}$ *and* $\tilde{\mathbf{1}} = \mathbf{1}$.
(f) $\tilde{a} = \mathbf{1} \to a$ *and* $a \to \mathbf{1} = \mathbf{1}$.
(g) $a \otimes (a \sim b) \leq \tilde{b}$.
(h) $a \otimes b \leq \tilde{a} \otimes \tilde{b} \leq a \sim b$.
(i) $b \leq \tilde{b} \leq a \to b$.
(j) $((a \wedge b) \sim (c \wedge d)) \otimes (a \sim a') \otimes (b \sim b') \otimes (c \sim c') \otimes (d \sim d') \leq$
$\leq (a' \wedge b') \sim (c' \wedge d')$.

Lemma 6. *(a)* $(c \to (a \wedge b)) \otimes (a \sim d) \leq (c \to (d \wedge b))$.
(b) $(c \to a) \otimes (a \sim d) \leq (c \to d)$.
(c) $((a \wedge b) \to c) \otimes (a \sim d)^2 \leq ((d \wedge b) \to c)$.
(d) $(d \to c) \otimes (a \sim d)^2 \leq (a \to c)$.

Lemma 7. *If* $a \leq b \to c$ *then* $a \otimes b \leq \tilde{c}$.

Definition 2. *EQ-algebra is* separated[1] *if for all* $a, b \in L$ [7]

$$E \quad a \sim b = \mathbf{1} \quad \textit{implies} \quad a = b.$$

Thus, it follows from Lemma 5(d) that EQ-algebra is separated if

$$a = b \quad \text{iff} \quad a \sim b = \mathbf{1}, \qquad a, b \in L.$$

3.2 Some Examples

Lemma 8. *Let* $\mathcal{L} = \langle L, \vee, \wedge, \otimes, \Rightarrow, \mathbf{0}, \mathbf{1} \rangle$ *be a residuated lattice and* $f : L \longrightarrow L$ *be a* $\wedge$*-homomorphism such that* $a \leq f(a)$ *holds for all* $a \in L$. *Put*

$$a \sim b = f(a) \Leftrightarrow f(b)$$

where $\Leftrightarrow$ *is the biresiduum defined in (3). Then* $\langle L, \wedge, \otimes, \sim, \mathbf{1} \rangle$ *is an EQ-algebra.*

As a consequence of this lemma, we may present the following example.

Example 1. Let $\otimes$ be the Łukasiewicz conjunction and define $f : [0,1] \longrightarrow [0,1]$ by $f(x) = 1 \wedge (x + k)$ for some $k \in [0,1]$. Finally, put

$$x \sim y = 1 - |f(x) - f(y)|, \qquad x, y \in [0,1]. \tag{7}$$

Then

$$\mathcal{L} = \langle [0,1], \wedge, \otimes, \sim, 1 \rangle$$

is an EQ-algebra. We have $\tilde{x} = f(x)$.

Example 2. Let $\mathcal{L} = \langle L, \wedge, \vee, \otimes, \Rightarrow, \mathbf{0}, \mathbf{1} \rangle$ be a residuated lattice and put

$$a \Leftrightarrow b = (a \Rightarrow b) \wedge (b \Rightarrow a),$$
$$a \mathrel{\dot{\Leftrightarrow}} b = (a \Rightarrow b) \otimes (b \Rightarrow a).$$

Then both $\mathcal{L} = \langle L, \wedge, \otimes, \Leftrightarrow, \mathbf{1} \rangle$ as well as $\mathcal{L} = \langle L, \wedge, \otimes, \dot{\Leftrightarrow}, \mathbf{1} \rangle$ are separated EQ-algebras.

[1] In [12], the property (E2) of the fuzzy equality $\sim$ is called 1-faithfulness. The term "separated" has been earlier introduced by U. Höhle in [9].

3.3 Filters in EQ-Algebras

Definition 3. *Let $\mathcal{L} = \langle L, \wedge, \otimes, \sim, \mathbf{1} \rangle$ be an EQ-algebra. A filter is a subset $F \subset L$ such that the following is fulfilled:*

(i) If $a \in F$ and $a \leq b$ then $b \in L$.
(ii) If $a, b \in F$ then $a \otimes b \in F$.
(iii) If $a \rightarrow b \in F$ for some $a, b \in L$ then $\{a \otimes c \rightarrow b \otimes c \mid c \in L\} \subset F$.

One may see that this definition of a filter is the standard one taken from residuated lattices but extended by the third condition (iii) which is necessary to obtain the theorem below. Clearly, $\mathbf{1} \in F$ holds for each nonempty filter.

A filter F is *proper* if $F \neq L$. A proper filter F is *maximal* if there is no proper filter $G \subset L$ such that $F \subset G$.

As usual, we define equivalence relation on L, given a filter $F \subset L$ by

$$a \approx_F b \quad \text{iff} \quad a \sim b \in F. \tag{8}$$

Lemma 9. *The relation $\approx_F$ is the equivalence relation on L.*

Let $F \subset L$ be a filter. If $a \in L$ then the equivalence class w.r.t. $\approx_F$ will be denoted by $[a]$ (or, if necessary, by $[a]_F$). Furthermore, we will define a factor-algebra

$$\mathcal{L}|F = \langle L|F, \wedge, \otimes, \sim_F, \mathbf{1} \rangle \tag{9}$$

in the standard way. The support is $L|F = \{[a] \mid a \in L\}$. The operations are defined by $[a] \wedge [b] = [a \wedge b]$ and similarly for the other operations. Both $\wedge$ and $\otimes$ are denoted in the same way as in the original algebra $\mathcal{L}$. To avoid misunderstanding, however, we will denote

$$[a] \sim_F [b] = [a \sim b].$$

The top element is $[\mathbf{1}]$.

The ordering in $\mathcal{L}|F$ will be defined using the derived meet operation as follows:

$$[a] \leq [b] \quad \text{iff} \quad [a] \wedge [b] = [a] \quad \text{iff} \quad a \wedge b \approx_F a \quad \text{iff} \quad a \wedge b \sim a = a \rightarrow b \in F.$$

If $a \leq b$ then $a \rightarrow b = \mathbf{1}$ by Lemma 4. Since $\otimes$ is isotone, $a \otimes c \leq b \otimes c$ holds for any $c \in L$ and so, $a \otimes c \rightarrow b \otimes c = \mathbf{1}$. Let $F = \{\mathbf{1}\}$. Then $\mathbf{1} \leq \mathbf{1}$ and we conclude that F is a (proper) filter on L.

Lemma 10. *To every $a \neq \mathbf{1}$ there is a maximal filter $F \subset L$ such that $a \notin F$.*

Theorem 1. *The algebra (9) is a separated EQ-algebra and $f : a \mapsto [a]$ is a homomorphism of $\mathcal{L}$ onto $\mathcal{L}|F$.*

3.4 Another View on EQ-Algebras

Let $\mathcal{L} = \langle L, \wedge, \otimes, \sim, \mathbf{1} \rangle$ be an EQ-algebra. We can seen it as a set endowed with a classical partial order with classical equality and a top element, and a fuzzy equality together with a fuzzy ordering, i.e. a structure

$$\langle L, =, \leq, \sim, \lesssim, \mathbf{1}\rangle \tag{10}$$

where $\sim, \lesssim \in L^{L\times L}$.

Indeed, $\sim$ is clearly a fuzzy equality by (E3) and Lemma 3(a) and (b). Therefore, "=" is, by Lemma 5(d), a special case of "$\sim$".

Furthermore, let us denote $a \lesssim b = a \to b$[2]. From Lemma 4(a) we get: $a \leq a$ implies $(a \lesssim a) = \mathbf{1}$, i.e. $\lesssim$ is reflexive. Furthermore, by Lemma 5(c) we get

$$(a \lesssim b) \otimes (b \lesssim a) \leq a \sim b,$$

i.e. $\lesssim$ is antisymmetric w.r.t. $\sim$. Finally, by Lemma 3(b)

$$(a \lesssim b) \otimes (b \lesssim c) \leq (a \lesssim c), \tag{11}$$

i.e. $\lesssim$ is also transitive. Therefore, the fuzzy relation $\lesssim$ in (10) is a fuzzy ordering. Note that when writing (5) as

$$a \lesssim b = (a \wedge b) \sim a \tag{12}$$

and understanding $\sim$ as a fuzzy equality then (12) becomes just a generalization of the classical definition: $a \leq b$ iff $a \wedge b = a$. Do not forget, however, that $a \lesssim b$ in (12) means a *degree* in which a is smaller than or equal to b.

3.5 Special EQ-Algebras

Let $\mathcal{L}$ contain also the bottom element $\mathbf{0}$. Then we may put

$$\neg a = a \sim \mathbf{0}, \qquad a \in L. \tag{13}$$

Lemma 11. *(a)* $\neg\mathbf{1} = \tilde{\mathbf{0}}$, $\neg\mathbf{0} = \mathbf{1}$.
(b) $\mathbf{0} \to a = \mathbf{1}$, $\neg a = a \to \mathbf{0}$.
(c) If $a \leq b$ *then* $\neg b \leq \neg a$.
(d) $\neg\tilde{\mathbf{0}} = \neg\neg\mathbf{1} \leq \mathbf{1}$
(e) $a \otimes \neg a \leq \tilde{\mathbf{0}}$, $\tilde{a} \otimes \tilde{\mathbf{0}} \leq \neg a$, $\neg a \otimes \tilde{\mathbf{0}} \leq \tilde{a}$ *and* $a \otimes \mathbf{0} = \mathbf{0}$.
(f) $\neg a \otimes \neg b \leq a \to b$.

Definition 4. *(i) EQ-algebra is* spanned *if* 2

$$E\ \tilde{\mathbf{0}} = \mathbf{0}.$$

(ii) EQ-algebra is good *if for all* $a \in L$, i

$$E\ a \sim \mathbf{1} = a.$$

(iii) EQ-algebra is residuated *if for all* $a, b, c \in L$, ii

$$E\ (a \otimes b) \wedge c = a \otimes b \quad \textit{iff} \quad a \wedge ((b \wedge c) \sim b) = a.$$

If the EQ-algebra is good then it is spanned but not vice-versa. Clearly, (Eiii) can be written in a classical way as $a \otimes b \leq c$ iff $a \leq b \to c$.

In good EQ-algebras, many properties from Lemmas 5 and 11 become the standard properties known from the theory of residuated lattices. Let us now introduce the following induced operations:

$$a \leftrightarrow b = (a \to b) \wedge (b \to a), \tag{14}$$

$$a \mathrel{\dot{\leftrightarrow}} b = (a \to b) \otimes (b \to a). \tag{15}$$

[2] This notation is introduced only for transparency.

Lemma 12. *The following holds in every EQ-algebra $\mathcal{L}$:*

(a) $(a \wedge b) \leftrightarrow a = (a \wedge b) \mathrel{\hat{\leftrightarrow}} a = a \to b$.
(b) $a \mathrel{\hat{\leftrightarrow}} b \leq a \sim b \leq a \leftrightarrow b$.
(c) Both $\leftrightarrow$ as well as $\hat{\leftrightarrow}$ are fuzzy equalities fulfilling axioms (E3)–(E7).
(d) If $\mathcal{L}$ is linearly ordered then $a \leftrightarrow b = a \mathrel{\hat{\Leftrightarrow}} b = a \sim b$.

Lemma 13. *(a) In every good EQ-algebra* $a \leftrightarrow \mathbf{1} = a \mathrel{\hat{\leftrightarrow}} \mathbf{1} = a$.
(b) A good EQ-algebra $\mathcal{L}$ is residuated if

$$(a \otimes b) \wedge c = a \otimes b \quad \textit{implies} \quad a \wedge ((b \wedge c) \sim b) = a$$

for all $a, b, c \in L$.
(c) An EQ-algebra $\mathcal{L}$ is good iff

$$a \otimes (a \sim b) \leq b \tag{16}$$

for all $a, b \in L$.
(d) If a good EQ-algebra fulfils

$$(a \sim a') \leftrightarrow (b \sim b') \leq (a \sim b) \leftrightarrow (a' \sim b') \tag{17}$$

for all $a, b, a', b' \in L$ then $\sim = \leftrightarrow$.

Lemma 14. *Let $\mathcal{L}$ be a residuated EQ-algebra. Then it is good and separated.*

The EQ-algebra is *complete* if it is a complete $\wedge$-semilattice. This immediately implies (see [2]) that, since it contains a top element, a complete EQ-algebra is at the same time a complete lattice.

Lemma 15. *The following holds in every complete EQ-algebra:*

(a) $a \to \bigwedge_{i \in I} b_i \leq \bigwedge_{i \in I}(a \to b_i)$.
(b) $\bigvee_{i \in I}(a_i \to b) \leq (\bigwedge_{i \in I} a_i \to b)$.
(c) $\bigvee_{i \in I}((a_i \to b_i) \otimes (a_i \to \bigwedge_{i \in I} a_i) \otimes (b_i \to \bigwedge_{i \in I} b_i) \leq \bigwedge_{i \in I} a_i \to \bigwedge_{i \in I} b_i$.

Definition 5. *A* lattice EQ-algebra *(ℓEQ-algebra) is an EQ-algebra that is a lattice and, moreover, the following additional substitution axiom holds:* [iii]

E $((a' \vee b) \sim c) \otimes (a' \sim a) \leq ((a \vee b) \sim c)$.

Obviously, a complete EQ-algebra is a complete ℓEQ-algebra. A finite EQ-algebra is ℓEQ-algebra.

Theorem 2. *Let $\mathcal{L} = \langle L, \wedge, \vee, \otimes, \sim, \mathbf{1}, \mathbf{0} \rangle$ be a residuated ℓEQ-algebra with the bottom element. Then $\langle L, \wedge, \vee, \otimes, \to, \mathbf{1}, \mathbf{0} \rangle$ is a residuated lattice.*

Clearly, a complete residuated ℓEQ-algebra is a complete residuated lattice.

Lemma 16. *If $\mathcal{L}$ is an ℓEQ-algebra then*

$$a \to b = (a \vee b) \sim b. \tag{18}$$

It follows from this lemma that the fuzzy ordering $\lesssim$ in the ℓEQ-algebra can be defined alternatively, either by $a \lesssim b = (a \wedge b) \sim a$ or by $a \lesssim b = (a \vee b) \sim b$, just as classical ordering in the lattice.

Lemma 17. *Let $\mathcal{L}$ be an ℓEQ-algebra. Then*

(a) $((b \wedge c) \sim a) \otimes ((b \vee c) \to b) \leq c \sim a.$
(b) $((a \wedge (b \vee c)) \sim d) \otimes (a \sim b) \leq b \sim d.$

4 Conclusion

In this paper, we have introduced a special algebra called EQ-algebra which is aimed to become the algebra of truth values for fuzzy type theory. The main operation is a fuzzy equality $\sim$ which is natural interpretation of the main connective in FTT.

We have introduced the basic definition and several special kinds of EQ-algebras and studied their basic properties. Much algebraic study must still be done to be able to fulfil the above task — to develop FTT based on EQ-algebra.

Acknowledgments. The research was supported by project MSM 6198898701.

References

1. Andrews, P.: 2002, *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*. Dordrecht: Kluwer.
2. Blyth, T.: 2005, *Lattices and Ordered Algebraic Structures*. London: Springer.
3. Cignoli, R. L. O., I. M. L. D'ottaviano, and D. Mundici: 2000, *Algebraic Foundations of Many-valued Reasoning*. Dordrecht: Kluwer.
4. Cintula, P.: 2003, 'Advances in the ŁΠ and Ł$\Pi\frac{1}{2}$ logics'. *Archive of Math. Logic* **42**, 449–468.
5. Esteva, F. and L. Godo: 2001, 'Monoidal t-norm based logic: towards a logic for left-continuous t-norms'. *Fuzzy Sets and Systems* **124**, 271–288.
6. Esteva, F., L. Godo, and F. Montagna: 2001, 'The ŁΠ. and Ł$\Pi\frac{1}{2}$ logics: two complete fuzzy systems joining Łukasiewicz and product logics'. *Archive of Math. Logic* **40**, 39–67.
7. Hájek, P.: 1998, *Metamathematics of Fuzzy Logic*. Dordrecht: Kluwer.
8. Henkin, L.: 1950, 'Completeness in the theory of types'. *J. Symb. Logic* **15**, 81–91.
9. Höhle, U.: 1988, 'Quotients with respect to similarity relations'. *Fuzzy Sets and Systems* **27**, 31–34.
10. Kabzinński, J. and A. Wroński: 1975, 'On equivalential algebras'. In: *Proc. Int. Symp. Multiple-Valued Logic*. Indiana University, Bloomington, USA, pp. 419–428.
11. Novák, V.: 2005, 'Fuzzy Type Theory As Higher Order Fuzzy Logic'. In: *Proc. 6^{th} Int. Conference on Intelligent Technologies (InTech'05), Dec. 14-16, 2005*. Bangkok, Thailand, pp. 21–26.
12. Novák, V.: 2005, 'On Fuzzy Type Theory'. *Fuzzy Sets and Systems* **149**, 235–273.
13. Novák, V.: 2006a, 'EQ-algebras: primary concepts and properties'. In: *Proc. Czech-Japan Seminar, Ninth Meeting. Kitakyushu& Nagasaki, August 18–22, 2006*. pp. 219–223.
14. Novák, V.: 2006b, 'Which logic is the real fuzzy logic?'. *Fuzzy Sets and Systems* **157**, 635–641.

A Fuzzy Approach for the Sequencing of Didactic Resources in Educational Adaptive Hypermedia Systems

Mario García Valdez[1], Guillermo Licea Sandoval[2], and Oscar Castillo[1], and Arnulfo Alanis Garza[1]

[1] Division of Graduate Studies and Research,
Calzada Tecnolǿgico, S/N, Tijuana, Mexico
{mariog,ocastillo,alanis}@tectijuana.mx

[2] University of Baja California, Chemistry and Engineering School,
Calzada Tecnolǿgico, 14418, Tijuana, Mexico
glicea@uabc.mx

Abstract. Research in educational adaptive hypermedia systems has been concerned with the generation of personalized courses, in this work we focus on a task of these kind of systems: the semi-automatic sequencing of didactic resources. Sequencing defines the order in which topics (and didactic resources) in a course will be presented to learners, considering for this, their previous knowledge and particular objectives. This task is based on subjective information, for example the learner knowledge, preferences, learning style, and even assessment results are perceived differently depending the context. In this paper we define an architecture for sequencing of didactic materials using fuzzy attributes and rules.

Keywords: Adaptive Hypermedia, Fuzzy Systems.

1 Introduction

Computational Intelligence methods have been successfully applied in industrial process control systems [16]. Adaptive controllers are implemented using fuzzy logic, genetic algorithms and artificial neural networks working in synergy. As an example of this kind of controllers we can mention a neuro-fuzzy controller. This are hybrid systems in which the fuzzy controller deals with inexact data, and the neural network is used for learning from previous knowledge and preparing the controller for future events. In this paper we use a fuzzy controller to implement a hypermedia adaptive system, this system is used for presentation of educational content. It is important to clarify that the term *Adaptive* is used at different levels by the computational intelligence and adaptive hypermedia communities. In this work we use a non adaptive fuzzy system to implement an adaptive hypermedia system. The definition of *adaptive system* that we use in this work, is given by Garcia et al. [14] (an adaptive) *"system alters something in such a way that the result of the alteration corresponds to the most suitable solution*

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 885–892, 2007.
springerlink.com

in order to fulfill some specific needs". An adaptive system can be seen as a service provider, fulfilling the need of its clients. If the clients are persons, an adaptive system can try to alter his services to each individual personal needs. If we see this system as a controller, the feedback and operations of the system are as complex as human nature. Also learning (as in a neuro-fuzzy controller) is also difficult because each person is different and each person changes over time. Intelligent tutors are adaptive systems that have tackled the complex problem of helping learners in their process since the early years of AI. This research activity continues today, with additional technologies, and perspectives. In this paper, we propose a framework for adaptive educational hypermedia based on fuzzy rules. First a brief presentation of the methods and technologies used are presented in section 2, the overall architecture of our framework is presented in section 3.

2 Educational Adaptive Hypermedia Systems

The goal of Adaptive Hypermedia (AH) systems research is to enhance the functionality of hypermedia, tailoring to each users needs the navigation and presentation of resources. According to Brusilovsky [4], [2] these hypermedia systems have a user model, this model is used by the system to personalize (or adapt) the navigation and presentation of hypermedia. The user model contains the users goals, preferences and knowledge. One of the main application of AH systems are educational systems [2]. In these systems the goal is a learning objective, to achieve this goal the user follows a path of didactic activities and resources. The user model includes information about the user‘s learning process. For modeling the user, knowledge representation methods are used, mainly with support of uncertain, incomplete and probabilistic knowledge. Bayesian representation is used in [10], [12] , Fuzzy Logic or Neural Networks in [8],[9], [1]. Ontologies in [13],[7]. The adaptation mechanism includes a pedagogical strategy, defined by instructors. This is a multidisciplinary field and with a very active research community [4]. AH systems can support users in their navigation by limiting browsing space, suggesting most relevant links to follow, or providing adaptive comments to visible links.

2.1 Learning Objects and Metadata

One important ingredient of the majority of current educational systems is the use of reusable content, in the form of learning objects. Based on the object oriented paradigm learning objects are typically defined as components of instruction material which can be reused in multiple contexts. Instruction designers can create and maintain these components independent of each other and share them over the Internet [3]. Learning objects can range from complex simulations, to videos, images, quizzes, or simple text. Learning objects are the basic elements of current Learning Management Systems (LMS) and are the focus of standardization initiatives whose goal is defining open technical standards and their characteristic metadata [6]. The most important initiatives are the Advanced Distributed Learning Initiative (ADL-SCORM) [18], the Instructional Management System Project (IMS) [19], the Alliance of Remote Instructional Authoring Distribution Networks of Europe (ARIADNE)[20], and the IEEE Learning Technology Standards Committee [21]. The main objective of these open standards is to enable the interoperability of learning objects between different LMSs and Learning Objects Repositories (LORS). Basic metadata schema specifications for learning objects include:

Learning Object Metadata (LOM). Based on the Dublin Core metadata [22] this specification defines a set of meta-data elements that can be used to describe learning resources. Includes educational, relation, technical, and classification elements.

Content Aggregation Model (CAM). CAM defines a package for the aggregation, distribution, management, and deployment of learning objects. Defines an organization element which contains information about one particular, passive organization of the material, the organization for now is limited to a tree structure.

Learner Information (LI). A collection of information about a learner or a producer of learning content, the elements are based upon accessibilities; activities; affiliations; competencies; goals; identifications; interests; qualifications, certifications and licenses; relationship; security keys; and transcripts.

Sequence and Navigation (SN). SN defines a method for representing the intended behavior of an authored learning experience such that any Learning Technology system (LTS) can sequence discrete learning activities in a consistent way. Provides a rule based sequencing of behaviors.

These standards have been the basis for various research projects in Eduacational Adaptive Hypermedia systems: [5],[13],[1],[17]. We refer to learning objects as didactic resources in this paper to avoid confusion. Educational Adaptive Hypermedia systems use learning objects in several ways:

- Adaptive Navigation According to a certain user model, the format or availability of links is modified, sometimes a link to a learning object is not visible until some goal has been reached or links to learning objects previously visited change to a different format. Is also possible to alter the path of navigation in accordance to the learning style or preferences of the user. Navigation patterns of other learners also can be used to define the navigation of similar learners.
- Adaptive Content Content can be customized considering the user model, adjusting the level of detail, style or changing the media of presentation. Words, paragraphs or even metaphors can adapt to the cultural background of the user. This type of adaptation also refers to the way a learning object can interact with the user, for instance an assessment object can change the questions considering the users knowledge or learning style.

3 Architecture for Educational Adaptive Hypermedia

The architecture is implemented using an object oriented paradigm, object are referred from fuzzy rules used by instructors to specify the path a learner can take. Objects are designed to be extensible, so the instructor can add new properties depending on the pedagogy he or she uses; new properties can be crisp or fuzzy.

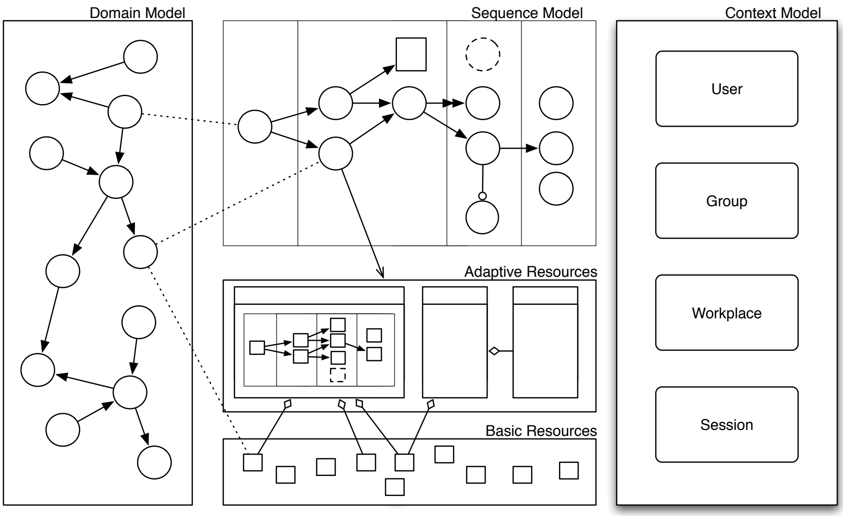


Fig. 1. Proposed architecture for eduacational adaptive hypermedia

The proposed architecture consists of the basic components of educational AH systems [4] with some distinctive features (Fig. 1):

– Domain Model. Implemented as a semantic network.
– Sequence Model. Defines an alternative path a user can take.
– Context Model. Includes the models of the User, Group, Workplace and Session.
– Adaptive Resources. Reusable resources of instruction with rules for adaptive presentation.
– Basic Resources. Basic resources, non aggregated.

3.1 User Model

The User Model is an important component of an adaptive hypermedia system. This model contains information about the user, that has been collected or inferred by the system. Brisulovsky [2] identifies five features (of the user) used by adaptive hypermedia systems: users goals, knowledge, background, hyperspace experience, and preferences. For educational systems psychological factors are also included for example: cognitive style, cognitive controls, and learning style.

There is however discussion about the effectiveness of this models and the adaptation process. Coffield, F. et al. [15] describes some of the main objections of learning style theories:

- The measure of the learning preferences are based on subjective judgments students make about themselves.
- The items in questioners are often ambiguous or not consider the different contexts or cultures where the items can have other meaning.
- Learning style is only one of the factors that influence learning, but is not necessarily the most significant.

Information about the learner is normally represented with uncertainty. In this paper we focus on the fuzziness of some of the students properties. This approach is also found in other projects: Issack and Senteni [1] proposed the use of a fuzzy or belief variable that is rationalizaed to take in to account in the adaptation process. Dattolo and Vicenzo [8] use fuzzy numbers to represent the users (stereotypes) different degrees of concern for the abstract objects included in the hypermedia. INSPIRE [11] also deals with uncertainty exploiting ideas of fuzzy logic and multicriteria decision making. Our approach resembles the model described by Stathacopoulou et al. [9] where fuzzy logic is used to describe the teachers subjective linguistic description of student behavior. The user model we propose is an object oriented representation of a user with fuzzy properties associated with a fuzzy linguistic variable. This linguistic variable has linguistic terms and corresponding membership functions. This variables are used in fuzzy propositions included in rules where instructors specify the sequence and selection of resources, this is described in section 3.4.

3.2 Context Model

When the user is interacting with the system, there are other factors Instructors can take into account. The user could be learning with other students, having a group knowledge. The Workplace model can include information about the physical or virtual place where the user is learning, for example: hardware, temperature, and bandwidth. This model can also be used to implement an adaptive behavior at a different level.

3.3 Domain Module

The domain model is a semantic network, implemented as a RDF model [23]. In this model the knowledge that is to be learned by users is represented. RDF resources include: learning objectives, topics, concepts, people and activities. The model does not include information about how to learn this concepts or how to achieve the learning objectives. This is specified in the sequencing model. The model is used to relate the concepts or as a glossary. It relates the sequence model with the didactic resources.

3.4 Sequence Model

The goal of the sequence model is to define a lesson for a target group. The IMS [19] Simple Sequencing (SS) specification defines sequencing as a mean to represent information needed to sequence learning activities in a variety of ways, it is neutral with regard of pedagogy and instructional strategies. The main components defined by the specification are Learning Activities and an Activity Tree. A Learning Activity is a pedagogically neutral unit of instruction, knowledge, assessment, etc. Learning Activities can have sub-activities nested to an arbitrary deep level. Learning Activities are organized in a tree structure called the Activity Tree. Sequencing defines a traversal of the tree structure, the default traversal path can be modified by the association of sequencing rules created by a learner designer. Learning Activity nodes may have associated content resources. With the conditional rules, complex behavior may be achieved. The Sequence Model (SM) we propose is shown in Fig. 1. The SM is basically implemented as a directed graph inside a timeline, where nodes are concepts or learning activities and are represented as circles. Alternatively rectangular nodes can also represent didactic resources, for cases where only one version of a didactic resource is required and also to be compatible the SS standard. Also there are nodes not linked to the graph, but their order in the sequence is implicitly represented by their position in the timeline. Optional learning activities are distinguished by a dotted line. The timeline is represented as contiguous vertical boxes, each representing a frame. This frames are ordered and can represent time, units, chapters in a book, credits, sessions, etc. There are several types of links in the SS:

Dependency. Represented by a solid arrow sequence links define optional paths a learner could follow. The direction implies a dependency between activities. An instructor suggest that a Learning should be needed to better understand the next activity. This are fuzzy links, Dependency is defined as a fuzzy variable with the terms: `HIGH`, `MEDIUM` or `LOW`. A fuzzy inference system is used to determine the dependency suggested for particular cases. For example:

```
IF user.age IS YOUNG and user.learningstyle.thoerist IS HIGH
THEN link.dependency IS LOW;
IF user.age IS ADULT and user.learningstyle.thoerist = LOW
THEN link.dependency IS MEDIUM;
IF user.age IS ADULT and user.learningstyle.thoerist = HIGH
THEN link.dependency IS HIGH;
```

There are two optional links for learning activities that are not needed for reaching the learning objectives of the sequence:

Additional Links. Additional links are optional learning activities that provide additional resources for the students, this links are represented by a line ending with a circle.

Advanced Links. These links connect to advanced learning activities and are represented by double solid arrows.

External Links. These links are from learning activities to other modules.

Implementation Links. A learning activity can be implemented with many adaptive resources, in a SS diagram this links point to this resources.Thess links are represented with a simple arrow.

Semantic association. These links associate a learning activity with an entity of the Domain model.

Adaptive Resources. These are didactic resources, that have adaptive presentation, the same sequence model is used internally. There is a collection of adaptive resources in a repository, and the more appropriate for a student is selected. Adaptive resources are aggregations of basic resources.

4 Conclusions

In this paper we have defined an architecture for sequencing of didactic materials. A prototype system is currently in development.

References

1. Santally, M., Senteni A.: A Learning Object Approach to Personalized Web-based Instruction European Journal of Open, Distance and Learning (2005)
2. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction Vol. 6 (1995) 87-129
3. Wiley, D.: Connecting Learning Objects to Instructional Design Theory: A Definition, a Metaphor, and a Taxonomy In: The Instructional Use of Learning Objects, Bloomington: Association for Educational Communocations and Technology (200) 3-23
4. Brusilovsky, P.: Adaptive Hypermedia User Modeling and User-Adapted Interaction Vol. 11 (2001)
5. Atif, Y., Benlamri, R.,Berr J.: Learning Objects Based Framework for Self-Adaptive Learning Education and Information Technologies Vol. 8 (2004) 345-368
6. Redeker G: An Educational Taxonomy for Learning Objects IEEE International Conference on Advanced Learning Technologies (ICALT'03) (2003)
7. Denaux, R., Dimitrova, V., Aroyo, L.: Interactive Ontology-Based User Modeling for Personalized Learning Content Management
8. Dattolo, A., Vicenzo, L.: A Fuzzy approach to Adaptive Hypermedia Proceedings. 4th Int. Conf. on K.B. Intelligent Engineering Systems and Allied Technologies Vol. 2 (2000) 700-703
9. Stathacopoulou, R., Grigoriadou, M., Magoulas, G., Mitropoulo, D.: A Neuro-fuzzy Approach in Student Modeling In: (P. Brusilovsky et al.) Eds: LNAI 2702 Springer-Verlag Berlin Heidelberg (2003) 345-368
10. Henze, N.: Adaptive Hyperbooks: Adaptation for Pro ject-Based Learning Resources PhD Thesis U. Hannover (2000)
11. Papanikolau, K. and Grigoriadou, M. and Kornilakis, H. and Magoulas, G.: Personalizing the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE User Modeling and User-Adapted Interaction Vol. 13 (2001) 213-267
12. Castillo, G., Gama, J., Breda A.M.: Adaptive Bayes for a Student Modeling Prediction Task Based on Learning Styles (P. Brusilovsky et al.) Eds: LNAI 2702 Springer-Verlag Berlin Heidelberg (2003) 328-332

13. Fischer, S.: Course and Exercise Sequencing Using Metadata in Adaptive Hypermedia Learning SystemsACM:Journal of Educational Resources in Computing Vol. 1 (2001)
14. García-Barrios, V. M., Mödritscher, F., Gütl, C.: Personalisation versus Adaptation? A User-centred Model Approach and its Applications KNOW05 (2005) 120-127
15. Coffield, F. and Moseley, D. and Hall, E. and Ecclestone, K.: Learning styles and pedagogy in post-16 learning. A systematic and critical review London Learning and Skills Research Centre
16. Konar, A.: Computational Intelligence: Principles, Techniques and Applications Springer, Berlin (2004)
17. Knolmay, G. F.: Decision Support Models for Composing and Navigating through e-Learning Objects Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS03) (2003)
18. SCORM: Sharable Content Object Reference Model http://www.adlnet.org
19. IMS: Instructional Management System Project http://www.imsproject.org
20. ARIADNE: Alliance of Remote Instructional Authoring and Distribution Networks of Europe http://www.ariadne.unil.ch.
21. LOM: Draft Standard for Learning Object Metadata, IEEE 1484.12.1-2002, Jul-15-2002.
22. DCMI: Dublin Core Metadata Initiative http://www.dublincore.org
23. Resource Description Framework (RDF): W3C Recommendation, February 10, 2004, Dave Beckett, ed.

Author Index