Vito Di Gesù
Sankar Kumar Pal
Alfredo Petrosino (Eds.)

LNAI 5571

Fuzzy Logic and Applications

8th International Workshop, WILF 2009
Palermo, Italy, June 2009
Proceedings

Lecture Notes in Artificial Intelligence 5571

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Vito Di Gesù Sankar Kumar Pal
Alfredo Petrosino (Eds.)

Fuzzy Logic and Applications

8th International Workshop, WILF 2009
Palermo, Italy, June 9-12, 2009
Proceedings

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Vito Di Gesù
Università di Palermo
Dipartimento di Matematica e Applicazioni
Via Archirafi 34, 90123 Palermo, Italy
E-mail: vito.digesu@unipa.it

Sankar Kumar Pal
Indian Statistical Institute
Center for Soft Computing Research
Machine Intelligence Unit
203 Barrackpore Trunk Road, Kolkata 700 108, India
E-mail: sankar@isical.ac.in

Alfredo Petrosino
Università di Napoli "Parthenope"
Dipartimento di Scienze Applicate
Centro Direzionale, Isola C4, 80143 Napoli, Italy
E-mail: alfredo.petrosino@uniparthenope.it

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2.3, I.5, I.4, F.4, G.2, J.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-02281-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-02281-4 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12695805 06/3180 5 4 3 2 1 0

Preface

The 8th International Workshop on Fuzzy Logic and Applications (WILF 2009) held in Palermo (Italy), June 9-12, 2009, covered topics related to theoretical and experimental areas of fuzzy sets and systems with emphasis on different applications.

This event represents the continuation of an established tradition of biannual interdisciplinary meetings. The previous editions of WILF were held, with an increasing number of participants, in Naples (1995), Bari (1997), Genoa (1999), Milan (2001), Naples (2003), Crema (2005) and Camogli (2007). Each event focused on distinct main thematic areas of fuzzy logic and related applications.

WILF 2009 aimed to highlight connections and synergies of fuzzy sets theory with nonconventional computing (e.g., neural networks, evolutionary computation, support vector machines, molecular computing, quantum computing) and cognitive science, in order to reach a better understanding of both natural and artificial complex systems as well as computing systems, inspired by nature, which are able to solve complex problems. From this perspective one of the main goals of the WILF workshops is to bring together researchers and developers from both academia and high-tech companies.

WILF 2009 received more than 60 paper submissions from all over the world, including Algeria, Belgium, Benin, Brazil, Canada, China, France, Greece, India, Iran, Italy, Japan, Poland, Romania, Slovakia, Spain and the USA. A rigorous peer-review selection process was applied to ultimately select nearly 40 high-quality manuscripts to be published in this volume.

Moreover, the volume also includes presentations from three keynote speakers Etienne Kerre (Ghent University, Belgium), Sankar K. Pal (ISI, India) and Enric Trillas (ECSC, Spain).

The success of this conference is to be credited to the contributions of many people. Special thanks go to the Program Committee members for their commitment to the task of providing high-quality reviews, and to the local Organizing Committee.

We cannot conclude without expressing our deepest gratitude to Giosué Lo Bosco and Cesare Valenti from the University of Palermo, who contributed to the organization of the workshop with the same passion and rigor that they apply in their scientific research. Without their help such a great and very special event could not have been organized.

March 2009

Vito Di Gesù
Sankar Kumar Pal
Alfredo Petrosino

To Our Friend Vito Di Gesú

Professor Vito Di Gesú passed away on March 15, 2009 at the age of 63 after a courageous struggle with a terminal illness.

Vito Di Gesú was one of the leading researchers in image analysis and pattern recognition after he joined the computer science group at the University of Palermo, where he served as a Full Professor and founded the PReDiGe research group.

In 1994 he was honored with the IAPR Fellowship at the very first assignment and with the Mahalanobis prize. Former President of GIRPR (the IAPR Italian Society), Vito Di Gesú initiated from 1986 its IAPR-TC13 on Pattern Recognition in Astronomy and Astrophysics. He was responsible for several projects for the Italian Ministry of Scientific Research and the Italian Space Agency; moreover, he was the national coordinator and member of the governing boards of several European projects. He served as member of the governing board of CINI (Italian Interuniversity Consortium in Informatics) and, at Palermo University, as research director of the CINI research unit, taking responsibility for the consortium COMETA and for the Socrates/Erasmus program on computer science.

Vito Di Gesú had a wide impact on our fields of research and pattern recognition at large, through authoring a large number of quality papers and co-editing more than two dozen books, organizing a dozen conferences and participating in about 60 conferences as a member of different committees.

His career was marked by remarkable productivity. His research interests originated from computer architectures like pyramids for computer vision to symmetry theory, passing through relevant results about intelligent clustering. As for applications, on top of his maintained interest in astronomy and astrophysics, to which he was applying his research on data mining and information fusion, we can cite his contributions in remote sensing and biomedical engineering, e.g., discrete tomography and diagnosis from image analysis.

He contributed significantly to spreading results through the Computer Architecture and Machine Perception (CAMP) workshops, sponsored by IEEE, and International Conferences on Image Analysis and Processing (ICIAP), sponsored by IAPR, of which he was a constant supportive participant and organizer. His interests in new disciplines and cross-fertilization between various fields took their meaning in organizing a series of workshops on Human – Machine Perception (HMP), where his contribution was really decisive, up to the very successful Workshop on Data Analysis in Astronomy (ADA) in 2007 where cosmology got connected with bioinformatics.

His current research dealt mainly with the origin of the synergy between the fields of soft computing and artificial vision, not only doing research by himself

in fuzzy clustering or genetic algorithms, but also participating actively in the success of our WILF conference series, since his first chairing in Naples in 2003.

He had been associated with the Indian Statistical Institute (ISI) for the last 20 years. He had a magical power of establishing both personal and professional friendship along with scientific collaboration with scientists of different institutions spread all over the world. He first visited ISI in 1988 in connection with an international conference and since then visited at least ten times, including his recent participation in PReMI-05 and PReMI-07 as invited speaker. He was actively connected with the "Center for Soft Computing Research: A National Facility" at ISI since its inception in 2004.

Vito Di Gesú was basically a man of many dimensions. He touched every one of us in a different way, both personally and professionally. We very much regret that some of you will not have the occasion to ever meet him. He would have surely deeply touched every one of you!

Vito, although of poor health, organized WILF 2009 with the enthusiasm of always.

We hope that WILF 2009 will serve both as a reminder of our loss and an inspiration as we venture toward other frontiers.

His friendly demeanor and hearty laugh are difficult to forget.

Sankar Kumar Pal
Alfredo Petrosino

Organization

Conference Chairs

Vito Di Gesù	University of Palermo, Italy
Sankar Kumar Pal	Indian Statistical Institute, Kolkata, India
Alfredo Petrosino	University of Naples Parthenope, Italy

Program Committee

Sanghamitra Bandyopadhyay	Indian Statistical Institute, Kolkata, India
Jim Bezdek	University of West Florida, Pensacola, USA
Isabelle Bloch	ENST-CNRS, Paris, France
Piero Bonissone	GE CRD, Schenectady, USA
Gianpiero Cattaneo	University of Milan Bicocca, Italy
Mario Enea	University of Palermo, Italy
Ashish Ghosh	Indian Statistical Institute, Kolkata, India
Ugur Halici	Middle East Technical University, Ankara, Turkey
Katsuhiro Honda	Osaka Prefecture University, Japan
Janusz Kacprzyk	Polish Academy of Sciences, Warsaw, Poland
Etienne Kerre	University of Ghent, Belgium
Erich Peter Klement	Johannes Kepler University, Linz, Austria
Malay Kumar Kundu	Indian Statistical Institute, Kolkata, India
Sushmita Mitra	Indian Statistical Institute, Kolkata, India
Witold Pedrycz	University of Alberta, Canada
Elie Sánchez	University of Aix-Marseille, France
George Sergiadis	Aristotle University of Thessaloniki, Greece
Michio Sugeno	Doshisha University, Kyoto, Japan
Roberto Tagliaferri	University of Salerno, Italy
Domenico Tegolo	University of Palermo, Italy
Andrea Tettamanzi	University of Milan, Italy
Settimo Termini	University of Palermo and CNR, Italy
Ioannis Vlachos	Aristotle University of Thessaloniki, Greece
Ronald Yager	Iona College, New Rochelle, USA
Bertrand Zavidovique	University of Paris-Sud, France

Steering Committee

Andrea Bonarini	Politecnico of Milan, Italy
Vito Di Gesù	University of Palermo, Italy
Antonio Di Nola	University of Salerno, Italy

Francesco Masulli	University of Genova, Italy
Gabriella Pasi	University of Milan Bicocca, Italy
Alfredo Petrosino	University of Naples Parthenope, Italy

Scientific Secretaries

Giosuè Lo Bosco	University of Palermo, Italy
Filippo Millonzi	University of Palermo, Italy
Marco Elio Tabacchi	University of Palermo, Italy
Cesare Fabio Valenti	University of Palermo, Italy

Sponsoring Institutions

Università degli Studi di Palermo, Italy
Università degli Studi di Napoli "Parthenope", Italy
Centro Interdipartimentale di Tecnologie della Conoscenza - Palermo, Italy
Indian Statistical Institute, India
Gruppo Italiano Ricercatori in Pattern Recognition, Italy

Table of Contents

Advances in Theory of Fuzzy Sets

Advances in Intuitionistic Fuzzy Sets

Fuzzy Classification and Clustering

Fuzzy Image Processing and Analysis

Fuzzy Systems

Advanced Applications

Non Contradiction, Excluded Middle, and Fuzzy Sets*

Enric Trillas**

European Centre for Soft Computing, 33600 Mieres, Spain
enric.trillas@softcomputing.es

Abstract. By means of a syntactic concept of self-contradiction, the aristotelian principles of non-contradiction and excluded-middle are posed in some very simple algebraic structures. Once linked with an algebraic representation of the relation If/then, such framework allows to represent both principles, and to prove that there is always the smallest relation for which they do hold. Finally, in agreement with the semantics of the relation If/then, the principles are stated with fuzzy sets, and some progress is reached in this case for what concerns the verification of the two principles.

Keywords: Self-contradiction, Non-contradiction, Excluded-middle, De Morgan algebras, Algebras of fuzzy sets.

1 Introduction

For some thinkers, the failure of either the principle of non-contradiction, or that of the excluded-middle, means that the corresponding theoretical developments are based on 'trembling grounds'. But things are what they are, and the principles fail in many cases.

Much earlier than logic were formalized, Aristotle linguistically stated the principle of non contradiction in [1], by a statement that can be shortened as 'A and not A is impossible', and he also added that such a principle is not susceptible to demonstration.

The term 'impossible' is usually taken as a synonym of 'false', whose reference is of a semantic character, and as far as this author knows, and with the exception of [6], no purely syntactic approach has been done to that principle. Nevertheless, by interpreting 'impossible' as 'self-contradictory' the principle can be algebraically translated, and also 'proven' with no previous conditions on the involved operations and relation, that is, in a totally general framework arisen from 'A and not A, self-contradictory'. This paper is devoted to such a goal, and the ideas in it come from the references [3], and [6].

* This paper is partially supported by the Foundation for the Advancement of Soft Computing (Asturias, Spain), and CICYT (Spain) under grant TIN2008-06890-C02-01.

** In fond remembrance of my Ph.D. Adviser, the late Prof. F. d'A. Sales i Vallès (1914-2005), who introduced me in the wellhead of ideas from which this paper comes, the 1964 Bodiou's book [3].

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 1–11, 2009.

For what concerns the principle of excluded-middle, Aristotle was actually less clear about it, although it was furtherly stated by 'not (A and not A) is always', or by 'A or not A is always'. It is not without some doubt, that this principle is here understood as 'not (A or not A) is impossible', and represented through 'not (A or not A), is self-contradictory'.

In order to have an algebraic representation of the above principles, it is considered a set L whose elements $a, b, c, \ldots$ represent statements, endowed with three operations $\cdot, +, '$, representing the connectives and, or , and not, respectively. To represent the conditionals 'If A, then B' it will be taken a binary relation $\vDash$ in L, allowing to interpret self-contradiction ('If A, then not A'). In this way the two principles can be syntactically considered, and the particular cases of De Morgan and fuzzy set algebras studied. With all that, some 'progress' in the principles' verification is reached.

2 Basic Concepts

2.1

Let L be a non-empty set whose elements $a, b, c, \ldots$ do represent statements A,B,... Furthermore let

- $' : L \to L$, is a mapping translating not. A statement like 'not A' is represented by a'.
- $\cdot : L \times L \to L$, is an operation translating and. A statement like 'A and B' is represented by $a \cdot b$.
- $+ : L \times L \to L$, is an operation translating or. A statement like 'A or B' is represented by $a + b$.
- $\vDash \subset L \times L$ is a binary relation translating If, then. A statement like 'If A, then B' is represented by $a \vDash b$.

In what follows, triplets $(L, \cdot, ')$ and $(L, +, ')$, as well as quartets $(L, \cdot, +, ')$ will be considered.

Definition 1. *$a \in L$ is self-contradictory provided $a \vDash a'$*

This definition translates 'If A, then not A'.

Definition 2. *A triplet $(L, \cdot, ')$ verifies the non-contradiction principle with respect to $\vDash$, provided all elements of the form $a \cdot a'$ are self-contradictory, that is*

$$a \cdot a' \vDash (a \cdot a')',$$

for all $a \in L$ (see [6]).

This definition translates 'If A and not A, then not (A and not A)'.

Definition 3. *A triplet $(L, +, ')$ verifies the excluded-middle principle with respect to $\vDash$, provided all elements of the form $(a + a')'$ are self-contradictory, that is*

$$(a + a')' \vDash ((a + a')')',$$

for all $a \in L$ (see [6]).

This definition translates 'If not (A or not A), then not(not (A or not A))'.

2.2

Examples of quartets $(L, \cdot, +, ')$ are given by ortholattices and De Morgan algebras (see [2])

An ortholattice $(L, \cdot, +, '; 0, 1)$, is a lattice $(L, \cdot, +)$ with first (0) and last (1) elements, endowed with an orthocomplement, that is, with a mapping $' : L \to L$ verifying: 1) $0' = 1$; 2) $(a')' = a$, for all $a \in L$; 3) $a \cdot a' = 0$ for all $a \in L$; and 4) $a \cdot b = ((a' + b')'$, for all $a, b \in L$. Orthomodular lattices are those ortholattices where $a \leq b \Rightarrow b = a + a' \cdot b$. Distributive ortholattices are Boolean algebras.

A De Morgan algebra $(L, \cdot, +, '; 0, 1)$,is a distributive lattice $(L, \cdot, +)$ with first (0) and last (1) elements, endowed with a pseudo-complement, that is, with a mapping $' : L \to L$ verifying: 1) $0' = 1$; 2) $(a')' = a$, for all $a \in L$; and 3) $a \cdot b = ((a' + b')'$, for all $a, b \in L$.

3 The Verification of the Principles

Denote by R_{NC} the set of binary relations $\vDash \subset L \times L$ for which the triplet $(L, \cdot, ')$ verifies the non-contradiction principle, and by R_{EM} that for which the triplet $(L, +, ')$ verifies the excluded-middle principle.

Theorem 1. *For no triplet $(L, \cdot, ')$ the set R_{NC} is empty.*

Proof. Define the binary relation in L, $\vDash_{NC} = \{(a \cdot a', (a \cdot a')'); a \in L\}$. Obviously, $\vDash_{NC} \in R_{NC}$.

Theorem 2. *It is* $\vDash_{NC} = \bigcap_{\vDash \in R_{NC}} \vDash$.

Proof. The relation $\bigcap \vDash$ is the smallest in R_{NC}, and obviously $\vDash_{NC} \subset \vDash$, for all $\vDash \in R_{NC}$.

Theorem 3. *For no triplet $(L, +, ')$ the set R_{EM} is empty.*

Proof. Define the binary relation in L, $\vDash_{EM} = \{((a + a')', ((a + a')')'); a \in L\}$. Obviously, $\vDash_{EM} \in R_{EM}$.

Theorem 4. *It is* $\vDash_{EM} = \bigcap_{\vDash \in R_{EM}} \vDash$.

Proof. The relation $\bigcap \vDash$ is the smallest in R_{EM}, and obviously $\vDash_{EM} \subset \vDash$, for all $\vDash \in R_{EM}$.

Theorem 5. *For all quartet $(L, \cdot, +, ')$, at least it exits the relation $\vDash_{NC} \cup \vDash_{EM}$, for which the principles of non-contradiction and excluded-middle do hold.*

Proof. Obvious.

Remark 1. Provided the mapping $'$ (negation) is *strong*, that is, verifies $(a')' = a$, for all $a \in L$, the *principle of excluded-middle can be stated by*

$$(a + a')' \vDash a + a',$$

for all $a \in L$.

Remark 2. If $'$ is strong, the two duality formulas $a \cdot b = (a' + b')'$, and $a + b = (a' \cdot b')'$, are equivalent. From $a \cdot b = (a' + b')'$, for all a, b in L , follows $a' \cdot b' = ((a)' + (b')')' = (a + b)'$, and $(a' \cdot b')' = a + b$. From $a + b = (a' \cdot b')'$ follows $a' + b' = (a \cdot b)'$, and $(a' + b')' = a \cdot b$.

Remark 3. If $'$ is strong, $\cdot$ is commutative, and $+$ is dual of $\cdot$, that is, $(a')' = a$, $a \cdot b = b \cdot a$, $a + b = (a' \cdot b')'$, for all a, b in L, the principle of excluded-middle follows from that of non-contradiction: $(a + a')' = ((a' \cdot (a)')' = a' \cdot a = a \cdot a'$, and $((a + a')')' = (a \cdot a')'$. Hence,

$$a \cdot a' \vDash (a \cdot a')' \Rightarrow (a + a')' \vDash ((a + a')')',$$

for all $a \in L$.

If $'$ is strong, $+$ is commutative, and $\cdot$ is dual of $+$, the principle of non-contradiction follows from that of excluded-middle: $(a + a')' = (a' + a)' = a \cdot a'$, and $((a + a')')' = a + a' = a' + a = (a \cdot a')'$.

Notice that from the duality and the commutativity of $+$, it follows the commutativity of $\cdot$.

Remark 4. As a consequence of remarks 3, it is clear that, *provided $'$ is strong, under the commutativity of $\cdot$ and $+$, and the respective duality of both $\cdot$ or $+$, the principles of non-contradiction and excluded-middle are equivalent.* In this case, it is obvious that

$$\vDash_{NC} = \vDash_{EM} = \{(a \cdot a', a + a'); a \in L\}$$

Remark 5. It is worth to notice that, independently of the properties operations $', \cdot, +$ can have, there are at least the relations $\vDash_{NC}, \vDash_{EM}$, and $\vDash_{NC} \cup \vDash_{EM}$, allowing the verification of either one, or two Aristotelian principles.

Nevertheless, and although this shows the general validity of the principles under no conditions, *at each problem, the relation $\vDash$ allowing them should be directly related with the particular way of representing* 'If A then B', *namely with the meaning or use of the conditional statements* 'If A, then not A'. A relationship that either could or could not exist (see [6]).

4 The Cases of Ortholattices, De Morgan Algebras and Fuzzy Sets

In [8], the verification of the principles is studied in the case of the major systems of three-valued logic. It is also interesting to study the cases of ortholattices, De Morgan algebras (see section 2.2), and fuzzy sets.

4.1

Let $(L, \cdot, +, '; 0, 1)$ be an ortholattice [2]. As it is well known, the two principles are there usually interpreted by the two axioms: $a \cdot a' = 0$, and $a + a' = 1$, for

all $a \in L$. Hence, $\vDash_{NC} = \vDash_{EM} = \{0,1\} \subset \leq$, with the lattice's order $\leq$ defined by $a \leq b \Leftrightarrow a \cdot b = a \Leftrightarrow a + b = b$.

In addition, $\leq$ is the greatest relation in L allowing the two principles, since

$$a \cdot a' = 0 \leq 1 = (a \cdot a')', \text{ and } (a + a')' = 1' = 0 \leq 1 = a + a'.$$

for all $a \in L$. Of course, these results also hold in the particular cases of orthomodular and boolean lattices.

4.2

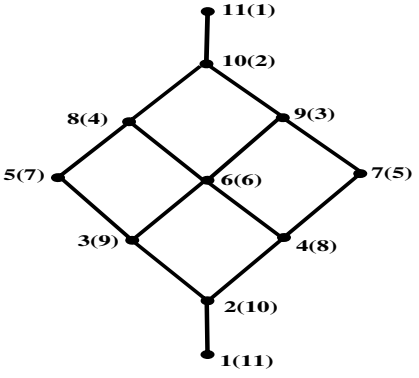
Let $(L, \cdot, +, '; 0, 1)$ be a De Morgan algebra [2]. Since it is neither $a \cdot a' = 0$, nor $a + a' = 1$ for all $a \in L$, the two principles cannot be stated as in ortholattices. Since $(a + a')' = a \cdot a'$ and $a + a' = (a \cdot a')'$, it is

$$\vDash_{NC} = \vDash_{EM} = \{(a \cdot a', a + a'); a \in L\} \subset \leq$$

with the lattice's order $\leq$ defined by $a \leq b \Leftrightarrow a \cdot b = a \Leftrightarrow a + b = b$. In addition,

- $a \cdot a' \leq a, a \cdot a' \leq a' \Rightarrow a' \leq (a \cdot a')', a \leq (a \cdot a')' \Rightarrow a \cdot a' \leq (a \cdot a')'$,
- $a \leq a + a', a' \leq a + a' \Rightarrow (a + a')' \leq a', (a + a')' \leq a \Rightarrow (a + a')' \leq a + a'$,

for all $a \in L$. Hence $\vDash_{NC} = \vDash_{EM} \subset \leq$, and $\leq$ is the greatest relation allowing the two principles.



In the finite De Morgan algebra in the figure, where the pseudo-complement a' of each a appears between parenthesis, it is easy to check that

$$\vDash_{NC} = \vDash_{EM} = \{(1,11), (2,10), (3,9), (4,8), (6,6), \} \subset \leq$$

4.3

In the De Morgan algebra $([0,1], \min, \max, 1 - \mathrm{id})$, is $a \cdot a' = \min(a, 1-a) \leq 0.5$, and $(a \cdot a')' = 1 - a \cdot a' \geq 1 - 0.5 = 0.5$, that is, $(a \cdot a', (a \cdot a')') \in [0, 0.5] \times [0.5, 1]$. Analogously, since $a + a' = \max(a, 1-a) \geq 0.5$, it also follows $((a + a')', ((a + a')')') \in [0, 0.5] \times [0.5, 1]$. Hence,

$$\vDash_{NC} = \vDash_{EM} \subset [0, 0.5] \times [0.5, 1] \subsetneq \leq$$

4.4

The set of fuzzy sets in a universe X, $[0,1]^X = \{\mu; \mu : X \to [0,1]\}$, is pointwise ordered by the partial order $\leq$ given by

$$\mu \leq \sigma \Leftrightarrow \mu(x) \leq \sigma(x),$$

for all $x \in X$. This order induces the pointwise identity $\mu = \sigma \Leftrightarrow \mu \leq \sigma$ and $\sigma \leq \mu \Leftrightarrow \mu(x) = \sigma(x)$ for all $x \in X$.

Among the constant fuzzy sets $\mu_r(x) = r$, for all $x \in X$, with $r \in [0,1]$, the ones μ_1 (the characteristic function of X), and μ_0 (the characteristic function of $\emptyset$), verify $\mu_0 \leq \mu \leq \mu_1$, for all $\mu \in [0,1]^X$.

A *minimal algebra of fuzzy sets* [5] is any quartet $([0,1]^X, \cdot, +, ')$, with $\cdot : [0,1]^X \times [0,1]^X \to [0,1]^X$, $+ : [0,1]^X \times [0,1]^X \to [0,1]^X$, and $' : [0,1]^X \to [0,1]^X$, verifying the four axioms:

1. $\mu \cdot \mu_1 = \mu_1 \cdot \mu = \mu$, $\mu + \mu_0 = \mu_0 + \mu = \mu$, for all $\mu \in [0,1]^X$
2. If $\mu \leq \sigma$, then $\mu \cdot \lambda \leq \sigma \cdot \lambda$, $\lambda \cdot \mu \leq \lambda \cdot \sigma$, for all $\lambda \in [0,1]^X$
3. If $\mu \leq \sigma$, then $\mu + \lambda \leq \sigma + \lambda$, $\lambda + \mu \leq \lambda + \sigma$, for all $\lambda \in [0,1]^X$
4. If $\mu \leq \sigma$, then $\sigma' \leq \mu'$.

From these axioms follows $\mu \cdot \mu_0 = \mu_0 \cdot \mu = \mu_0$, $\mu + \mu_1 = \mu_1 + \mu = \mu_1$, and $\mu \cdot \sigma \leq \min \circ (\mu \times \sigma) \leq \max \circ (\mu \times \sigma) \leq \mu + \sigma$, for all $\mu, \sigma \in [0,1]^X$. No minimal algebra of fuzzy sets is an ortholattice, but the one with $\mu \cdot \sigma = \min \circ (\mu \times \sigma)$, $\mu + \sigma = \max \circ (\mu \times \sigma)$, $\mu' = 1 - \mu$, is a De Morgan algebra. All these algebras verify the Kleene's law $\mu \cdot \mu' \leq \sigma + \sigma'$, for all $\mu, \sigma \in [0,1]^X$, and in general it is neither $\mu \cdot \mu' = \mu_0$, nor $\mu + \mu' = \mu_1$, for all $\mu, \sigma \in [0,1]^X$.

Theorem 6. *Minimal algebras of fuzzy sets verify the principles of non-contradiction and excluded-middle with respect to the ordering relation* $\leq$.

Proof. Since $\mu \cdot \mu' \leq \mu, \mu \cdot \mu' \leq \mu'$, it follows $\mu' \leq (\mu \cdot \mu')'$, and $\mu \cdot \mu' \leq (\mu \cdot \mu')'$. Since $\mu \leq \mu + \mu', \mu' \leq \mu + \mu'$, it follows $(\mu + \mu')'(\mu + \mu')' \leq \mu'$, and $(\mu + \mu')' \leq (\mu')'$. Hence, also $(\mu')' \leq ((\mu + \mu')')'$, and $(\mu + \mu')' \leq ((\mu + \mu')')'$.

When $\mu \cdot \sigma = T \circ (\mu \times \sigma), \mu + \sigma = S \circ (\mu \times \sigma), \mu' = N \circ \mu$, with T a continuous t-norm, S a continuous t-conorm, and N a strong negation, the algebras $([0,1]^X, \cdot, +, ') = ([0,1]^X, T, S, N)$ are called *standard algebras* of fuzzy sets.

5 Some Basic Ideas on Fuzzy Conditionals

Let us consider an standard algebra of fuzzy sets $([0,1]^X, T, S, N)$, where ordinal sums of t-norms or t-conorms will not be considered.

Sometimes, $\mu \leq \sigma$ is read μ is included in σ, since when μ, σ are in $\{0,1\}^X$ (crisp sets), $\mu \leq \sigma$ is equivalent to $\mu \subset \sigma$.

There are many forms of representing the conditional statements 'If μ, then σ' ($\mu \to \sigma$), but fuzzy logic always deals with functionally expressible representations of the form

$$\mu \to \sigma = J \circ (\mu \times \sigma),$$

with $J : [0,1]\times[0,1] \to [0,1]$, a T_0-conditional, that is, verifying $T_0(a, J(a,b) \leq b$, for all a, b in $[0,1]$, for some continuous t-norm T_0 (see [7]). Among these functions the most usual are the following:

- $J(a,b) = S(N(a),b)$, or S-implications (S for 'strong')
- $J(a,b) = J_T(a,b) = \sup\{z \in [0,1]; T(a,z) \leq b\}$, or R-implications (R for 'residuated')
- $J(a,b) = S(N(a), T(a,b))$, or Q-implications (Q for Quantum)
- $J(a,b) = T(a,b)$, or Mamdani-Larsen conditionals

Examples of such J are, respectively,

- $J(a,b) = \max(1-a,b)$, Kleene-Dienes implication, a T_0-conditional with $T_0 = W$
- $J(a,b) = 1$, if $a \leq b$; $J(a,b) = b/a$, otherwise, Goguen implication, with $T_0 = prod$
- $J(a,b) = \max(1-a, \min(a,b))$, Early-Zadeh implication, $T_0 = W$
- $J(a,b) = \min(a,b)$, Mamdani-conditional, $T_0 = \min$
- $J(a,b) = a \cdot b$, Larsen-conditional, $T_0 = \min$.

To check at least one T_0, observe that with the Kleene-Dienes implication is, $W(a, \max(1-a,b)) = \max(0, \max(1-a,b) + a - 1) = \max(0, a+b-1) = W(a,b) \leq b$, but neither $\min(a, \max(1-a,b)) \leq b$, nor $a.\max(1-a,b)) \leq b$, since, for example, $\min(0.7, \max(1-0.7, 0.1)) = 0.3 > 0.1$, and $0.7.\max(1-0.7, 0.1) = 0.21 > 0.1$.

6 The Case with Fuzzy Sets

Let us suppose an standard algebra $([0,1]^X, T, S, N)$ in which the conditional 'If μ, then σ' is functionally expressible by means of a T_0-conditional J, that is, by $(\mu \to \sigma)(x,y) = J(\mu(x), \sigma(y))$, for all x, y in X [7].

To relate the principles of non-contradiction and excluded-middle to the meaning of 'If μ, then σ' , that is, to the semantics of the corresponding T_0-conditional, define the relation

$$\mu \vDash^* \sigma \Leftrightarrow (\mu \to \sigma)(x,x) = 1 \Leftrightarrow J(\mu(x), \sigma(x)) = 1,$$

for all $x \in X$, *reducible to the numerical one*

$$a \vDash b \Leftrightarrow J(a,b) = 1,$$

for all a, b in $[0,1]$, by $\mu \vDash^* \sigma \Leftrightarrow \mu(x) \vDash \sigma(x)$, for all $x \in X$. The pairs $(a,b) \in \vDash$ are nothing else than the solutions of the two-variables equation $J(a,b) = 1$. In what follows only the cases considered in section 5 are taken into account.

6.1 The Case of S-Implications $J(a,b) = S(N(a),b)$

6.1.1 $S = \max$: $J(a,b) = \max(N(a),b) = 1 \Leftrightarrow a = 0$ or $b = 1$. Hence, $\vDash = \{(a,1); a \in [0,1]\} \cup \{(0,b); b \in [0,1]\}$, and $\vDash \subset \leq$, where $\leq$ is the linear order of $[0,1]$.

6.1.2 $\boldsymbol{S = prod^*_\varphi}$**:** $J(a,b) = \varphi^{-1}(\varphi(N(a)) + \varphi(b) - \varphi(N(a)).\varphi(b)) = 1 \Leftrightarrow \varphi(N(a))(1-\varphi(b)) = 1-\varphi(b) \Leftrightarrow b = 1$ or $(b < 1$ and $a = 0)$. Hence

$$\vDash = \{(a,1); a \in [0,1]\} \cup \{(0,b); b \in [0,1]\} \subset \leq .$$

6.1.3 $\boldsymbol{S = W^*_\varphi}$**:** $J(a,b) = \varphi^{-1}(\min(1, \varphi(N(a)) + \varphi(b))) = 1 \Leftrightarrow 1 \leq \varphi(N(a)) + \varphi(b) \Leftrightarrow \varphi^{-1}(1-\varphi(b)) \leq N(a)$, or $N_\varphi(b) \leq N(a)$.

6.1.3.1 If $N = N_\varphi, J(a,b) = 1 \Leftrightarrow N_\varphi(b) \leq N_\varphi(a) \Leftrightarrow a \leq b$.
Hence, $\vDash = \leq$. In this case, $J(a,b) = \varphi^{-1}(\min(1, 1-\varphi(a)+\varphi(b)))$,
is the R-implication J_T with $T = W_\varphi$.

6.1.3.2 If $N \neq N_\varphi$, then different solutions can arise. For example, with $N(a) = \sqrt{1-a^2}$, and $\varphi = \mathrm{id}$ $(N_\varphi = 1 - \mathrm{id})$, is $J(0.5, 0.4) = W^*(\sqrt{1-0.25}, 0.4)$ $= W^*(0.866, 0.4) = 1$, that is, $0.5 \vDash 0.4$, but not $0.5 \leq 0.4$. Hence, in this case, it is not $\vDash \subset \leq$.
With $N(a) = \frac{1-a}{1+a}$, and $\varphi = \mathrm{id}$, is $J(a,b) = W^*(\frac{1-a}{1+a}, b) = \min(1, \frac{1-a}{1+a} + b)$ $= 1 \Leftrightarrow (1-b)(1+a) \leq (1-a) \Leftrightarrow 2a \leq b(1-a) \Rightarrow 2a \leq 2b$, or $a \leq b$.
Hence, in this case, $\vDash \subset \leq$.

6.2 The Case of R-Implications $J_T(a,b) = \sup\{z \in [0,1]; T(a,z) \leq b\}$

From $J_T(a,b) = 1 \Leftrightarrow a \leq b$, follows $\vDash = \leq$.

6.3 The Case of Mamdani-Larsen Conditionals $J(a,b) = T(a,b)$

Since for either $T = \min, prod_\varphi$, and W_φ, is $T(a,b) = 1 \Leftrightarrow a = b = 1$, follows $\vDash = \{1,1\} \subset \leq$.

6.4 The Case of Q-Conditionals $J(a,b) = S(N(a), T(a,b))$

6.4.1 $\boldsymbol{S = \max}$ **:** $\max(N(a), T(a,b)) = 1 \Leftrightarrow a = 0$ or $a = b = 1$, hence $\vDash = \{(0,b); b \in [0,1]\} \cup \{1,1\} \subset \leq$.

6.4.2 $\boldsymbol{S = prod^*_\varphi}$ **:** $\varphi(N(a)) + \varphi(T(a,b)) - \varphi(N(a)).\varphi(T(a,b)) = 1 \Leftrightarrow \varphi(N(a)).(1-\varphi(T(a,b))) = 1 - \varphi(T(a,b)) \Leftrightarrow \begin{cases} T(a,b) = 1 \Leftrightarrow a = b = 1 \\ T(a,b) < 1, \text{ and } a = 0. \end{cases}$ Hence, $\vDash = \{1,1\} \cup \{(0,b); b \in [0,1]\} \subset \leq$.

6.4.3 $\boldsymbol{S = W^*_\varphi}$ **:** $\min(1, \varphi(N(a)) + \varphi(T(a,b))) = 1 \Leftrightarrow 1 \leq \varphi(N(a)) + \varphi(T(a,b)) \Leftrightarrow \varphi^{-1}(1-\varphi(T(a,b)) \leq N(a) \Leftrightarrow N_\varphi(T(a,b)) \leq N(a)$, whose study can be done like that in 6.1.3. For example, if $N = N_\varphi$, it results $a \leq T(a,b)$, that is $T(a,b) = a$ for all a, b in $[0,1]$, thus

- $T = \min$, gives $a \leq b$. Hence, $a \vDash b \Leftrightarrow a \leq b$, and $\vDash = \leq$.
- $T = prod_\varphi$, gives $\varphi(a).\varphi(b) = \varphi(a) \Leftrightarrow a = 0$, or $b = 1$, and $\vDash = \{(0,b); b \in [0,1]\} \cup \{(a,1); a \in [0,1]\} \subset \leq$.
- $T = W_\varphi$, gives $\max(0, \varphi(a) + \varphi(b) - 1) = \varphi(a) \Leftrightarrow a = 0$, or $b = 1$, and $\vDash = \{(0,b); b \in [0,1]\} \cup \{(a,1); a \in [0,1]\} \subset \leq$.

Remark 6. From 6.1 to 6.4 *it can be concluded the following*

- It is $\vDash = \leq$, in the case of all R-implications, since in the case 6.4.3, the Q-implications of the form $J(a,b) = W^*_\varphi(N_\varphi(a), \min(a,b)) = J_{W_\varphi}(a,b)$, are just R-implications.
- In all the other cases, it is $\vDash \subset \leq$, with the singularity of the Mamdani-Larsen conditionals, where $\vDash = \{1,1\} \subset \leq$.

6.5

Theorem 7. *For all triplet* (T,S,N) *it is* $T(a,N(a)) \leq N(T(a,N(a)))$ *and* $N(S(a,N(a))) \leq S(a,N(a))$ *for all* $a \in [0,1]$.

Proof. For all pairs (T,S) is $T \leq S$ (see [5]), an equality that also holds if $S = N \circ T \circ (N \circ N)$ -the N-dual t-conorm of T-, or if $T = N \circ S \circ (N \circ N)$ -the N-dual t-norm of S. Hence,

- $T(a,N(a)) \leq N(T(N(a),N(N(a)))) = N(T(N(a),a)) = N(T(a,N(a)))$, for all $a \in [0,1]$.
- $N(S(a,N(a))) \leq N(T(N(a),N(N(a)))) = T(N(a),a) = T(a,N(a)) \leq S(a,N(a))$, for all $a \in [0,1]$.

Corollary 1. *In all the cases in which* $\vDash = \leq$, $([0,1]^X, T, S, N)$ *does verify the principles of non-contradiction and excluded-middle.*

Proof. Follows immediately from theorem 7, and also from theorem 6,

- $\mu \cdot \mu' \vDash^* (\mu \cdot \mu')' \Leftrightarrow (\mu \cdot \mu')(x) \vDash (\mu \cdot \mu')'(x) \Leftrightarrow (\mu \cdot \mu')(x) \leq (\mu \cdot \mu')'(x) \Leftrightarrow T(\mu(x), N(\mu(x))) \leq N(T(\mu(x), N(\mu(x))))$
- $(\mu + \mu')' \vDash^* ((\mu + \mu')')' \Leftrightarrow (\mu + \mu')'(x) \vDash (\mu + \mu')(x) \Leftrightarrow (\mu + \mu')'(x) \leq (\mu + \mu')(x) \Leftrightarrow N(S(\mu(x), N(\mu(x)))) \leq S(\mu(x), N(\mu(x)))$,

for all $x \in X$

In the cases in which $\vDash \nsubseteq \leq$, *the principles do not hold.* To see it, consider the following relations (section 6.4) :

1. $\vDash = \{(0,b);(a,1);a,b \in [0,1]\}$
2. $\vDash = \{(0,b);(a,1);a \in [0,1]a,b \in [0,1)\}$
3. $\vDash = \{(0,b);b \in [0,1]\} \cup \{(1,1)\}$
4. $\vDash = \{(0,b);(a,1);a \in (0,1]a,b \in [0,1]\}$
5. $\vDash = \{(0,b);(a,1);a \in [0,1]a,b \in [0,1]\}$
6. $\vDash = \{(1,1)\}$.

Since $\mu \vDash^* \sigma \Leftrightarrow (\mu(x),\sigma(x)) \in \vDash$, for all $x \in X$, it respectively results

- 1.) $\mu \cdot \mu' \vDash^* (\mu \cdot \mu')' \Leftrightarrow (\mu \cdot \mu')(x) \vDash (\mu \cdot \mu')'(x), \forall x \in X \Leftrightarrow [(\mu \cdot \mu')(x) = 0 \ \& \ (\mu \cdot \mu')'(x) \in [0,1]]$ or $[(\mu \cdot \mu')(x) \in [0,1] \& (\mu \cdot \mu')'(x) = 1]$, for all $x \in X$. Hence, it should be $\mu \cdot \mu'(x) = 0$ for all $x \in [0,1]$, or $\mu \cdot \mu' = \mu_0$. The principle does not hold, and the only μ in $[0,1]^X$ verifying it are those that $\mu \cdot \mu' = \mu_0$ (for example, the crisp sets).

– 2.), 3.), 4.) and 5.). It follows the same result in (1.)
– 6.) $\mu \cdot \mu' \vDash^* (\mu \cdot \mu')' \Leftrightarrow (\mu \cdot \mu')(x) = 1 \;\&\; (\mu \cdot \mu')'(x) = 1, \forall x \in X \Leftrightarrow \mu \cdot \mu' = \mu_1 \;\&\; \mu \cdot \mu' = \mu_0$, that is absurd. In this case, not only the principle does not hold, but it is not verified by any $\mu \in [0,1]^X$.

6.6

For the principle of excluded-middle, an analogous process of reasoning can be done.

6.7

Let us remember [5] that the only standard algebras in which it is $\mu \cdot \mu' = \mu_0$ are those with $T = W_\varphi$ and $N \leq N_\varphi$, and the only for which it is $\mu + \mu' = \mu_1$ are those with $S = W^*_\psi, N \geq N_\psi$.

Hence, the only standard algebras where the above two formulas do hold are those with $T = W_\varphi$, $S = W^*_\psi$, and $N_\psi \leq N \leq N_\varphi$.

7 Conclusions

7.1

There is a simple way to always have $\vDash = \leq$. If suffices to change J, by

$$J^*(a,b) = \begin{cases} 1, & \text{if } a \leq b \\ J(a,b), & \text{otherwise} \end{cases}, \text{ with } J \text{ a } T_0\text{-conditional.}$$

With this change it results

– J^* is also a T_0-conditional: $T_0(a, J^*(a,b)) = \left\{ \begin{array}{ll} T_0(a,1) = a, & \text{if } a \leq b \\ T_0(a, J(a,b)) \leq b, & \text{otherwise} \end{array} \right\}$
$\leq b$, for all a, b in $[0,1]$.
– $a \vDash b \Leftrightarrow J^*(a,b) = 1 \Leftrightarrow a \leq b$.

The problem could lie in the corresponding change in the meaning of $\mu \to \sigma$, that could be non acceptable.

7.2

The relation $\vDash_{NC}$ (section 2.2) is the smallest one for which the principle of non-contradiction holds, and obviously $\vDash_{NC} \subset \leq$. *Can it be* $\vDash_{NC} = \leq$?

The answer is negative, since $\vDash_{NC} = \leq$ would mean that , once given T and N, for all pair $(x,y) \in \leq$ it should exist $a \in [0,1]$ such that $x = T(a, N(a))$ and $y = N(T(a, N(a)))$, equalities implying $y = N(x)$ for all $x \leq y$, that is never the case. For example, with $N = 1-\text{id}$, and $(0.6, 0.7)$ it is $0.7 \neq 1 - 0.6 = 0.4$.

It can be analogously proceeded with $\vDash_{EM}$.

7.3

To go a little further from $\vDash_{NC} \not\subseteq \leq$, and $\vDash_{EM} \not\subseteq \leq$, let us show that both relations are included in $[0,n] \times [n,1] \not\subseteq \leq$, with $N(n) = n \in (0,1)$, the fix point of the strong negation N [4]. From:

- $T(a, N(a)) \leq n$, for all $a \in [0,1]$, follows $n \leq NT(a, N(a)))$, and $\vDash_{NC} \subset [0, n] \times [n, 1]$
- $n \leq S(a, N(a))$, for all $a \in [0,1]$, follows $NT(a, N(a))) \leq n$, and $\vDash_{EM} \subset [0, n] \times [n, 1]$

For example, with $N(a) = \frac{1-a}{1+a}$, for which $n = \sqrt{2} - 1 = 0.4142$, the point $(0.5, 0.8)$ is not in $[0, \sqrt{2} - 1] \times [\sqrt{2} - 1, 1]$, and hence is neither in $\vDash_{NC}$, nor in $\vDash_{EM}$. A particular case is shown in section 4.3.

7.4

Let us finish by saying which is the 'progress' for the principles' verification that is obtained in this paper.

a. Independently of the semantics of 'If μ, then σ' ($\mu \rightarrow \sigma$), and as it was advanced in [6], by taking $\vDash = \leq$ all standard algebras of fuzzy sets (either dual or not) do verify the two principles.
b. By taking into account the semantics of $\mu \rightarrow \sigma$,
   b.1 The principles are verified in all the cases in which $\vDash_J = \leq$, and they fail if $\vDash_J \neq \leq$.
   b.2 If it is semantically acceptable to change J by J^* (7.1), since $\vDash_{J^*} = \leq$, all standard algebras do verify the principles

Acknowledgements. The author thanks Prof. Claudio Moraga and Dr. Eloy Renedo (Mieres) by his comments and kind help during the preparation of this paper.

References

1. Aristotle: Metaphysics. Book IV, 4. Translated by Ross, W.D., eBooks@Adelaide (2007)
2. Birkhoff, G.: Lattice Theory, 3rd edn. (7th. print) Amer. Math. Society, Colloq. Publs. (1993)
3. Bodiou, G.: Théorie dialectique des probabilités (englobant leurs calculs classique et quantique). Gauthier-Villars (1964)
4. Nguyen, H.T., Walker, E.A.: A First course in Fuzzy Logic. Chapman & Hall/Crc, Boca Raton (2000)
5. Pradera, A., Trillas, E., Guadarrama, S., Renedo, E.: On fuzzy set theories. In: Wang, P.P., Ruan, D., Kerre, E. (eds.) Fuzzy Logic. A spectrum of Theoretical and Practical Issues, vol. 215, pp. 15–47 (2007)
6. Trillas, E., Alsina, C., Pradera, A.: Searching for the roots of non-contradiction and excluded-middle. International Journal of General Systems 31, 499–513 (2002)
7. Trillas, E., Alsina, C., Pradera, A.: On mpt-implication functions for fuzzy logic. Rev. R. Acad. Cien. Serie A. Mat. 98, 259–271 (2004)
8. Trillas, E., Moraga, C., Renedo, E.: On Aristotle's NC and EM principles in three-valued logic. Technical report, European Centre for Soft Computing. In: IFSA 2009 (available upon request) (submitted, 2009)

Approximate Parallelism between Fuzzy Objects: Some Definitions

Maria Carolina Vanegas[1,2], Isabelle Bloch[1], Henri Maître[1], and Jordi Inglada[2]

[1] Institut Telecom, TELECOM ParisTech , CNRS-LTCI UMR 5141, Paris, France
{carolina.vanegas,isabelle.bloch,henri.maitre}@enst.fr
[2] CNES, Toulouse, France
jordi.inglada@cnes.fr

Abstract. Assessing the parallelism between objects is an important issue when considering man-made objects such as buildings, roads, etc. In this paper, we address this problem in the fuzzy set framework and define novel approximate parallelism notions, for fuzzy segments and non-linear objects or groups of objects. The proposed definitions are in agreement with the intuitive perception of this spatial relation, as illustrated on real objects from satellite images.

1 Introduction

We discuss the problem of defining parallelism between objects and fuzzy objects. This work is motivated by the importance of this spatial relation for describing human made objects such as buildings, roads, railways, observed in satellite images. Parallelism has been widely studied in the computer vision community in the perceptual organization domain, since it is an important feature of the grouping principles of the Gestalt theory [4].

Parallelism between linear segments was studied in several works, for example [7,6,5,8]. In [7] the parallelism is detected by assigning a significance value to determine that the detected parallelism has not been accidentally originated. A fuzzy approach is proposed in [6,5], leading to a measure of the degree of parallelism between two linear segments. The parallelism between curves was studied in [4,9], where it was treated as a shape matching problem.

The previous works focus on parallelism between crisp segments. We propose a definition to evaluate parallelism between fuzzy segments. Then we extend it to non-linear objects and groups of objects (crisp or fuzzy) taking into account the semantic meaning of the relation. The properties of our definitions are different from those desired in perceptual organization and are adapted to our purpose.

This paper is organized as follows. Section 2 contains considerations that should be taken into account when modeling the parallel spatial relation. A model for fuzzy segments is proposed in Sec. 3 and for non-linear objects in Sec. 4. Experimental results are shown in Sec. 5.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 12–19, 2009.

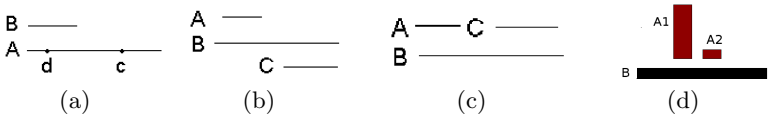
2 Considerations for Modeling Parallelism

Parallelism can be of interest in multiple situations, between objects or their boundaries, between lines, between groups of objects, etc.

For linear segments to be parallel, we expect a constant distance between them, or that they have the same normal vectors and the same orientation. Although classical parallelism in Euclidean geometry is a symmetric and transitive relation, these properties are subject to discussion when dealing with image segments of finite length. When segments have different extensions as in Fig. 1(a), where B can be the boundary of a car, and A the boundary of a road, the symmetry becomes questionable. The statement "B is parallel to A" can be considered as true, since from every point in B it is possible to see (in the normal direction) a point of A, and the normal vectors at both points are the same. On the contrary, the way we perceive "B parallel to A" will change depending on our position: from point d it is possible to see a point of B in the normal direction with the same normal vector, while this is not possible from point c. In both cases (symmetrical and non symmetrical ones) the transitivity is lost. For example, in Fig. 1(b) and 1(c) the statements "A is parallel to B" and "B is parallel to C" hold, but "A is not parallel to C" since it is not possible to see C from A in the normal direction to A. This example also illustrates the interest of considering the degree of satisfaction of the relation instead of a crisp answer (yes/no). Then the relation "B parallel to A" will have a higher degree than "A parallel to B" in Fig. 1(a).

Now, when considering objects, parallelism is often assessed visually for elongated objects, based on the portions of their boundaries that are facing each other. Figure 1(d) shows two configurations where the portions of the boundary of object $A1$ and of $A2$ that face B have the same length. Do we want to assign the same degree of satisfaction to the parallel relations in both situations? Therefore, the dimensions of the objects can also influence the way we perceive parallelism.

The parallel relation can also be considered between a group of objects $\{A_i\}$ and an object B, typically when the objects in the group are aligned and B is elongated. For example a group of boats and a deck in a port. When evaluating the relation "$\{A_i\}$ is parallel to B", actually we are evaluating that the whole set $\{A_i\}$ and the boundary of B that faces $\{A_i\}$ have a similar orientation, and that there is a large proportion of $\cup_i A_i$ that sees the boundary of B in the normal direction to the group. Similar considerations can be derived when considering



(a) (b) (c) (d)

Fig. 1. (a),(b),(c) Examples where parallelism should preferably be considered as a matter of degree, and should not be necessarily symmetrical and transitive. (d) Example of parallelism between objects with different dimensions.

the relation "B is parallel to $\{A_i\}$" or the relation between two groups of objects. All these considerations form the basis for the formal models provided in the next sections.

Notations: Let $\mathcal{S}$ be the image space, and $\mathcal{F}$ the set of fuzzy sets defined over $\mathcal{S}$. Let A denote a fuzzy set, defined through its membership function $\mu_A : \mathcal{S} \rightarrow [0, 1]$. Let $\boldsymbol{u}_{\theta_A}$ denote the normal unit vector to the principal axis of A, with angle θ_A with respect to the x-axis. Fuzzy conjunctions (t-norms) and disjunctions (t-conorms) are denoted by t and T respectively. In this work we make use of some definitions of fuzzy mathematical morphology and spatial relations, such as the directional dilation. The directional dilation of a fuzzy set μ in a direction $\boldsymbol{u}_\theta$ is defined as [3]:

$$D_{\nu_\theta}(\mu)(x) = \sup_y t[\mu(y), \nu_\theta(x - y)] \ , \tag{1}$$

where ν_θ is a fuzzy directional structuring element chosen so as to have high membership values in the direction $\boldsymbol{u}_\theta$ and its value at a point $x = (r, \alpha)$ (in polar coordinates) is a decreasing function of $|\theta - \alpha|$ modulo 2π (see Fig. 2(b)).

Another notion that will be useful is the admissibility of a segment: a segment $]a, b[$, with $a \in A$ and $b \in B$ (for A and B closed), is said to be admissible if it is included in $A^C \cap B^C$ [2]. In the fuzzy case, this extends to a degree of admissibility denoted by $\mu_{adm}(a, b)$.

3 Parallelism between Fuzzy Segments

In this section we propose a definition of parallelism between fuzzy boundaries or fuzzy lines, including the particular case of crisp linear segments, and taking into account the considerations of Sec. 2. The degree of satisfaction of the relation "A is parallel to B" should depend on the proportion of μ_A that sees μ_B in the normal direction of μ_A, and be high if the visible part of μ_B has a similar orientation to the one of μ_A. The degree to which a point $x \in \mu_A$ sees μ_B in the direction $\boldsymbol{u}_{\theta_A}$ is equivalent to the degree to which the point is seen by μ_B in the direction $\boldsymbol{u}_{\theta_A+\pi}$. To determine this degree we use the directional dilation (Eq. 1), which provides a fuzzy set, where the membership value of a point $x \in \mathcal{S}$ corresponds to the degree to which this point is visible from μ in the direction $\boldsymbol{u}_\theta$ [1,2].

Definition 1. *Let $\mu_A, \mu_B \in \mathcal{F}$. The subset of μ_A that sees μ_B in the direction $\boldsymbol{u}_{\theta_A}$ is denoted by $\mu^B_{A_\theta}$ and is equivalent to the intersection of μ_A and the fuzzy directional dilation of μ_B in direction $\boldsymbol{u}_{\theta_A+\pi}$. It has the following membership function:*

$$\forall x \in \mathcal{S}, \ \ \mu^B_{A_{\theta_A}}(x) = t[\mu_A(x), D_{\nu_{\theta_A+\pi}}(\mu_B)(x)] \ . \tag{2}$$

The set $\mu^B_{A_{\theta_A}}$ can be interpreted as the projection of μ_B onto μ_A. The proportion of μ_A that sees μ_B in the normal direction $\boldsymbol{u}_{\theta_A}$ is given by the relation $\mu_P(\mu_A, \mu_B)$ expressed as: $\mu_P(\mu_A, \mu_B) = |\mu_{A^B_{\theta_A}}|/|\mu_A|$.

(a) Original Segments (b) ν_{θ_A} (c) $D_{\nu_{\theta_A}}(\mu_A)$ (d) $\mu^B_{A_\theta}$ and $\mu^A_{B_{\theta_A+\pi}}$

Fig. 2. Illustration of the computation of parallelism between segments using directional dilation. Membership values vary from 0 (white) to 1 (black).

We have $\mu_P(\mu_A, \mu_B) = 1$ if and only if $\forall x \in S\ \mu^B_{A_{\theta_A}}(x) = \mu_A(x)$. This occurs when the projection of the segment μ_B onto μ_A is equal to μ_A.

In a similar way, we define the portion of μ_B visible from μ_A as $\forall x \in \mathcal{S},\ \mu^A_{B_{\theta_A+\pi}}(x) = t(\mu_B(x), D_{\nu_{\theta_A}}(\mu_A)(x))$ (See Fig. 2).

Definition 2. *The relation "A is parallel to B" is given by the following measure:*

$$\mu_{\|N}(\mu_A, \mu_B) = t[\mu_P(\mu_A, \mu_B), \mu_\alpha(\mu^A_{B_{\theta_A+\pi}}, \mu_A)]\ , \tag{3}$$

where $\mu_\alpha(\mu, \mu')$ is a function that penalizes large orientation differences between the orientations of μ and μ', for example:

$$\mu_\alpha(\mu, \mu') = \begin{cases} 1 & \text{if } 0 \le |\theta_\mu - \theta_{\mu'}| < a, \\ (b - |\theta_\mu - \theta_{\mu'}|)/(b-a) & \text{if } a \le |\theta_\mu - \theta_{\mu'}| < b, \\ 0 & \text{if } b \le |\theta_\mu - \theta_{\mu'}| \end{cases} \tag{4}$$

In some contexts a symmetrical relation is needed (for example in perceptual organization), and is then expressed as "A and B are parallel". In such cases, we verify that each set is visible from the other in the normal direction and that the orientations of both sets are similar, leading to the following definition.

Definition 3. *The degree of satisfaction of the symmetrical relation, "A and B are parallel" is expressed by:*

$$\mu_{\|S}(\mu_A, \mu_B) = t[T[\mu_P(\mu_B, \mu_A), \mu_P(\mu_A, \mu_B)], \mu_\alpha(\mu^B_{A_{\theta_B+\pi}}, \mu_B), \mu_\alpha(\mu^A_{B_{\theta_A+\pi}}, \mu_A)], \tag{5}$$

Proposition 1. *Both relations (Definitions 2 and 3) are invariant with respect to geometric transformations (translation, rotation, scaling).*

None of the relations is transitive, as discussed in Sec. 2. But we have the following partial result in the crisp case:

Proposition 2. *Let A, B, C be linear crisp segments, if $\mu_{\|N}(A, B) = 1$, $\mu_{\|N}(B, C) = 1$ and $\theta_A = \theta_B = \theta_C$, then $\mu_{\|N}(A, C) = 1$.*

This result shows that in the crisp case we have transitivity. To have the transitivity property, it is necessary that $\theta_A = \theta_B = \theta_C$, since $\mu_\alpha(A, B) = 1$ and

$\mu_\alpha(B,C) = 1$ do not imply $\mu_\alpha(A,C) = 1$ due to the tolerance value a of the function μ_α (See Eq. 4). To have the transitivity without imposing the condition $\theta_A = \theta_B = \theta_C$, it is necessary that μ_α is a linear function (i.e $a = 0$). But, this is restrictive.

It is clear that both relations are reflexive. However, depending on the context we may not want to consider intersecting objects as parallel. In this case, it is necessary to combine in a conjunctive way the previous degree (Def. 2 or Def. 3) with a degree of non-intersection between the two sets.

4 Parallelism between Objects

As explained in Sec. 2, parallelism can occur between more than two objects. The following paragraphs detail each situation of interest.

Parallelism between Two Objects
For objects of similar spatial extension, we evaluate the relation between the boundaries that are facing each other. These boundaries correspond to the boundaries of the objects that delimit the region between the objects, and are defined as the extremities of the admissible segments [2]. We call this portion of the boundary, the admissible boundary. When the admissible boundary of each object can be approximated by one segment the degree of satisfaction of the relation is evaluated using one of the equations presented in Sec. 3.

For the case where the admissible boundary is approximated by several segments we concentrate on the non symmetric relation. A is considered parallel to B if for every segment of the admissible boundary of A there exists a segment of the admissible boundary of B that is parallel to it.

Definition 4. *Let A and B be two fuzzy sets, defined through their membership functions μ_A and μ_B. Let $\{\mu_{\delta A_i}\}_{i=0}^{I}$ and $\{\mu_{\delta B_j}\}_{j=0}^{J}$ be the approximation by fuzzy sets of the admissible boundary of μ_A with respect to μ_B, and vice-versa. The degree of satisfaction of the relation "A is parallel to B" is defined as:*

$$\mu_{\|N}(\mu_A, \mu_B) = \sum_i |\mu_{\delta A_i}| \max_j \mu_{\|N}(\mu_{\delta A_i}, \mu_{\delta B_j}) / |T(\mu_{\delta A_0}, \ldots, \mu_{\delta A_I})| \ . \tag{6}$$

The degree to which each $\mu_{\delta A_i}$ is parallel to $\mu_{\delta B}$ is equal to $\max_j \mu_{\|N}(\mu_{\delta A_i}, \mu_{\delta B_j})$. Then this degree is weighted by the importance of $\mu_{\delta A_i}$ in the admissible boundary of A.

To calculate the degree $\mu_{\|N}(\mu_{\delta A_i}, \mu_{\delta B_j})$ Eq. 2 can be used with a modification of $\mu^B_{A_{\theta_A}}$ to take into account potential hidden parts due to concavities or corners of the objects. A point $x \in \mu_{\delta A}$ will see a point $y \in \mathcal{S}$ in the direction $\boldsymbol{u}_{\theta_A}$ if it is visible according to Eq. 1 and also if the segment $]x, y[$ is admissible (with respect to μ_A and μ_B). This is expressed as:

$$\forall x \in \mathcal{S}, \ \tilde{\mu}^B_{A_{\theta_A}}(x) = t[\mu_A(x), D_{\nu_{\theta_A+\pi}}(\mu_B)(x), \mu_{adm}(]x, y[)] \ . \tag{7}$$

When objects have different spatial extensions the boundaries that should be considered are different if we want to take into account the dimension of the

object (see Sec. 2). In this case, we can use the admissible or closest boundaries and/or include a term that expresses the relation between the principal axis of both objects.

A Ggroup of Objects Parallel to an Object

Let $\mathcal{A} = \{A_i\}_{i=0}^{I}$ be a finite set of fuzzy sets with membership functions μ_{A_i}. Let B be another fuzzy set with membership function μ_B.

For $\mathcal{A}$ to be parallel to B it is necessary that the objects of $\mathcal{A}$ are aligned. Considering each object of the group as a point (typically its center of mass), we can say that they are aligned if for every couple of points the orientation of the vector that joins them is equal to the orientation of the vector that joins the first and last points.

Definition 5. *Let m_i be the center of mass of each μ_{A_i}. Suppose that the set $\mathcal{A} = \{\mu_{A_i}\}_{i=0}^{I}$ is organized by a lexicographic order of its centers. Let μ_{align} be the relationship of alignment between fuzzy sets. This relationship is defined as:*

$$\mu_{align}(\mathcal{A}) = \min_{i<I} \mu_\alpha(T(\mu_{A_0}, \ldots, \mu_{A_I}), \{\mu_{A_i}(m_i), \mu_{A_{i+1}}(m_{i+1})\}), \tag{8}$$

where the set $\{\mu_{A_i}(m_i), \mu_{A_{i+1}}(m_{i+1})\}$ has two points and its central axis is the vector joining the two centers. The function $\mu_{\alpha'}$ has same shape as the function used in Eq. 4, and it penalizes large orientation differences.

The values of tolerance for $\mu_{\alpha'}$ can be different from those used for the parallel relation. This definition considers that objects have similar dimensions, and it does not take into account the distance between the objects.

To evaluate the degree of satisfaction of "$\mathcal{A}$ is parallel to B", we calculate for every i the portion of the closest boundary of μ_{A_i} to μ_B, which we denote $\mu_{\gamma A_i}$. And for μ_B we consider the linear boundary $\mu_{\gamma B}$ that is closest to the group.

Definition 6. *The degree of satisfaction of the relation "$\mathcal{A}$ is be parallel to B" is given by:*

$$\begin{aligned}\mu_{\|N}(\mathcal{A}, \mu_B) = t[&\mu_P(T(\mu_{\gamma A_0}, \ldots, \mu_{\gamma A_I}), \mu_{\gamma B}),\\ &\mu_\alpha(T(\mu_{A_0}, \ldots, \mu_{A_I}), \mu_{\gamma B}), \mu_{align}(\mathcal{A})].\end{aligned} \tag{9}$$

An Object Parallel to a Group of Objects

Using the same notations as above, let us assume that the set $\{\mu_{A_i}\}_{i=0}^{I}$ is organized by a lexicographic order of its centers. Let $\beta_A \in \mathcal{F}$ be the region composed of the union of the regions between two consecutive elements of $\mathcal{A}$ (see [2]). For "B is parallel to $\mathcal{A}$ " to be true, it is necessary that the objects in the group are aligned, that $\mu_{\gamma B}$ and the group of objects have a similar orientation and that there is a large proportion of $\mu_{\gamma B}$ that sees the group of objects or β_A:

Definition 7. *The degree of satisfaction of the relation "B is parallel to $\mathcal{A}$" is given by:*

$$\mu_{\|N}(\mu_B, \mathcal{A}) = t[\mu_P(\mu_{\gamma B}, \mu_{\gamma A'}), \mu_\alpha(T(\mu_{A_0}, \ldots, \mu_{A_I}), \mu_{\gamma B}), \mu_{align}(\mathcal{A})] \ , \tag{10}$$

where $\mu_{\gamma A'}$ denotes the admissible boundaries of $T(\mu_{A_0}, \ldots, \mu_{A_I}, \beta_A)$.

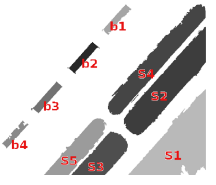
Using the same notation as in Def. 7, we can define the parallelism between two finite sets of fuzzy sets $\mathcal{A} = \{A_i\}_{i=0}^{I}$ and $\mathcal{B} = \{B_j\}_{j=0}^{J}$:

Definition 8. *The degree of satisfaction of the relation "$\mathcal{A}$ is parallel to $\mathcal{B}$" is given by:*

$$\mu_{\|N}(\mathcal{A},\mathcal{B}) = t[\mu_P(T(\mu_{\gamma A_0},\dots\mu_{\gamma A_I}),\mu_{\gamma B'}),\mu_{align}(\mathcal{A}),\mu_{align}(\mathcal{B}), \\ \mu_\alpha(T(\mu_{A_0},\dots,\mu_{A_I}),T(\mu_{B_0},\dots,\mu_{B_J}))] \ , \tag{11}$$

5 Results

We evaluated the parallel relation between two objects for the labeled objects of Figs. 3 and 4. For these examples we used $a = \pi/12$ and $b = \pi/6$ in Eq. 4.



A	B	$\mu_{\|\|N}(A,B)$	$\mu_{\|\|N}(B,A)$
b2	S4	0.94	0.55
b3	S5	0.97	0.87
b4	S5	0.89	0.66
S2	S4	0.97	0.97
S4	S1	0.87	0.94
S5	S3	0.90	0.95
S3	S1	0.78	0.43
b1	S4	0.90	0.69

Fig. 3. Original image, segmented image and results

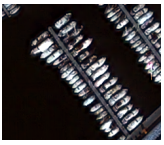
To evaluate the relation between objects with different spatial extensions, we used the closest boundaries that had a similar orientation to the principal axis of the objects. From Fig. 3 we observe that the obtained results fit with the intuition. For objects that have similar spatial extensions ($S2$ and $S4$ or $S3$ and $S5$), similar values were obtained for $\mu_{||N}(A,B)$ and $\mu_{||N}(B,A)$. The results for $b2$ and $S4$, $b1$ and $S4$ and $b4$ and $S5$ reflect that when objects have different extensions the results are not symmetric.

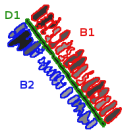
A	B	$\mu_{\|\|N}(A,B)$	$\mu_{\|\|N}(B,A)$
R1	R2	0.93	0.85
R2	H1	0.14	0.29
R2	H2	0.42	0.69
R2	H3	0.22	0.74
R1	H3	0.00	0.38

Fig. 4. Original image, segmented image and results

Figure 4 shows a fuzzy segmentation of the image. Again the objects with similar spatial extension $R1$ and $R2$ have similar values for $\mu_{||N}(A,B)$ and $\mu_{||N}(B,A)$.

Our definitions for the parallelism between a group of objects and an object (Def. 7 and 6) were applied to the objects in Fig. 5. We used $a = \pi/18$ and $b = \pi/6$ in $\mu_{\alpha'}$ of Def. 5. As the boundaries involved in the relation of $B1$, $B2$ and $D1$ have similar spatial extensions the results are almost symmetrical, agreeing with intuition.





A	B	$\mu_{\|\|_N}(A,B)$	$\mu_{\|\|_N}(B,A)$
B1	D1	0.94	0.94
B2	D1	0.95	0.95
B1	B2	0.85	0.87

Fig. 5. Original image, segmented image and results

6 Conclusion

In this work we discussed the considerations that should be taken into account when modeling the parallel relation. We highlighted that the parallel relation depends on the situation and the context. We presented a definition of parallelism between two objects of similar spatial extensions, and briefly discussed the case of objects with different spatial extensions. Illustrations on real objects show the interest of the proposed definition.

Future work aims at extending the notion of alignment and parallelism to objects of different sizes and to more complex situations.

Acknowledgement. This work was done within the CNES-DLR-ENST Competence Center.

References

1. Bloch, I.: Fuzzy spatial relationships for image processing and interpretation: a review. Image and Vision Computing 23(2), 89–110 (2005)
2. Bloch, I., Colliot, O., Cesar Jr., R.M.: On the ternary spatial relation between. IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics 36(2), 312–327 (2006)
3. Bloch, I., Maître, H.: Fuzzy mathematical morphologies: A comparative study. Pattern Recognition 28(9), 1341–1387 (1995)
4. Ip, H.H.S., Wong, W.H.: Detecting perceptually parallel curves: Criteria and force driven optimization. Computer Vision and Image Understanding 68(2), 190–208 (1997)
5. Penedo, M.G., Ortega, M., Roucu, J., Penas, M., Alonso-Montes, C.: Certainty measure of pairwise line segment perceptual relations using fuzzy logic. In: Rueda, L., Mery, D., Kittler, J. (eds.) CIARP 2007. LNCS, vol. 4756, pp. 477–486. Springer, Heidelberg (2007)
6. Kang, H.-B., Walker, E.L.: Perceptual grouping based on fuzzy sets. In: IEEE International Conference on Fuzzy Systems, pp. 651–659 (1992)
7. Lowe, D.G.: Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence 31(3), 355–395 (1987)
8. Sarkar, S., Boyer, K.L.: Perceptual organization in computer vision: a review and a proposal for a classificatory structure. IEEE Transactions on Systems, Man, and Cybernetics 23(2), 382–399 (1993)
9. Shen, D., Wong, W., Ip, H.H.S.: Affine-invariant image retrieval by correspondence matching of shapes. Image and Vision Computing 17(7), 489–499 (1999)

Barycentric Algebras and Gene Expression

Anna B. Romanowska[1] and Jonathan D.H. Smith[2]

[1] Faculty of Mathematics and Information Sciences
Warsaw University of Technology
00-661 Warsaw, Poland
aroman@alpha.mini.pw.edu.pl
[2] Department of Mathematics
Iowa State University
Ames, Iowa, 50011, USA
jdhsmith@math.iastate.edu

Abstract. Barycentric algebras have seen widespread application in the modeling of convex sets, semilattices, and quantum mechanics. Recently, they were developed further to encompass Boolean logic and if-then-else algebras. This paper discusses an application of barycentric algebras to systems biology. Here, they provide a calculus for the conversion from simplified Boolean models of gene transcription to fuzzy models that give a more realistic tracking of the biochemistry. Indeed, it appears that logic gates experimentally observed in cells actually follow the barycentric algebra format.

1 Introduction

Barycentric algebras (as defined in §2.3 below) are universal algebras used for modeling convex sets, semilattices, geometry, hierarchical statistical mechanics, and quantum mechanics [5,6,12,13,14,15,16,17,18]. Recently [17], they have been developed further (as *abstract barycentric algebras*) by use of the LΠ-algebras of fuzzy logic [3,10,11], incorporating B-sets [2,20,21] and if-then-else algebras [8,9]. The aim of the current paper is to show how the calculus of barycentric algebras may be used in systems biology, to provide a virtually automatic translation from simplified Boolean models of gene expression to continuous, fuzzy logic models that give a much more realistic picture of the biochemical processes involved. Experimentally observed logic gates in cells do not follow the pattern directly suggested by standard Boolean models, but their features concur exactly with the models obtained using the barycentric algebra approach [19, Fig. 3b].

The bulk of the paper comprises two parts. Section 2 gives a direct account of the algebra required. For readers who may be unfamiliar with universal algebra, §2.2 discusses concatenations of binary operations. The two key incarnations of abstract barycentric algebras, namely classic "fuzzy" barycentric algebras and their crisp Boolean counterparts, are described in §2.3.

Section 3 then focusses on the systems biology. For readers unfamiliar with molecular biology, §3.1 gives a brief account of the way cells use transcription

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 20–27, 2009.

factors to respond to signals and regulate gene expression. Subsequent paragraphs formulate the crisp and fuzzy models of gene regulation in the language of barycentric algebras. Once this formulation is established, Eqn. (12) provides the automatic conversion from Boolean models to fuzzy models. In §3.4, the conversion process is illustrated by the example of the AND gate. The final paragraph explains how fuzzy logic gates that have been observed experimentally in cells actually follow the barycentric algebra format.

2 Algebra

2.1 Operations on Real Numbers and Binary Digits

Although the algebra of real numbers is traditionally performed in terms of field operations such as the addition $p+q$ and product pq of real numbers p and q, the algebra discussed in this paper requires different operations, which specialize to more familiar Boolean operations on the subset $\{0,1\}$ of the reals. In fact, this specialization will also work in any field. In particular, it works if the set $\{0,1\}$ of binary digits is interpreted as the two-element (Galois) field GF(2) or field of integers modulo 2.

For a real number p, define the *complementation* $p' = 1 - p$ specializing to the Boolean $\neg p$ or NOT p on the set $\{0.1\}$ of binary digits. Note that the complementation is *involutive*: $p'' = p$. For real numbers p and q, define the *product*

$$p \cdot q = pq \tag{1}$$

specializing to the Boolean $\wedge$ or AND on $\{0,1\}$. Define the *dual product*

$$p \circ q = p + q - pq \tag{2}$$

specializing to the Boolean $\vee$ or OR on $\{0,1\}$. Note that the dual product may be defined in terms of the product and complementation using *de Morgan's law* $p \circ q = (p'q')'$ or $(p \circ q)' = p'q'$. Define the *implication*

$$p \to q = \textbf{if } (p = 0) \textbf{ then } 1 \textbf{ else } q/p \tag{3}$$

specializing to the Boolean implication $p \to q = (\neg p) \vee q$ on $\{0,1\}$. Note that the implication (3) is always defined in any field, while the division q/p is not defined for $p = 0$.

2.2 Binary Operations

If x and y are elements of a real vector space, and p is a real number, it is convenient to define

$$xy\,\underline{p} = x(1-p) + yp = xp' + yp\,, \tag{4}$$

so that $\underline{p}$ is understood as a binary operation combining the arguments x and y. Schematically, the binary operation may be understood as a circuit element or

"black box" combining the inputs x and y to produce the output $xy\,\underline{p}$. For a second real number q and vector z, one may concatenate circuit elements to yield

$$xy\,\underline{p}\,z\,\underline{q} = (xp' + yp)z\underline{q} = xp'q' + ypq' + zq\,. \tag{5}$$

Alternatively one may concatenate the circuit elements to yield

$$x\,yz\,\underline{p}\;\underline{q} = x(yp' + zp)\underline{q} = xq' + yp'q + zpq\,. \tag{6}$$

Note that the parsing of the left hand sides of (5) and (6) is unique, even without the insertion of any brackets. This is one of the many advantages of the algebraic notation (4) for binary operations.

If p is an element of the closed real unit interval $I = [0,1] = \{p \mid 0 \leq p \leq 1\}$, then the operation (4) makes sense when the inputs x and y lie in some convex set C, for example some interval on the real line.

If p is a binary digit 0 or 1, the operation (4) makes sense when the inputs x and y are elements of some arbitrary set S, with

$$xy\,\underline{p} = \textbf{if } (p = 1) \textbf{ then } y \textbf{ else } x\,.$$

Recalling that the *truth value* $[\![P]\!]$ of a proposition P is 0 if P is false, and 1 if P is true, one obtains

$$xy\,\underline{[\![P]\!]} = \textbf{if } P \textbf{ then } y \textbf{ else } x\,. \tag{7}$$

Given an arbitrary set S, consider the convex set C of all finite probability distributions on S, identifying each element x of S with the distribution putting weight 1 on x. For elements x and y of S, and p in I, the operation (4) produces the distribution selecting y with probability p and x with probability p'.

2.3 Barycentric Algebras and If-Then-Else Algebras

An *abstract barycentric algebra* is defined as a set A that is equipped with binary operations $xy\,\underline{p}$ satisfying *idempotence* $xx\,\underline{p} = x$ for x in A, *skew-commutativity*

$$xy\,\underline{p} = yx\,\underline{p'} \tag{8}$$

for x, y in A, and *skew-associativity* $xy\,\underline{p}\,z\,\underline{q} = x\,yz\,\big(\underline{p \circ q \to q}\big)\,\underline{p \circ q}$ for x, y, z in A. There are two classical interpretations:

- Taking the operators p, q from the open real unit interval

$$I^\circ =]0,1[= \{p \mid 0 < p < 1\}$$

  yields a *barycentric algebra* [14,15,16].
- Taking Boolean operators p, q — elements of a Boolean ring such as GF(2) or its powers — yields *B-sets* [2,20], including certain types of *if-then-else algebras* [8,20].

Within abstract barycentric algebras, concatenations of the type (6) serve to implement the "AND" product (1) as

$$x\,xy\,\underline{p}\;\underline{q} = xy\,\underline{p\cdot q}\,, \tag{9}$$

while concatenations of the type (5) implement the dual "OR" product (2) as $xy\,\underline{p}\,y\,\underline{q} = xy\,\underline{p\circ q}$. Of course, skew-commutativity gives a direct implementation of the complement.

3 Systems Biology

3.1 Transcription Factors

Cells survive and develop by producing proteins in response to various signals that they receive. We describe a simplified model that will be adequate for the purposes of this paper. For fuller details, see [1,7]. A specific protein Y is produced by the expression of a corresponding part of the cell's DNA, namely the gene that encodes for protein Y. The gene is first *transcribed* to messenger RNA (mRNA). The mRNA is then *translated* into the required protein. The transcription process, synthesis of the mRNA, is facilitated by the enzyme RNA polymerase (RNAp). The enzyme binds itself to a regulatory region of the DNA, adjacent to the gene, known as the *promoter site*.

Signals that are of importance to a cell may be physical, such as a change in temperature, or chemical, such as the presence of a nutrient like glucose. Received signals switch proteins known as *transcription factors* from a dormant to an active state. Active transcription factors attach themselves to the promoter site, where they change the binding probability of the RNAp. If a transcription factor is an *activator*, it will increase the binding probability of the RNAp, thereby increasing the rate of transcription and protein production. Other transcription factors, known as *repressors*, have the opposite effect of inhibiting the expression of certain genes.

3.2 Crisp Logic

Fig. 1 displays sample dependencies of the transcription rate for production of a protein on the relative concentration x/k of an activator X. In the absence of the activator, the transcription rate assumes a *residual base level* v_0, in this case 0.1. (Often, a value of $v_0 = 0$ is appropriate.) If the activator is present in high concentrations, the transcription rate assumes a *maximal expression level* v_1, in this case 1.0.

The step function displays a crisp logical dependence of the transcription rate on the dimensionless ratio x/k between the actual concentration x of the activator X, and a critical threshold concentration level k. The transcription rate may be written as

$$v_0 v_1 \underline{\left[\!\left[\, 1 > \frac{k}{x} \,\right]\!\right]} \tag{10}$$

in the Boolean notation of (7). If the transcription factor X were a repressor rather than an activator, the corresponding transcription rate would appear in any of the forms

$$v_0 v_1 \underline{\left[\!\left[\, 1 > \frac{k}{x} \,\right]\!\right]}' = v_1 v_0 \underline{\left[\!\left[\, 1 > \frac{k}{x} \,\right]\!\right]} = v_0 v_1 \underline{\left[\!\left[\, 1 > \frac{x}{k} \,\right]\!\right]} \tag{11}$$

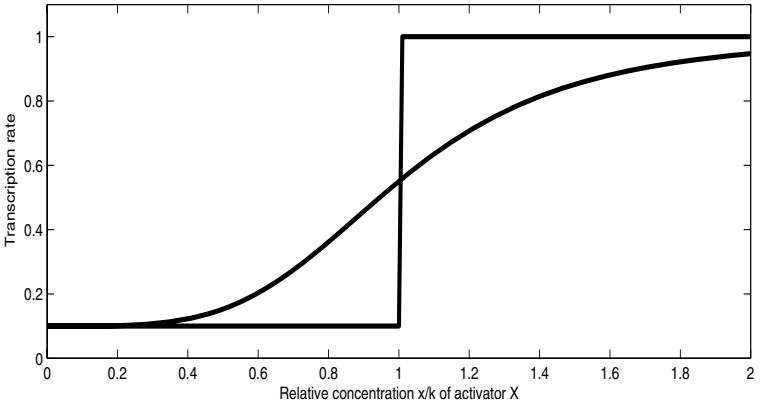


Fig. 1. Dependence of transcription rate on activator X

that are equivalent by virtue of the skew-commutativity (8) that implements complementation. (The last form neglects the improbable equality $x = k$.)

3.3 Fuzzy Logic

Because of its convenience, crisp logic has been used widely for the construction of network models in systems biology [4,22]. However, the curved graph of Fig. 1 displays a more realistic description of the dependence of transcription rates on the relative concentrations of transcription factors. One of the main theses of this paper is the way that the formalism of abstract barycentric algebras allows one to convert easily from the crisp functions of §3.2 to more realistic fuzzy functions. The crisp activator dependence (10) is replaced by the classic barycentric-algebraic expression

$$v_0 v_1 \underline{\left[\, 1 + \left(\frac{k}{x}\right)^n \,\right]^{-1}}$$

— a so-called *hill function* in the terminology of [1] — interpreted in the closed interval $[v_0, v_1]$, a convex set. Fig. 1 illustrates the case $n = 4$. For $n = 1$ (and $v_0 = 0$), the hill function implements Michaelis-Menten kinetics [1, A.7]. The case $n > 1$ corresponds to *cooperative reactions*. The crisp repressor dependencies (11) are replaced by either of the equivalent forms

$$v_1 v_0 \underline{\left[\, 1 + \left(\frac{k}{x}\right)^n \,\right]^{-1}} = v_0 v_1 \underline{\left[\, 1 + \left(\frac{x}{k}\right)^n \,\right]^{-1}}.$$

From these expressions, it is clear that the passage from crisp to fuzzy logic is formally achieved by the replacement

$$[\![1 > \lambda]\!] \qquad \longrightarrow \qquad [1 + \lambda^n]^{-1} \,. \tag{12}$$

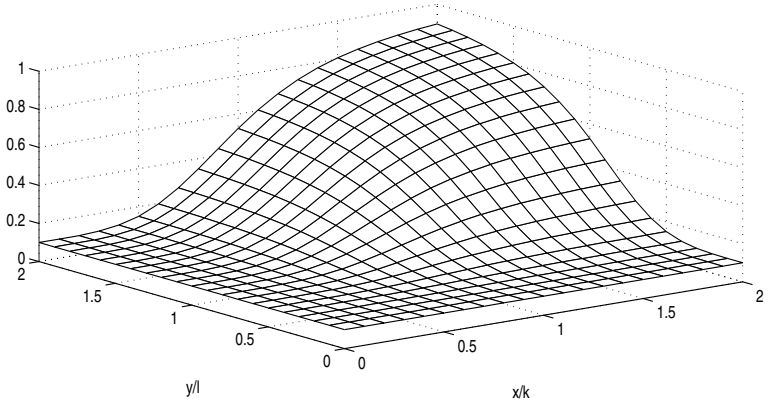


Fig. 2. Fuzzy AND gate

Here, the dimensionless quantity λ is taken as k/x for activators and x/k for repressors. The conversion process is illustrated in the following paragraph.

3.4 Logic Gates

Transcription rates may depend on logical combinations of different transcription factors. For example, the dependence

$$v_0 v_1 \underline{\left[\!\left[\, 1 > \frac{k}{x} \,\right]\!\right] \cdot \left[\!\left[\, 1 > \frac{l}{y} \,\right]\!\right]} \tag{13}$$

requires high concentrations of each of two transcription factors X and Y. The concentration x of X must exceed the critical threshold k; the concentration y of Y must exceed the critical threshold l. Using (9), the crisp logical expression (13) may be rewritten as the concatenation

$$v_0\, v_0 v_1 \underline{\left[\!\left[\, 1 > \frac{k}{x} \,\right]\!\right]}\ \underline{\left[\!\left[\, 1 > \frac{l}{y} \,\right]\!\right]}$$

which then translates to

$$v_0\, v_0 v_1 \underline{\left[1 + \left(\frac{k}{x}\right)^n\right]^{-1}}\ \underline{\left[1 + \left(\frac{l}{y}\right)^n\right]^{-1}} \tag{14}$$

under the replacement (12). With the previously used parameter values $v_0 = 0.1$, $v_1 = 1$, $k = 1$, $n = 4$, along with $l = 1$, this fuzzy AND gate is displayed in Fig. 2.

3.5 Some Experimental Observations

The fuzzy AND gate presented in (14) has the format $v_0\, v_0 v_1\, \underline{p}\ \underline{q}$ of (9). Here, the concatenated barycentric algebra operations have arguments (corresponding

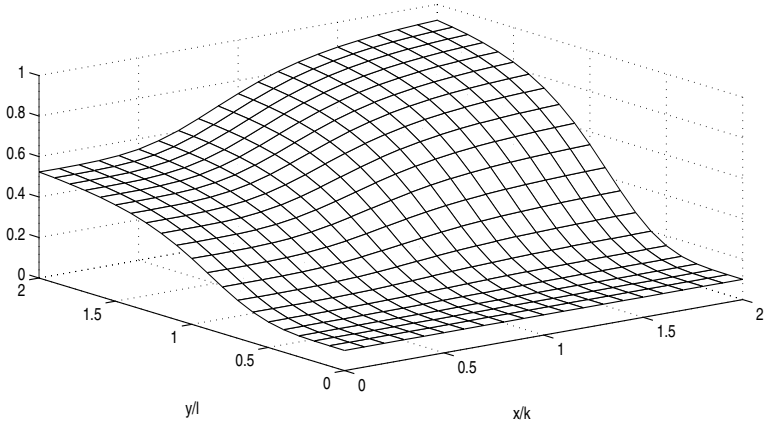


Fig. 3. Modified fuzzy AND gate

to transcription rates) that are repeated exactly. Exact repeats of this kind are improbable in biology. At first glance, it might appear that this would argue against the barycentric algebra approach. However, it turns out that real fuzzy AND gates as observed experimentally [19, Fig. 3b] actually have the format $v_0\, v_{01} v_1\, \underline{p}\, \underline{q}$ of (6) with distinct transcription rates $v_0 < v_{01} < v_1$, as illustrated in Fig. 3 using an intermediate expression level $v_{01} = 0.55$. It thus emerges that the barycentric algebra formulation gives a natural framework for the dependence of expression levels on transcription factor concentrations.

References

1. Alon, U.: An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, Boca Raton (2006)
2. Bergman, G.M.: Actions of Boolean rings on sets. Algebra Universalis 28, 153–187 (1991)
3. Esteva, F., Godo, L., Montagna, F.: The LΠ and LΠ1/2 logics: two complete fuzzy systems joining Łukasiewicz and product logics. Arch. Math. Logic 40, 39–67 (2001)
4. Glass, L., Kauffman, S.A.: The logical analysis of continuous, non-linear biochemical control networks. J. Theor. Biol. 39, 103–129 (1973)
5. Gudder, S.P.: Convex structures and operational quantum mechanics. Comm. Math. Phys. 29, 249–264 (1973)
6. Ignatov, V.V.: Quasivarieties of convexors (in Russian). Izv. Vyssh. Uchebn. Zaved. Mat. 29, 12–14 (1985)
7. Klipp, E., Herwig, R., Kowald, C., Wierling, C., Lehrach, H.: Systems Biology in Practice. Wiley, New York (2005)
8. Manes, E.G.: Adas and the equational theory of if-then-else. Algebra Universalis 30, 373–394 (1993)
9. McCarthy, J.A.: A basis for a mathematical theory of computation. In: Braffort, P., Hirschberg, D. (eds.) Computer Programming and Formal Systems, pp. 33–70. North-Holland, Amsterdam (1993)

10. Montagna, F.: An algebraic approach to propositional fuzzy logic. Journal of Logic, Language and Information 9, 91–124 (2000)
11. Montagna, F.: Subreducts of MV-algebras with product and product residuation. Algebra Universalis 53, 109–137 (2005)
12. Neumann, W.D.: On the quasivariety of convex subsets of affine spaces. Arch. Math. 21, 11–16 (1970)
13. Ostermann, F., Schmidt, J.: Der baryzentrische Kalkül als axiomatische Grundlage der affinen Geometrie. J. Reine Angew. Math. 224, 44–57 (1966)
14. Romanowska, A.B., Smith, J.D.H.: Modal Theory. Heldermann, Berlin (1985)
15. Romanowska, A.B., Smith, J.D.H.: On the structure of barycentric algebras. Houston J. Math. 16, 431–448 (1990)
16. Romanowska, A.B., Smith, J.D.H.: Modes. World Scientific, Singapore (2002)
17. Orłowska, E., Romanowska, A.B., Smith, J.D.H.: Abstract barycentric algebras. Fund. Informaticae 81, 257–273 (2007)
18. Skornyakov, L.A.: Stochastic algebras. Izv. Vyssh. Uchebn. Zaved. Mat. 29, 3–11 (1985)
19. Setty, Y., Mayo, A.E., Surette, M.G., Alon, U.: Detailed map of a cis-regulatory input function. Proc. Nat. Acad. Sci. 100, 7702–7707 (2003)
20. Stokes, T.: Sets with B-action and linear algebra. Algebra Universalis 39, 31–43 (1998)
21. Stokes, T.: Radical classes of algebras with B-action. Algebra Universalis 40, 73–85 (1998)
22. Thieffry, D., Thomas, R.: Qualitative analysis of gene networks. In: Pac. Symp. Biocomput., vol. 3, pp. 77–88 (1998)

Fuzzy Quantification Using Restriction Levels

Daniel Sánchez, Miguel Delgado, and María-Amparo Vila

Dept. Computer Science and A.I., University of Granada
E.T.S.I.I.T., C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
daniel@decsai.ugr.es, mdelgado@ugr.es, vila@decsai.ugr.es

Abstract. We introduce a model for the evaluation of fuzzy quantified expressions involving imprecise concepts. Imprecise concepts are assumed to be represented via restriction levels, a recently introduced representation of imprecision that extends the representation of fuzzy sets and introduces new operators. The proposal verifies all the properties required for the evaluation of quantified sentences, including all the properties involving negation. Particularly, the evaluation of "Some of A are ¬A" is definitely 0 for any fuzzy set A.

Keywords: Fuzzy quantification, restriction level representation.

1 Introduction

Fuzzy quantification extends classical quantification by considering (fuzzy) linguistic quantifiers, a generalization of the ordinary quantifiers $\exists$ and $\forall$ of first order logic [12]. A large number of applications can be found in the literature in areas like quantifier-guided aggregation, linguistic summarization, computing with words and quantification in fuzzy description logics, among many others.

The most usual quantified sentences considered in the literature are of the form "Q of X are A" or "Q of D are A", called type I and type II sentences, respectively. Here Q is a linguistic quantifier, X is a (finite) crisp set, and A,D are fuzzy subsets of X. Linguistic quantifiers are normal, convex fuzzy subsets of $\mathbb{Z}$ (absolute quantifiers) or $[0,1]$ (relative quantifiers).

There are many different approaches for evaluating quantified sentences [12, 10, 2, 5, 4, 1, 6, 11, 3, 7]. Properties that any suitable method should verify have been proposed in [5, 4, 6]. However, to the best of our knowledge, no existing method verifies all the properties. Particularly difficult is to verify simultaneously the properties of low computational complexity and those involving negation of properties and quantifiers, and quantifier antonyms.

In this paper we introduce an alternative approach to the evaluation of quantified sentences on the basis of restriction levels (RL), a recently introduced representation of imprecision that extends the representation of fuzzy sets and introduces new operators [8, 9]. RL-representation is similar to the representation of imprecise concepts by means of a collection of α-cuts, but differs from it in that crisp operations and definitions of any kind are extended to the imprecise case by operating on each level independently. This way, the result is not

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 28–35, 2009.

necessarily a fuzzy set, the latter being a particular case of RL-representation. We employ this approach since, among other properties, operations extended to the imprecise case using RL-representations verify all the properties of the crisp case and, in addition, their computational complexity is very close to that of the crisp case when a finite, fixed number of levels is considered, as usual in practice.

2 Restriction-Level Representation

2.1 Representation

The RL-representation of an imprecise property is a collection of crisp sets, each crisp set corresponding to a crisp realization of the property under a *restriction* rule. We distinguish between *atomic* and *derived* properties. Atomic properties are those that cannot be defined in terms of other properties in our problem. Derived properties are defined by logical operations on other properties.

In [8] we consider that atomic imprecise properties are represented by fuzzy sets, and hence restrictions are of the form *degree* $\geq \alpha$ with $\alpha \in (0, 1]$, and restriction levels are associated to values $\alpha \in (0, 1]$. In the same case, the crisp realization of an atomic imprecise property represented by a fuzzy set A in the restriction level α corresponds to the α-cut A_α.

For every property we assume that there is a finite set of restriction levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$. We call such sets *RL-sets*. The consideration that a RL-set is finite is not a practical limitation since humans are able to distinguish a limited number of restriction or precision levels and, in practice, the limit in precision and storage of computers allows us to work with a finite number of degrees (and consequently, of levels) only. In practice, the RL-set for an atomic property represented by a fuzzy set A on an universe X is

$$\Lambda_A = \{A(x) \mid x \in \mathit{support}(A)\} \cup \{1\} \tag{1}$$

The RL-set employed to represent a derived property is obtained as the union of the RL-sets of the atomic properties in terms of which the property is defined. Finally, the RL-representation of an imprecise property on X is defined in [8] as follows:

Definition 1. *A RL-representation is a pair* (Λ, ρ) *where* Λ *is a RL-set and* ρ *is a function*

$$\rho : \Lambda \rightarrow \mathcal{P}(X) \tag{2}$$

The function ρ indicates the crisp realization that represents the imprecise property for each restriction level. As an example, the RL-representation for an atomic imprecise property defined by a fuzzy set A is the pair (Λ_A, ρ_A), where Λ_A is obtained using equation (1), and $\rho_A(\alpha) = A_\alpha \ \forall \alpha \in \Lambda_A$.

Given an imprecise property P represented by (Λ_P, ρ_P), we define the set of crisp representatives of P, Ω_P, as

$$\Omega_P = \{\rho_P(\alpha) \mid \alpha \in \Lambda_P\} \tag{3}$$

For an atomic property A, the set of crisp representatives Ω_A is the set of significant α-cuts of A, as we have seen. However, notice that in definition 1 there is no restriction about the possible crisp representatives for non-atomic properties. In particular, as a consequence of operations, they don't need to be nested, so the final RL-representation of a derived property is not always equivalent to the α-cut representation of fuzzy sets.

In order to define properties by operations, it is convenient to extend the function ρ to any RL $\alpha \in (0,1]$. Let (Λ, ρ) be a RL-representation with $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$. Let $\alpha \in (0,1]$ and $\alpha_i, \alpha_{i+1} \in \Lambda$ such that $\alpha_i \geq \alpha > \alpha_{i+1}$. Then

$$\rho(\alpha) = \rho(\alpha_i) \tag{4}$$

Finally, let us remark that a RL-representation (Λ, ρ) on X is a crisp set $A \subseteq X$ iff $\forall \alpha \in \Lambda,\ \rho(\alpha) = A$.

2.2 Interpretation in Terms of Evidence

Given a RL-representation (Λ_A, ρ_A) for an atomic property A, the values of Λ_A can be interpreted as values of possibility of a possibility measure defined $\forall \rho_A(\alpha_i) \in \Omega_A$ as

$$Pos(\rho_A(\alpha_i)) = \alpha_i. \tag{5}$$

Following this interpretation we define a basic probability assignment in the usual way:

Definition 2. *Let (Λ, ρ) be a RL-representation with crisp representatives Ω. The associated probability distribution $m : \Omega \to [0,1]$ is*

$$m(Y) = \sum_{\alpha_i \mid Y = \rho(\alpha_i)} \alpha_i - \alpha_{i+1}. \tag{6}$$

The basic probability assignment m_F gives us information about how representative of the property F is each crisp set in Ω_F. For each $Y \in \Omega_F$, the value $m_F(Y)$ represents the proportion to which the available evidence supports the claim that the property F is represented by Y. From this point of view, a RL-representation can be seen as a basic probability assignment in the sense of the theory of evidence, *plus a structure indicating dependencies between the possible representations of different properties.*

2.3 RL-Numbers

On the basis of RL-representations and operations, RL-numbers were introduced in [9] as a representation of imprecise quantities.

Definition 3. *A RL-real number is a pair $(\Lambda, \mathcal{R})$ where Λ is a RL-set and $\mathcal{R} : (0,1] \to \mathbb{R}$.*

We shall note $\mathbb{R}_{RL}$ the set of RL-real numbers. The RL-real number R_x is the representation of a (precise) real number x iff $\forall \alpha \in \Lambda_{R_x}$, $\mathcal{R}_{R_x}(\alpha) = x$. We shall denote such RL-real number as R_x or, equivalently, x, since in the crisp case, the set Λ_{R_x} is unimportant. Operations are extended as follows:

Definition 4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $R_1 \dots R_n$ be RL-real numbers. Then $f(R_1, \dots, R_n)$ is a RL-real number with*

$$\Lambda_{f(R_1,\dots,R_n)} = \bigcup_{1 \le i \le n} \Lambda_{R_i} \tag{7}$$

and, $\forall \alpha \in \Lambda_{f(R_1,\dots,R_n)}$

$$\mathcal{R}_{f(R_1,\dots,R_n)}(\alpha) = f(\mathcal{R}_{R_1}(\alpha), \dots, \mathcal{R}_{R_n}(\alpha)) \tag{8}$$

This approach offers two main advantages:

- RL-numbers are representations of imprecise quantities that can be easily obtained by extending usual crisp measurements to fuzzy sets.
- Arithmetic and logical operations on RL-numbers are straightforward and unique extensions of the operations on crisp numbers, verifying the following:
  - They verify all the usual properties of crisp arithmetic and logical operations.
  - The imprecision does not necessarily increase through operations, and can even diminish. The maximum imprecision is related to the number of restriction levels employed.

3 Evaluation of Quantified Sentences

We shall consider the evaluation of quantified sentences of type II because of lack of space, and since type I sentences are a particular case of type II sentences, under the following assumptions:

- Q is a fuzzy quantifier
- A,D are imprecise properties defined on a finite, crisp set X by RL-representations (Λ_A, ρ_A) and (Λ_D, ρ_D), respectively.

Notice that, since fuzzy sets are particular cases of RL-representations, the proposal in this section is also valid for the evaluation of quantified sentences when properties A,D are represented by fuzzy sets.

3.1 RL-Evaluation

Definition 5. *The evaluation of $E \equiv$ "Q of D are A" is a RL-truth degree, i.e., a RL-real number E in $[0, 1]$, defined by (Λ_E, ρ_E), where*

$$\Lambda_E = \Lambda_A \cup \Lambda_D \tag{9}$$

and, $\forall \alpha \in \Lambda_E$,

$$\rho_E(\alpha) = Q\left(\frac{|\rho_{A \wedge D}(\alpha)|}{|\rho_D(\alpha)|}\right) = Q\left(\frac{|\rho_A(\alpha) \cap \rho_D(\alpha)|}{|\rho_D(\alpha)|}\right) \tag{10}$$

Table 1. RL-representation of several properties derived from two atomic properties A and B

α	$\rho_A(\alpha)$	$\rho_{\neg A}(\alpha)$	$\rho_D(\alpha)$	$\rho_{\neg D}(\alpha)$	$\rho_{A\wedge\neg D}(\alpha)$	$\rho_{D\wedge\neg A}(\alpha)$
1	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\emptyset$	X	$\{x_1\}$	$\emptyset$
0.9	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\emptyset$	$\emptyset$
0.8	$\{x_1, x_2\}$	$\{x_3, x_4, x_5\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_2\}$	$\emptyset$
0.6	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_1, x_3\}$	$\{x_2, x_4, x_5\}$	$\{x_2\}$	$\emptyset$
0.5	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$	$\{x_2\}$	$\{x_4\}$
0.4	$\{x_1, x_2, x_3, x_5\}$	$\{x_4\}$	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$	$\{x_2, x_5\}$	$\{x_4\}$

Table 2. Some quantified sentences involving A and D

E_1	Q_{most} of A are D	E_3	Q_{most} of $\neg$ A are D	E_5	$\exists$ of A are $\neg$ A
E_2	Q_{most} of A are $\neg$ D	E_4	Q_{most} of $\neg$ A are $\neg$ D	E_6	$\exists$ of (A $\wedge$ $\neg$ D) are D

As an example, consider the atomic properties A and D defined by the following fuzzy sets:

$$A = 1/x_1 + 0.8/x_2 + 0.5/x_3 + 0.4/x_5$$

$$D = 0.9/x_1 + 0.6/x_3 + 0.5/x_4$$

Then we have $\Lambda_A = \Lambda_{\neg A} = \{1, 0.8, 0.5, 0.4\}$ and $\Lambda_D = \Lambda_{\neg D} = \{1, 0.9, 0.6, 0.5\}$. In Table 1 we can see several properties derived from A and B using $\wedge$, $\vee$, and negation. It is easy to see that the result is not always a fuzzy set, specifically when negation is involved (conjunction and disjunction of fuzzy sets via levels is equivalent to minimum and maximum, respectively, but not for any RL-representation in general). Notice that since RL-representations verify the classical properties, the representation of $A \wedge \neg A$ is $\emptyset$ in every level, i.e., a classical, crisp contradiction.

Table 2 shows a set of quantified sentences involving properties A and D, using the quantifier Q_{most} defined as in Equation 11. Table 3 shows the corresponding RL-evaluation of the sentences in table 2 and the same sentences but replacing the quantifiers by $Q(x) = x$ (E_7 is E_1 replacing the quantifier, and so on). The latter correspond to the measure of the relative cardinality of the corresponding properties as well.

$$Q_{most}(x) = \begin{cases} 0 & x \leq 0.5 \\ 1 & x \geq 0.75 \\ 4(x - 0.5) & \text{otherwise} \end{cases} \tag{11}$$

3.2 Properties

From definition 5 it is obvious that RL-evaluation performs a crisp evaluation on each level, where there are crisp representatives of the imprecise properties A and D. Since both logical and arithmetic operations are performed by definition

Table 3. RL-evaluation of sentences in Table 2 (E_1-E_6) and the same sentences with the quantifier $Q(x) = x$ (E_7-E_{12})

α	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	E_{12}
1	0	1	0	1	0	0	0	1	0	1	0	0
0.9	1	0	0	1	0	0	1	0	0	1	0	0
0.8	0	0	0	1	0	0	1/2	1/2	0	1	0	0
0.6	2/3	0	0	1	0	0	2/3	1/3	0	1	0	0
0.5	2/3	0	0	0	0	0	2/3	1/3	1/2	1/2	0	0
0.4	0	0	1	0	0	0	1/2	1/2	1	0	0	0

Table 4. Numerical evaluation of sentences in Table 2 (E_1-E_6) and the same sentences with the quantifier $Q(x) = x$ (E_7-E_{12})

$S(E_1)$	$S(E_2)$	$S(E_3)$	$S(E_4)$	$S(E_5)$	$S(E_6)$	$S(E_7)$	$S(E_8)$	$S(E_9)$	$S(E_{10})$	$S(E_{11})$	$S(E_{12})$
7/30	0.1	0.4	0.5	0	0	1.6/3	1.4/3	0.45	0.55	0	0

in each level independently, all the properties of quantification on crisp data are preserved. Due to the properties of RL-representations, two important properties are verified.

First, the evaluation is efficient in time, since it is a crisp evaluation ($O(1)$) in each level, the number of levels depending on either the amount of data (if we do not fix a precision for the degrees) or the precision considered. In the worst case, corresponding to A and D being fuzzy sets from which the RL-representation (representation as a set of α-cuts) must be obtained, the complexity is $O(nlogn)$, as shown in [4].

But in addition, properties involving the negation of properties are verified. In particular, there is an intuitive property that, to the best of our knowledge, is not verified by any other existing method: the evaluation of the sentences E_5, E_6, E_{11} and E_{12} is definitely 0.

3.3 Numerical Evaluation

The evaluation of quantified sentences yields usually a number in $[0, 1]$. We can obtain such summary of the evaluation when that is the final, expected result of our system; otherwise, following the ideas of RL-representation, we would proceed operating in each level independently.

In this paper we propose to summarize the information given by the RL-evaluation as follows:

Definition 6. *The numerical summary $S(E)$ of a RL-evaluation E defined by (Λ_E, ρ_E) is given by*

$$S(E) = \sum_{\beta \in \Omega_E} m_E(\beta) \cdot \beta \tag{12}$$

Table 4 shows the evaluation of quantified sentences E_1 - E_{12} following definition 6.

Let us remark again that, as the intuition suggests, the final evaluation of sentences E_5, E_6, E_{11} and E_{12} is 0, though any existing method would give a greater value. This is because when representing imprecision by means of fuzzy sets, $A \cap \overline{A} \neq \emptyset$ in general. Using RL-representations, we obtain a coherent result whilst representing properly the imprecision of the properties. Let us remark that, as can be seen in table 1, RL-representations allow to have the same object in the representation of A and ¬ A, though in different levels, and at the same time the representation of $A \wedge \neg A$ is $\emptyset$.

In the particular case when A,D are fuzzy sets, the following propositions are easy to show:

Proposition 1. *Let A,D be fuzzy sets and let E be the sentence "Q of D are A". Then*

$$S(E) = GD_Q(D/A) \tag{13}$$

where $GD_Q(D/A)$ is the evaluation of E as given by the method GD proposed in [4].

Proposition 2. *The following intuitive properties for the evaluation of quantified sentences [1,6] are verified by both the RL-evaluation and the numerical evaluation:*

1. *Correct generalization: in the crisp case, the evaluation of "Q of A are D" yields the expected result.*
2. *External negation: the evaluation of "(¬ Q) of A are D" is equal to 1 minus the evaluation of "Q of A are D"*
3. *Internal negation: the evaluation of "Q of A are D" is equal to the evaluation of "antQ of A are ¬ D", where antQ is the antonym of Q defined as antQ(x)=Q(1-x)*
4. *Duality: the evaluation of "Q of A are D" is equal to the evaluation of "(¬ antQ) of A are ¬ D".*

4 Conclusions

No existing method for evaluating linguistically quantified sentences verifies all the intuitive properties proposed by several authors. In particular, the evaluation of the sentence "∃ of A are ¬ A", that some authors write "Some A are ¬ A" yields always a value greater than 0. In this paper we have proposed alternatives based on the representation of imprecise properties (in particular, of fuzzy sets) via restriction levels [8]. First, RL-evaluation is defined as an imprecise number in [0,1] defined by a RL-number [9]. Then, a summary of that information on the basis of the basic probability assignment associated to each level is proposed, yielding a numerical result as usual. In both cases, the proposed methods verify many intuitive properties, including an evaluation definitely 0 of the sentence "Some A are ¬ A".

Many work remains to be done. First, we shall study the fulfilment of properties expected from the numerical summary of the RL-evaluation, like continuity

and monotonicity with respect to arguments. Let us remark that we expect to show those properties since, as we have shown in this paper, in the specific case of fuzzy sets without negation, the proposed numerical method is equivalent to method GD [4], that verifies these properties. In addition, we shall study syllogisms and reasoning on the basis of our approach. Finally, we plan to apply these results in data mining, quantification in description logics and summarization.

References

1. Barro, S., Bugarín, A., Cariñena, P., Díaz-Hermida, F.: Voting-model based evaluation of fuzzy quantified sentences: A general framework. Fuzzy Sets and Systems 146(1), 97–120 (2004)
2. Bosc, P., Lietard, L.: Monotonic quantified statements and fuzzy integrals. In: Proc. NAFIPS/IFIS/NASA Conference, pp. 8–12 (1994)
3. Cui, L., Li, Y.: Linguistic quantifiers based on choquet integrals. Int. Journal of Approximate Reasoning 48, 559–582 (2008)
4. Delgado, M., Sánchez, D., Vila, M.A.: Fuzzy cardinality based evaluation of quantified sentences. International Journal of Approximate Reasoning 23, 23–66 (2000)
5. Glöckner, I.: DFS - an axiomatic approach to fuzzy quantification. Technical Report TR97-06, Technical Faculty, University Bielefeld, 33501 Bielefeld, Germany (1997)
6. Glöckner, I.: Fuzzy Quantifiers: A Computational Theory. Studies in Fuzziness and Soft Computing. Springer, Heidelberg (2006)
7. Lietard, L., Rocacher, D.: Evaluation of quantified statements using gradual numbers. In: Galindo, J. (ed.) Handbook of Research on Fuzzy Information Processing in Databases, Hershey, PA, USA, pp. 246–269 (2008)
8. Sánchez, D., Delgado, M., Vila, M.A.: A restriction level approach to the representation of imprecise properties. In: IPMU 2008, pp. 153–159 (2008)
9. Sánchez, D., Delgado, M., Vila, M.A.: RL-numbers: An alternative to fuzzy numbers for the representation of imprecise quantities. In: FUZZ-IEEE 2008, pp. 2058–2065 (2008)
10. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on System, Man and Cybernetics 18(1), 183–190 (1988)
11. Ying, M.: Linguistic quantifiers modeled by sugeno integrals. Artificial Intelligence 170, 581–606 (2006)
12. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Computing and Mathematics with Applications 9(1), 149–184 (1983)

Least Squares Method for L-R Fuzzy Variables

Barbara Gładysz and Dorota Kuchta

Institute of Organization and Management, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
{barbara.gladysz,dorota.kuchta}@pwr.wroc.pl

Abstract. The least squares method is used to determine the fuzzy regression. The data for the regression equation are observations for the output and input variables. Analogous assumptions for those used in case of the classical regression are adopted - concerning the fuzzy random component of the model. It is shown how to determine the possibilistic distributions of the output variable and the model coefficients if the random component of the model is an L-R fuzzy variable and its generative probabilistic distribution is known.

Keywords: least squares method, fuzzy regression, L-R fuzzy variable, possibility theory.

1 Introduction

Since Tanaka, being the first author to do so, proposed a fuzzy econometric model [13], many different concepts of constructing fuzzy econometric models have appeared in the literature [8]. Three main streams can be observed. In the first one the authors take, while building the econometric model, fuzzy numbers as observations of the dependent (output) variable, e.g. [4],[12]. In the second concept the input are dependent variable realizations together with the corresponding λ-levels, e.g. [7]. In the third approach the authors, construct the regression using the dependent variable observations from a fixed λ-level, e.g. [9], [10], [11]. We are proposing an approach analogous to the classical regression concept, taking as input for the fuzzy econometric model the observations of the variables, both the independent and the dependent ones. We give an example of the application of the proposed method in the energy load forecasting.

2 Classical Regression Model

In the classical econometric approach the input data for the regression equation construction are observations $(y_t, x_{t1}, \ldots, x_{tk})$, $t = 1, \ldots, T$. The parameters of the linear dependency

$$y_t = \alpha_0 + \alpha_1 x_{t1} + \cdots + \alpha_k x_{tk} + \xi_t \tag{1}$$

where ξ_t is a random variable, are estimated by means of the least squares method (LS) [14]. A linear regression equation is obtained: $\hat{y}_t = a_0 + a_1 x_{t1} +$

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 36–43, 2009.

$\cdots + a_k x_{tk}$, where the coefficients a_j are estimators of the unknown parameters α_j, and $\hat{y}_t$ - estimators of the depended variable y_t.

In the LS method the coefficients a_j are determined in such a way that the sum of the squared deviations of the estimated dependent variable values $\hat{y}_t$ from its actual values y_t is minimal:

$$\sum_{t=1}^{T} u_t^2 = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \tag{2}$$

This function takes on its minimum in point $a = \left(X^T X\right)^{-1} X^T y$, where: y is the vector of dependent variable observations, X - the matrix of observations of the independent variables $x_1, \ldots, x_k$ [14].

The following asumptions are made as far as the random component ξ_t is concerned:

1. The expected value equals zero: $E(\xi_t) = 0$ for $t = 1, \ldots, T$.
2. Identical variance: $V(\xi_t) = \sigma^2$ for $t = 1, \ldots, T$.
3. Independency: ξ_{t_1} and ξ_{t_2} are independent for t_1 different from t_2.
4. Independency from x_j: ξ_t and x_j are independent for all ξ_t and x_j - this assumption follows immediately from the fact that x_j are non-random variables.
5. Normality: all ξ_t have a normal ditribution. This assumption, combined with assumptions 1, 2 and 3 leads to the conclusion that ξ_t have independent normal distributions with expected value equal zero and a variance σ^2 - $N(0, \sigma)$.

From assumptions 1 - 5 it follows [14] that estimators of coefficients α_j have normal distributions with expected values $a = \left(X^T X\right)^{-1} X^T y$ and variances $\sigma^2 d_{00}, \ldots, \sigma^2 d_{kk}$, where d_{jj} are the diagonal elements of matrix $\left(X^T X\right)^{-1}$, $j = 0, \ldots, k$.

The dependend variable y_t is a random variable with normal distribution $N(\alpha_0 + \alpha_1 x_{t1} + \cdots + \alpha_k x_{tk}, \sigma)$.

The distribution of the estimator $\hat{y}_t$ of dependent variable y_t is a normal distribution with the expected value and variance respectively $\alpha_0 + \alpha_1 x_{t1} + \cdots + \alpha_k x_{tk}$, $\sigma^2 \left(1 + x_0^T \left(X^T X\right)^{-1} x_0\right)$.

The estimator of the expected value of the random component is $\frac{\sum_{t=1}^{T} u_t}{T}$, and the estimator of the random component variance is $\frac{\sum_{t=1}^{T} u_t^2}{T-k-1}$.

3 Fuzzy Regression Model

3.1 Basic Notion

Let X be a single valued variable, whose value is not precisely known. There is given a normal, quasi concave and upper semicontinuos function $\mu_X : \Re \rightarrow$

$[0,1]$ called the possibility distribution for X. The value of $\mu_X(x)$ for $x \in \Re$ denotes the possibility of the event that X takes the value of x, i.e. $\mu_X(x) = Pos(X = x)$, [5], [16]. Such a variable X will be called a fuzzy variable. An L-R fuzzy variable is such a fuzzy variable that such $\underline{m}_X, \overline{m}_X$ and functions L, R (continuous, symmetric, attaining value 1 for argument 0 and increasing in the domain of non-negative numbers) exist that its possibility distribution has the following form:

$$\mu_X(x) = \begin{cases} L\left(\left(\underline{m}_X - x\right)/\alpha_X\right) & \text{for } x < \underline{m}_X \\ 1 & \text{for } \underline{m}_X \leq x \leq \overline{m}_X \\ R\left(\left(x - \overline{m}_X\right)/\beta_X\right) & \text{for } x > \overline{m}_X \end{cases} \tag{3}$$

Functions L and R determine the shape of the corresponding possibility distribution. The shape of possibility distribution (3) is preserved in linear transformations. A triangular fuzzy number $X = (m_X, \alpha_X, \beta_X)$ is such an L-R fuzzy number for which functions L and R are linear functions and $\underline{m}_X = \overline{m}_X = m_X$.

For a given fuzzy variable X and a given λ, the λ -level is defined as the closed interval $\left[\underline{x}(\lambda), \overline{x}(\lambda)\right] = \{x : \mu_X(x) \geq \lambda\}$.

Consider two fuzzy variables X, Y with possibility distributions μ_X, μ_Y respectively. The possibility distribution of fuzzy variables $Z = X+Y$ and $U = XY$ is defined by means of Zadeh extension principle respectively as follows [15]

$$\mu_Z(z) = sup_{z=x+y} min\left(\mu_X(x), \mu_Y(y)\right) \tag{4}$$

$$\mu_U(u) = sup_{u=x \cdot y} min\left(\mu_X(x), \mu_Y(y)\right). \tag{5}$$

We are interested in comparing X to Y, i.e. we want to characterize the possibility of the event that the value taken by X will be equal to the value taken by Y. It can be done by the following index [5], [6]:

$$Pos(X = Y) = min\left(Pos(X \geq Y), Pos(Y \geq X)\right), \tag{6}$$

where $Pos(X \geq Y) = sup_{x \geq y} min\left(\mu_X(x), \mu_Y(y)\right)$.

In the literature there are definitions of the mean and the variance of a fuzzy variable which follow from so called generative probability distributions of fuzzy variables [2], [3]. Carlsson and Fullér [2] define crisp possibilistic mean value and the crisp possibilistic variance as of a fuzzy variable X as

$$E(X) = \frac{1}{2} \int_0^1 \left(\underline{x}(\lambda) + \overline{x}(\lambda)\right) d\lambda \tag{7}$$

$$V(X) = \frac{1}{3} \int_0^1 \left[\left(\underline{x}(\lambda)\right)^2 + \underline{x}(\lambda)\,\overline{x}(\lambda) + \left(\overline{x}(\lambda)\right)^2\right] d\lambda + \left[\frac{1}{2} \int_0^1 \left(\underline{x}(\lambda) + \overline{x}(\lambda)\right) d\lambda\right]^2 \tag{8}$$

From the probabilistic view point, the possibilistic mean and variance of a fuzzy variable X can be viewed as the expected values of, respectively, the conditional mean value and variance under a given λ, of the random variable taking on values $\underline{x}(\lambda)$ and $\overline{x}(\lambda)$ with probability 0,5, if we treat the λ-level as a random variable having a Beta distribution Beta(2,1) (λ being the density of the distribution).

Chanas and Nowakowski [3] proposed a generative expected value and variance of a fuzzy variable corresponding to a compound random variable, where both the λ - level and the values from it are generated according to the uniform distribution.

$$E(X) = \frac{1}{2}\int_0^1 \left(\underline{x}(\lambda) + \overline{x}(\lambda)\right) d\lambda \tag{9}$$

$$V(X) = \frac{1}{3}\int_0^1 \left[\left(\underline{x}(\lambda)\right)^2 + \underline{x}(\lambda)\,\overline{x}(\lambda) + \left(\overline{x}(\lambda)\right)^2\right] d\lambda + \left[\frac{1}{2}\int_0^1 \left(\underline{x}(\lambda) + \overline{x}(\lambda)\right) d\lambda\right]^2 \tag{10}$$

where λ is the uniform random variable over $(0, 1]$.

The expected value and variance for fuzzy variables have the following properties [2], [3]:

$$\begin{array}{ll} E(aX) & = aE(X) \\ E(X+Y) & = E(X) + E(Y) \\ V(aX) & = a^2 V(X) \\ V(X+Y) & = V(X) + V(Y) \text{- if } X, Y \text{ are independent fuzzy variables.} \end{array} \tag{11}$$

For a triangular fuzzy variable $X = (m_X, \alpha_X, \beta_X)$ the expected value both according to the Carlsson and Fullér, the Chanas and Nowakowski definitions equals $E(X) = m_X + \frac{\beta_X - \alpha_X}{4}$. It is the Steiner point of a triangular fuzzy variable. The variance according to Carlsson and Fullér equals $V(X) = \frac{\alpha_X + \beta_X}{12}$, and according to Chanas and Nowakowski $V(X) = \frac{7\alpha_X^2 + 7\beta_X^2 + 2\alpha_X\beta_X}{144}$.

Zadeh [16] proposed the following indicator showing to which degree a possibility distribution fits a probability distribution having the density function $f(x)$: $\int_0^1 \mu(x) f(x)\, dx$. The higher the value of the indicator, to the greater degree the possibility distribution fits the probability distribution. If we are choosing the probability to generate the values of a triangular symmetric fuzzy variable $X = (m_X, \alpha_X, \alpha_X)$ from the class of normal distributions $N(m, \sigma)$, the highest fitness is achieved when the following equations are satisfied:

$$m_X = m \tag{12}$$

$$exp\left(\frac{\alpha_X^2}{2\sigma^2}\right) + \frac{\alpha_X^2}{2\sigma^2} + \alpha_X^2 - 1 = 0 \tag{13}$$

3.2 Least Squares Method for Fuzzy Variables

Let us assume again that we have observations $(y_t, x_{t1}, \ldots, x_{tk})$, $t = 1, \ldots, T$ and suppose that in (1) ξ_t are fuzzy variables. Similarly as in case of the classical regression, function S attains its minimum in point $a = \left(X^T X\right)^{-1} X^T y$.

Let us make assumptions 1-3 analogous to those in the classical case (chapter 2) about the random component ξ_t. Instead of assumpions 4 and 5, let us now assume that:

- 4' Independency from x_j: ξ_t and x_j are independent for all ξ_t and x_j - this assumption follows from the fact that x_j are non-fuzzy variables.
- 5'. All ξ_t have a L-R possibility distribution of the same membership function shape, e.g. triangular. This assumption, combined with assumptions 1, 2 and 3 leads to the conclusion that ξ_t have independent L-R possibility distributions of the same membership function shape with expected value equal zero and a variance σ^2.

Knowing the realisations u_t of the fuzzy random component, we can determine the estimators of the mean value and the variance of the model error: respectively $\frac{\sum_{t=1}^{T} u_t}{T}$, $\frac{\sum_{t=1}^{T} u_t^2}{T-k-1}$. Further, making use of (7) - (8) or (9) - (10) or (12) depending on the form of the probabilistic distribution of the random component, we can determine the parameters of the fuzzy variable if it is a symmetric triangular fuzzy variable $\xi = (m_\xi, \alpha_\xi, \alpha_\xi)$. For such a fuzzy variable for example from assumption 1 and (9) - (10) we will get the following equations system:

$$E(\xi) = 0, \;\; V(\xi) = \frac{\alpha_\xi^2}{9} \tag{14}$$

Thus for a symmetric triangular fuzzy number knowing the variance and the expected value allows to determine the membership function shape. If the random component of the model is not a symmetric triangular fuzzy variable, we can determine from the sample $u_1, \ldots, u_T$ the other parameters of the fuzzy variable, e.g. the median [1], obtaining an equations system which allows to calculate the respective possibilistic distribution parameters. If we get an infeasible system, it means that we have incorrectly estimated the shape of the fuzzy variable possibilistic distribution.

From assumptions 1 - 3, 4' - 5' and from (11) and taking into account that L-R fuzzy variable membership function shape is preserved in linear transformations it follows that:

- The estimators of coefficients α_j are fuzzy variables with the same membership function shape as ξ_t with expected values $a = \left(X^T X\right)^{-1} X^T y$ and variances $\sigma^2 d_{00}, \ldots, \sigma^2 d_{kk}$ $(j = 0, \ldots, k)$.
- The possibility distribution of the estimator $\hat{y}_t$ of the dependent variable y_t is of the same type (e.g. have the same memership function shape) as distribution of ξ_t with the expected value $\hat{y}_t = a_0 + a_1 x_{t1} + \cdots + a_k x_{tk}$ and variance $\sigma^2 \left(1 + x_0^T \left(X^T X\right)^{-1} x_0\right)$.

In order to verify whether individual fuzzy coefficients of model α_j $(j = 0, \ldots, k)$ are significiant we will make use of relation (6). We will assume that α_j is a significant if $Pos(\alpha_j = 0) \leq \lambda_0$. The value of λ_0 is selected arbitralily, like the significante level in statistic tests.

4 Example 1

The regression has been constructed for the energy load at midday for a power region in Poland. The data for the analysis are hour observations of the energy

load and air temperature throughout four years (January 1998 - September 2001). The observations from September 2001 were not taken into account while constructing the econometric model, they were used to verify the forecast.Using the least squares method, the following model was obtained:

$$\hat{L12}_t = 60565.91 + 0.9L7_t - 0.15L7_{t-1} + 0.1L12_{t-1}, \tag{15}$$

where Lh_t - energy load at the hour h at the day t.

Using statistical test we have verified that the random component of this model has a distribution where both the λ - level and the values form it are generated according to the uniform distribution with the expected value zero and the variance 1229309764. According to (14) we can determine the possibilistic distribution of the estimators of the fuzzy variables the possibilistic distributions of the coefficients were determined: $\alpha_0 = (60565.91, 10644.55, 10644.55)$, $\alpha_{L7_t} = (0.90, 0.06, 0.06)$, $\alpha_{L7_{t-1}} = (-0.15, 0.11, 0.11)$, $\alpha_{L12_{t-1}} = (0.10, 0.09, 0.09)$. Also the significance of the model coefficients has been verified: $Pos(\alpha_j = 0) = 0$ for $j = 0, 1, 2, 3$. The fuzzy autoregression model of the energy load (15) is thus a model, in which all the variable are significant. The determination coefficient of this model equals to $R^2 = (\hat{y}_t - \bar{y})^2 / (y_t - \bar{y})^2 = 0.77$. We can thus use it in forecasting. In Table 1 we show the possibilistic distribution of the energy load, determined according to (9),(10) for individual September days, as well as the actual values of the energy load $L12_t$ and the corresponding possibility, that the energy load according to model (15) takes on such a value - $Pos(\hat{L12}_t = L12_t)$. The lowest value of $Pos(\hat{L12}_t = L12_t)$ equals $0,48$. The model has thus proved to be useful in forecasting.

Table 1. Observations and possibilistic distributions of the energy load for September

Day	Day of the week	$L12_t$	$\hat{L12}_t$	$Pos\left(\hat{L12}_t = L12_t\right)$
1	Saturday	227059.8	(205679 , 40983.6 , 40983.6)	0.48
2	Sunday	190561.8	(194703 , 41025.3 , 41025.3)	0.90
3	Monday	218563.4	(217211 , 41014.8 , 41014.8)	0.97
4	Tuesday	219109.0	(219221 , 40970.1 , 40970.1)	1.00
5	Wednsday	221819.4	(222233 , 40966.8 , 40966.8)	0.99
6	Thursday	225335.0	(223303 , 40964.4 , 40964.4)	0.95
7	Friday	232474.0	(224943 , 40963.5 , 40963.5)	0.82

5 Example 2

Let us now consider the problem analyzed by Tanaka [13]. He investigated the dependency between y=prefabricated house sale price in Japan and x_1=quality of the construction material (1 - low, 2 - medium, 3 - high), x_2=area of the first floor, x_3=area of the second floor, x_4=total number of room, x_5=number of Japanese room, Table2. He obtained the following form of fuzzy regression: $\hat{y} = (0, 0, 0) + (245.2, 17.26, 17.26)\, x_1 + (5.85, 0, 0)\, x_2 + (4.79, 0, 0)\, x_3 + (0, 0, 0)\, x_4 + (0, 0, 0)\, x_5$.

Using the least squares method proposed in the paper, the following model was obtained $\hat{y} = -107.5 + 236.5x_1 + 9.4x_2 + 8.2x_3 - 38x_4 - 18.4x_5$. The residuals probabilistic distribution of this model is a normal one: $N(0, 39.7)$. Thus the possibilistic distributions of fuzzy coefficients of the fuzzy regression were determined according to formulae (12)-(13). The model constant and the coefficients at variables x_4, x_5 turned out to be insignificant. Variables x_4 and x_5 were thus excluded from the model and it was assumed that the constant is equal to zero. Consequently, according to (2) the following model was obtained:

$$\hat{y} = 270x_1 + 5.7x_2 + 3.9x_3. \tag{16}$$

The determination coefficient for this model equals 0.99, the model residuals have the normal distribution $N(0, 63.67)$. Using formulae (12)-(13) again, the possibilistic distributions of the coefficients were determined: $\alpha_1 = (270, 124, 124)$, $\alpha_2 = (5.7, 0, 0)$, $\alpha_3 = (3.9, 0, 0)$. Also the significance of the model coefficients has been verified: $Pos(\alpha_j = 0) = 0$ for $j = 1, 2, 3$. The obtained model includes the same variables as in the Tanaka model. The possibilistic distributions of the coefficients of both models are similar. Let us now compare the possibilistic dis-

Table 2. Data and possibilistic distributions of house prices

						Tanaka			LS		
y	x_1	x_2	x_3	x_4	x_5	$\hat{Y}$	$\alpha_{\hat{Y}}$	$Pos\left(\hat{Y} = y\right)$	$\hat{Y}$	$\alpha_{\hat{Y}}$	$Pos\left(\hat{Y} = y\right)$
606	1	38.09	36.43	5	1	642.5	75.3	0.52	627.8	225.6	0.90
710	1	62.10	25.50	6	1	730.6	75.3	0.73	722.4	217.9	0.94
808	1	63.76	44.71	7	1	832.3	75.3	0.68	806.1	215.7	0.99
826	1	74.52	38.09	8	1	863.6	75.3	0.50	841.9	218.1	0.93
865	1	75.38	41.10	7	2	883.0	75.3	0.76	858.4	220.2	0.97
852	2	52.99	26.49	4	2	927.2	150.5	0.50	944.2	225.5	0.59
917	2	62.93	26.49	5	2	985.4	150.5	0.55	1000.9	210.0	0.60
1031	2	72.04	33.12	6	3	1070.4	150.5	0.74	1078.4	202.1	0.77
1092	2	76.12	43.06	7	2	1141.9	150.5	0.67	1140.1	203.6	0.76
1203	2	90.26	42.64	7	2	1222.6	150.5	0.87	1219.1	209.2	0.92
1394	3	85.70	31.33	6	3	1386.9	225.8	0.97	1419.3	214.7	0.88
1420	3	95.27	27.64	6	3	1425.2	225.8	0.98	1459.6	215.3	0.82
1601	3	105.98	27.64	6	3	1487.9	225.8	0.50	1520.6	219.4	0.63
1632	3	79.25	66.81	6	3	1519.1	225.8	0.50	1519.6	268.2	0.58
1699	3	120.50	32.25	6	3	1594.9	225.8	0.54	1621.2	233.0	0.67

tributions of the house prices determined using both models, Table 2. In model determined using the least squares method the width ($\alpha_{\hat{Y}}$) of the possibilistic distribution of the houses prices is significantly higher in case of houses built with low grade materials than the corresponding width of the fuzzy price approximated according to the Tanaka model. In the model obtained using the least squares method the smallest value $Pos(\hat{Y} = y)$ equals 0.58 and in the Tanaka model 0.5.

6 Conclusions

We have proposed an application of the least squares method to determining fuzzy regression based on observations of output and input variables. We have presented the method using two examples. The first one is an autoregressive model of the energy load in Poland. The second one describes the dependency between the houses prices in Japan and their surface and the materials quality.

References

1. Bodjanova, S.: Median value and median interval of a fuzzy number. Information Sciences 172, 73–89 (2005)
2. Carlsson, C., Fullér, R.: On possibilistic mean value and variance of fuzzy numbers. Fuzzy Sets and Systems 122, 315–326 (2001)
3. Chanas, S., Nowakowski, M.: Single value simulation of fuzzy variable. Fuzzy Sets and Systems 25, 43–57 (1988)
4. Diamond, P.: Least squares and maximum likelihood regression for fuzzy linear models. In: Kacprzyk, J., Fedrizzi, M. (eds.) Fuzzy Regression Analysis, pp. 137–151. Omnitech Press Warsaw, Physica-Verlag, Heidelberg, Warsaw (1992)
5. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1988)
6. Dubois, D., Prade, H.: The Use of Fuzzy Number in Decision Analysis, Fuzzy Information and Decision Processes, pp. 309–321. North-Holland Publishing Company, Amsterdam (1982)
7. Guo, P., Tanaka, H.: Dual models for possibilistic regression analysis. Computational Statistics and Data Analysis 51, 253–266 (2006)
8. Kacprzyk, J., Fedrizzi, M.: Fuzzy Regression Analysis. Omnitech Press Warsaw, Physica-Verlag, Heidelberg, Warsaw (1992)
9. Nasrabadi, M.M., Nasrabadi, E., Nasrabadi, A.R.: Fuzzy linear regression analysis: A multi-objective programming approach. Applied Mathematics and Computation 163, 245–251 (2005)
10. Özelkan, E.C., Duckstein, L.: Multi-objective fuzzy regression: a general framework. Computers and Operations Research 27, 635–652 (2000)
11. Sakawa, M., Yano, H.: Multiobjective fuzzy linear regression analysis for fuzzy input-output data. Fuzzy Sets and Systems 47, 173–181 (1992)
12. D'Urso, P., Gastaldi, T.: An "ordewise"' polynomial regression procedure for fuzzy data. Fuzzy Sets and Systems 130, 1–19 (2002)
13. Tanaka, H.: Fuzzy data analysis by possibilistic linear models. Fuzzy Sets and Systems 24, 363–375 (1987)
14. Theil, H.: Principles of Econometrics. John Wiley, New York (1987)
15. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
16. Zadeh, L.A.: Fuzzy sets as a basis of theory of possibility. Fuzzy Sets and Systems 1, 3–29 (1978)

Modeling Interpretive Steps in Fuzzy Logic Computations*

Pedro J. Morcillo and Ginés Moreno

University of Castilla-La Mancha
Department of Computing Systems
02071, Albacete, Spain
{pmorcillo,gmoreno}@dsi.uclm.es

Abstract. Fuzzy logic programming is a growing declarative paradigm aiming to integrate fuzzy logic into logic programming (LP). In this setting, the multi-adjoint logic approach represents an extremely flexible fuzzy language with a procedural principle structured in two separate phases. During the operational one, *admissible steps* are systematically applied in a similar way to classical resolution steps in LP, thus returning an expression where all atoms have been exploited. This last expression is then interpreted under a given lattice during the so called interpretive phase. Whereas the operational phase has been successfully formalized in the past, more effort is needed to clarify the notion of *interpretive step*. In this paper we firstly introduce a refinement of this concept which fairly models at a very low level the computational behaviour of the interpretive phase. Then, we present a simple but powerful cost measure induced from such definition which helps to estimate the computational (interpretive) effort required to solve a goal. The resulting method is much more accurate and realistic than other simpler cost measures (like counting the number or the *weights* of interpretive steps) that we have proposed in the past for proving efficiency properties in program transformation tasks such as fold/unfold, partial evaluation, and so on.

Keywords: Fuzzy Logic Programming, Cost Measures, Aggregators.

1 Introduction

Fuzzy Logic Programming [1,2,4,8,13] is an interesting and still growing research area that agglutinates the efforts for introducing fuzzy logic into *Logic Programming* [9], in order to provide techniques and constructs for dealing with uncertainty and approximated reasoning in a natural way. Most of these systems replace the classical inference mechanism of SLD–Resolution with a fuzzy variant which is able to handle partial truth. This is the case of *Multi-adjoint logic programming* [11,10] where a program can be seen as a set of rules each one annotated by a truth degree and a goal is a query to the system plus a substitution (initially the empty substitution, denoted by *id*). *Admissible steps*

* This work was supported by the EU (FEDER), and the Spanish Science and Education Ministry (MEC) under grants TIN 2004-07943-C04-03 and TIN 2007-65749.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 44–51, 2009.

(a generalization of the classical *modus ponens* inference rule) are systematically applied on goals in a similar way to classical resolution steps in pure LP, thus returning an state composed by a computed substitution together with an expression where all atoms have been exploited. Next, this expression is interpreted under a given lattice, hence returning a pair $\langle$*truth degree*; *substitution*$\rangle$ which is the fuzzy counterpart of the classical notion of computed answer used in pure LP. As we showed in [6], this last interpretive process admits a formulation based on a transition system where each *interpretive step* solves a concrete connective of an state. In this paper, we improve such definition by explicitly expanding connective definitions and evaluating primitive (arithmetic) operators on states. The method does not alter the final set of solutions, but it has the extra ability of exhibiting the complexity of the interpretive phase in detail.

In connection with the last statement, we are specially interested in to accurately observe the behaviour of computations performed on programs obtained via some program transformation techniques developed in our group [5,3]. When analyzing efficiency, it is convenient to define abstract approaches to cost measurement, such as the usual method of counting the number of derivation steps required to reach a solution. In [7] we showed that, in the context of multi-adjoint logic programming, this method was inappropriate when considering the interpretive phase. The problem was partially solved there by defining a more refined interpretive cost measure based on the weights (that is, the number of connectives and primitive operators) of the connectives evaluated in each interpretive step of a given derivation. Anyway, such approach fails when dealing with complex connective definitions which possible might emerge at transformation time. By facing the problem via *small interpretive steps*, in Section 3 we definitively overcome these inconveniences in a natural and accurate way.

2 Multi-adjoint Logic Programming

This section summarizes the main features of multi-adjoint logic programming (see [11] for a complete formulation of this framework). We work with a first order language, $\mathcal{L}$, containing variables, constants, function symbols, predicate symbols, and several (arbitrary) connectives to increase language expressiveness: implication connectives $(\leftarrow_1, \leftarrow_2, \ldots)$; conjunctive operators (denoted by $\&_1, \&_2, \ldots$), disjunctive operators $(\vee_1, \vee_2, \ldots)$, and hybrid operators (usually denoted by $@_1, @_2, \ldots$), all of them are grouped under the name of "aggregators". Although these connectives are binary operators, we usually generalize them as functions with an arbitrary number of arguments. So, we often write $@(x_1, \ldots, x_n)$ instead of $@(x_1, \ldots, @(x_{n-1}, x_n), \ldots)$. By definition, the truth function for an n-ary aggregation operator $[\![@]\!] : L^n \to L$ is required to be monotonous and fulfills $[\![@]\!](\top, \ldots, \top) = \top$, $[\![@]\!](\bot, \ldots, \bot) = \bot$.

Additionally, our language $\mathcal{L}$ contains the values of a multi-adjoint lattice, $\langle L, \preceq, \leftarrow_1, \&_1, \ldots, \leftarrow_n, \&_n\rangle$, equipped with a collection of adjoint pairs $\langle \leftarrow_i, \&_i\rangle$, where each $\&_i$ is a conjunctor which is intended to the evaluation of *modus ponens* [11]. In general, L may be the carrier of any complete bounded lattice but, for readability reasons, in the examples we shall select L as the set of real numbers

in the interval $[0, 1]$. A L-expression is a well-formed expression composed by values and connectives of L, as well as variable symbols and *primitive operators* (i.e., arithmetic symbols such as $*, +, min$, etc...). In what follows, we assume that the truth function of any connective @ in L is given by its corresponding *connective definition*, that is, an equation of the form $@(x_1, \ldots, x_n) \triangleq E$, where E is a L-expression not containing variable symbols apart from $x_1, \ldots, x_n$.

A *rule* is a formula $H \leftarrow_i \mathcal{B}$, where H is an atomic formula (usually called the *head*) and $\mathcal{B}$ (which is called the *body*) is a formula built from atomic formulas $B_1, \ldots, B_n$ — $n \geq 0$ —, truth values of L, conjunctions, disjunctions and aggregations. A *goal* is a body submitted as a query to the system. Roughly speaking, a multi-adjoint logic program is a set of pairs $\langle \mathcal{R}; \alpha \rangle$ (we often write $\mathcal{R}$ *with* α), where $\mathcal{R}$ is a rule and α is a *truth degree* (a value of L) expressing the confidence of a programmer in the truth of the rule $\mathcal{R}$. By abuse of language, we sometimes refer a tuple $\langle \mathcal{R}; \alpha \rangle$ as a "rule".

The procedural semantics of the multi–adjoint logic language $\mathcal{L}$ can be thought as an operational phase (based on admissible steps) followed by an interpretive one. In the following, $\mathcal{C}[A]$ denotes a formula where A is a sub-expression which occurs in the –possibly empty– context $\mathcal{C}[]$. Moreover, $\mathcal{C}[A/A']$ means the replacement of A by A' in context $\mathcal{C}[]$, whereas $\mathcal{V}ar(s)$ refers to the set of distinct variables occurring in the syntactic object s, and $\theta[\mathcal{V}ar(s)]$ denotes the substitution obtained from θ by restricting its domain to $\mathcal{V}ar(s)$.

Definition 1 (Admissible Step). *Let $\mathcal{Q}$ be a goal and let σ be a substitution. The pair $\langle \mathcal{Q}; \sigma \rangle$ is a* state *and we denote by $\mathcal{E}$ the set of states. Given a program $\mathcal{P}$, an* admissible computation *is formalized as a state transition system, whose transition relation $\rightarrow_{AS} \subseteq (\mathcal{E} \times \mathcal{E})$ is the smallest relation satisfying the following* admissible rules *(where we always consider that A is the selected atom in $\mathcal{Q}$ and $mgu(E)$ denotes the* most general unifier *of an equation set E):*

1) $\langle \mathcal{Q}[A]; \sigma \rangle \rightarrow_{AS} \langle (\mathcal{Q}[A/v \,\&_i\, \mathcal{B}])\theta; \sigma\theta \rangle$ *if* $\theta = mgu(\{A' = A\})$, $\langle A' \leftarrow_i \mathcal{B}; v \rangle$ *in* $\mathcal{P}$ *and* $\mathcal{B}$ *is not empty.*
2) $\langle \mathcal{Q}[A]; \sigma \rangle \rightarrow_{AS} \langle (\mathcal{Q}[A/v])\theta; \sigma\theta \rangle$ *if* $\theta = mgu(\{A' = A\})$, *and* $\langle A' \leftarrow_i ; v \rangle$ *in* $\mathcal{P}$.

As usual, rules are taken renamed apart. Symbols $\rightarrow_{AS1}$ and $\rightarrow_{AS2}$ are used to distinguish between computation steps performed by the specific admissible rules. The application of a rule on a step will appear as a superscript of $\rightarrow_{AS}$.

Definition 2. *Let $\mathcal{P}$ be a program and let $\mathcal{Q}$ be a goal. An* admissible derivation *is a sequence $\langle \mathcal{Q}; id \rangle \rightarrow^*_{AS} \langle \mathcal{Q}'; \theta \rangle$. When $\mathcal{Q}'$ is a formula not containing atoms (i.e., a L-expression), the pair $\langle \mathcal{Q}'; \sigma \rangle$, where $\sigma = \theta[\mathcal{V}ar(\mathcal{Q})]$, is called an* admissible computed answer *(a.c.a.) for that derivation.*

Example 1. Let $\mathcal{P}$ be the following multi-adjoint logic program:

$\mathcal{R}_1 : p(X) \leftarrow_{\mathtt{P}} \&_{\mathtt{G}}(\vee_{\mathtt{L}}(q(X), 0.6), r(X))$ with 0.9
$\mathcal{R}_2 : q(a) \leftarrow$ with 0.8 $\qquad$ $\mathcal{R}_3 : r(X) \leftarrow$ with 0.7

where the labels L, G and P mean respectively for *Łukasiewicz logic*, *Gödel intuitionistic logic* and *product logic*, that is, $\vee_{\mathtt{L}}(x_1, x_2) \triangleq min(1, x_1 + x_2)$, $\&_{\mathtt{G}}(x_1, x_2) \triangleq min(x_1, x_2)$ and $\&_{\mathtt{P}}(x_1, x_2) \triangleq x_1 * x_2$. Now, we can generate the

following admissible derivation (we underline the selected atom in each step):
$\langle \underline{p(X)};\ id\rangle \rightarrow_{AS1}{}^{\mathcal{R}_1} \langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\vee_{\mathtt{L}}(\underline{q(X1)}, 0.6), r(X1)));\{X/X_1\}\rangle$
$\rightarrow_{AS2}{}^{\mathcal{R}_2} \langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\vee_{\mathtt{L}}(0.8, 0.6), \underline{r(a)}));\{X/a, X_1/a\}\rangle$
$\rightarrow_{AS2}{}^{\mathcal{R}_3} \langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\vee_{\mathtt{L}}(0.8, 0.6), \overline{0.7}));\{X/a, X_1/a, X2/a\}\rangle$
Here, the a.c.a. is the pair: $\langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\vee_{\mathtt{L}}(0.8, 0.6), 0.7)); \theta\rangle$, where $\theta = \{X/a, X_1/a, X2/a\}[\mathcal{V}ar(p(X))] = \{X/a\}$.

If we exploit all atoms of a goal, by applying admissible steps as much as needed during the operational phase, then it becomes a formula with no atoms (a L-expression) which can be then directly interpreted w.r.t. lattice L by applying the following definition we initially presented in [6]:

Definition 3 (Interpretive Step). *Let $\mathcal{P}$ be a program, $\mathcal{Q}$ a goal and σ a substitution. We formalize the notion of* interpretive computation *as a state transition system, whose transition relation $\rightarrow_{IS} \subseteq (\mathcal{E} \times \mathcal{E})$ is defined as the least one satisfying: $\langle Q[@(r_1, \ldots, r_n)]; \sigma\rangle \rightarrow_{IS} \langle Q[@(r_1, \ldots, r_n)/[\![@]\!](r_1, \ldots, r_n)]; \sigma\rangle$, where $[\![@]\!]$ is the truth function of connective $@$ in the lattice $\langle L, \preceq\rangle$ associated to $\mathcal{P}$.*

Definition 4. *Let $\mathcal{P}$ be a program and $\langle Q; \sigma\rangle$ an a.c.a., that is, $\mathcal{Q}$ is a goal not containing atoms (i.e., a L-expression). An* interpretive derivation *is a sequence $\langle Q; \sigma\rangle \rightarrow_{IS}^{*} \langle Q'; \sigma\rangle$. When $Q' = r \in L$, being $\langle L, \preceq\rangle$ the lattice associated to $\mathcal{P}$, the state $\langle r; \sigma\rangle$ is called a* fuzzy computed answer *(f.c.a.) for that derivation.*

Example 2. If we complete the previous derivation of Example 1 by applying 3 interpretive steps in order to obtain the final f.c.a. $\langle 0.63; \{X/a\}\rangle$, we generate the following interpretive derivation D_1: $\langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\underline{\vee_{\mathtt{L}}(0.8, 0.6)}, 0.7)); \theta\rangle \rightarrow_{IS}$ $\langle \&_{\mathtt{P}}(0.9, \underline{\&_{\mathtt{G}}(1, 0.7)}); \theta\rangle \rightarrow_{IS} \langle \underline{\&_{\mathtt{P}}(0.9, 0, 7)}; \theta\rangle \rightarrow_{IS} \langle 0.63; \theta\rangle$.

3 Small Interpretive Steps and Cost Measures

A classical, simple way for estimating the computational cost required to built a derivation, consists in counting the number of computational steps performed on it. So, given a derivation D, we define its:

- *operational cost*, $\mathcal{O}_c(D)$, as the number of admissible steps performed in D.
- *interpretive cost*, $\mathcal{I}_c(D)$, as the number of interpretive steps done in D.

Note the operational and interpretive costs of derivation D_1 performed in the previous section are $\mathcal{O}_c(D_1) = 3$ and $\mathcal{I}_c(D_1) = 3$, respectively. Intuitively, $\mathcal{O}_c$ informs us about the number of atoms exploited along a derivation. Similarly, $\mathcal{I}_c$ seems to estimate the number of connectives evaluated in a derivation. However, this last statement is not completely true: $\mathcal{I}_c$ only takes into account those connectives appearing in the bodies of program rules which are replicated on states of the derivation, but no those connectives recursively *nested* in the definition of other connectives. The following example highlights this fact.

Example 3. A simplified version of rule $\mathcal{R}_1$, whose body only contains an aggregator symbol is $\mathcal{R}_1^{*}$: $p(X) \leftarrow_{\mathtt{P}} @(q(X), r(X))$ with 0.9, where $@(x_1, x_2) \triangleq$

$\&_{\mathtt{G}}(\vee_{\mathtt{L}}(x_1, 0.6), x_2)$. Note that $\mathcal{R}_1^*$ has exactly the same meaning (interpretation) that $\mathcal{R}_1$, although different syntax. In fact, both of them have the same sequence of atoms in their head and bodies. The differences are regarding the set of connectives which explicitly appear in their bodies since in $\mathcal{R}_1^*$ we have moved $\&_{\mathtt{G}}$ and $\vee_{\mathtt{L}}$ (as well as value 0.6) from the body of the rule (see $\mathcal{R}_1$) to the connective definition of @. Now, we use rule $\mathcal{R}_1^*$ instead of $\mathcal{R}_1$ for generating the following derivation D_1^* which returns exactly the same f.c.a that D_1:

$$\begin{aligned}
\langle \underline{p(X)};\ id\rangle &\rightarrow_{AS1}{}^{\mathcal{R}_1^*} \langle \&_{\mathtt{P}}(0.9, @(\underline{q(X1)}, r(X1)); \{X/X_1\}\rangle \\
&\rightarrow_{AS2}{}^{\mathcal{R}_2} \langle \&_{\mathtt{P}}(0.9, @(0.8, \underline{r(a)})); \{X/a, X_1/a\}\rangle \\
&\rightarrow_{AS2}{}^{\mathcal{R}_3} \langle \&_{\mathtt{P}}(0.9, \underline{@(0.8, 0.7)}); \{X/a, X_1/a, X2/a\}\rangle \\
&\rightarrow_{IS} \quad \langle \underline{\&_{\mathtt{P}}(0.9, 0.7)}; \{X/a, X_1/a, X2/a\}\rangle \\
&\rightarrow_{IS} \quad \langle \overline{0.63; \{X/a}, X_1/a, X2/a\}\rangle
\end{aligned}$$

Note that, since we have exploited the same atoms with the same rules (except for the first steps performed with rules $\mathcal{R}_1$ and $\mathcal{R}_1^*$, respectively) in both derivations, then $\mathcal{O}_c(D_1) = \mathcal{O}_c(D_1^*) = 3$. However, although connectives $\&_G$ and $\vee_L$ have been evaluated in both derivations, in D_1^* such evaluations have not been explicitly counted as interpretive steps, and consequently they have not been added to increase the interpretive cost measure $\mathcal{I}_c$. This unrealistic situation is reflected by the abnormal result: $\mathcal{I}_c(D_1) = 3 > 2 = \mathcal{I}_c(D_1^*)$.

In order to solve this problem, we opt here for redefining the behaviour of the interpretive phase in a more accurate way than [6], as follows.

Definition 5 (Small Interpretive Step). *Let $\mathcal{P}$ be a program, $\mathcal{Q}$ a goal and σ a substitution. Assume that the (non interpreted yet) L-expression $\Omega(r_1, \ldots, r_n)$ occurs in $\mathcal{Q}$, where Ω is just a primitive operator or a connective defined in the lattice $\langle L, \preceq\rangle$ associated to $\mathcal{P}$, and $r_1, \ldots, r_n$ are elements of L. We formalize the notion of* small interpretive computation *as a state transition system, whose transition relation $\rightarrow_{SIS} \subseteq (\mathcal{E} \times \mathcal{E})$ is the smallest relation satisfying the following* small interpretive rules *(where we always consider that $\Omega(r_1, \ldots, r_n)$ is the selected L-expression in $\mathcal{Q}$):*

1) $\langle Q[\Omega(r_1, \ldots, r_n)]; \sigma\rangle \rightarrow_{SIS} \langle Q[\Omega(r_1, \ldots, r_n)/E']; \sigma\rangle$, *if Ω is a connective defined as $\Omega(x_1, \ldots, x_n) \triangleq E$ and E' is obtained from the L-expression E by replacing each variable (formal parameter) x_i by its corresponding value (actual parameter) r_i, $1 \leq i \leq n$, that is, $E' = E[x_1/r_1, \ldots, x_n/r_n]$.*
2) $\langle Q[\Omega(r_1, \ldots, r_n)]; \sigma\rangle \rightarrow_{SIS} \langle Q[\Omega(r_1, \ldots, r_n)/r]; \sigma\rangle$, *if Ω is a primitive operator such that, once evaluated with parameters $r_1, \ldots, r_n$, produces the result r.*

From now, we shall use the symbols $\rightarrow_{SIS1}$ and $\rightarrow_{SIS2}$ to distinguish between computation steps performed by applying one of the specific "small interpretive" rules. Moreover, when we use the expression *interpretive derivation*, we refer to a sequence of *small interpretive steps* (according to the previous definition) instead of a sequence of *interpretive steps* (regarding Definition 3). Note that this fact, supposes too a slight revision of Definition 4 which does not affect the essence of the notion of fuzzy computed answer: the repeated application of both kinds of

interpretive steps on a given state only affects to the length of the corresponding derivations, leading both ones to the same final states (fuzzy computed answers).

There are some remarks to do regarding Definition 5. Firstly, compared with Definition 1, we observe some correspondences among $\rightarrow_{AS1}$ and $\rightarrow_{AS2}$ with respect to $\rightarrow_{SIS1}$ and $\rightarrow_{SIS2}$. Note that both $\rightarrow_{AS2}$ and $\rightarrow_{SIS2}$ simply replace a portion of a goal $\mathcal{Q}$ (the selected atom or L-expression, respectively) by a single value of L (the truth degree of a program rule with empty body, or the result of evaluating a primitive operator, respectively). On the other hand, the replacements performed by $\rightarrow_{AS1}$ and $\rightarrow_{SIS1}$ can be seen as the expansion on the new state of some *definitions* appearing in the program (i.e., an atom is replaced by the body of a program rule, or a L-expression is replaced by the right hand side of a connective definition, respectively). This last similitude is even more evident if we formalize $\rightarrow_{SIS1}$ as follows: "$\langle Q[\Omega(r_1,\ldots,r_n)];\sigma\rangle \rightarrow_{SIS1} \langle (Q[\Omega(r_1,\ldots,r_n)/E])\theta];\sigma\theta\rangle$, if $\theta = mgu(\{\Omega(x_1,\ldots,x_n) = \Omega(r_1,\ldots,r_n)\})$ and $\Omega(x_1,\ldots,x_n) \triangleq E$ is a connective definition in $\mathcal{P}$". The reader may easily check that, although we have used unifiers in this alternative definition of $\rightarrow_{SIS1}$ in order to evoke Definition 1, we obtain exactly the same effect of Definition 5 (it is easy to see that the *matcher* θ we have just described is the set of links $x_1/r_1,\ldots,x_n/r_n$, which neither affects $\mathcal{Q}$ nor σ, but once applied to E generates the L-expression E' used in the original definition of small interpretive step).

Example 4. Recalling again the a.c.a. obtained in Example 1, we can reach the final fuzzy computed answer $\langle 0.63; \{X/a\}\rangle$ (achieved in Example 2 by means of interpretive steps) by generating now the following interpretive derivation D_2 based on "small interpretive steps":

$$
\begin{array}{ll}
\langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\underline{\vee_{\mathtt{L}}(0.8, 0.6)}, 0.7)); \{X/a\}\rangle & \rightarrow_{SIS1} \\
\langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\underline{min(1, 0.8 + 0.6)}, 0.7)); \{X/a\}\rangle & \rightarrow_{SIS2} \\
\langle \&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\underline{min(1, 1.4)}, 0.7)); \{X/a\}\rangle & \rightarrow_{SIS2} \\
\langle \&_{\mathtt{P}}(0.9, \underline{\&_{\mathtt{G}}(1, 0.7)}); \{X/a\}\rangle & \rightarrow_{SIS1} \\
\langle \&_{\mathtt{P}}(0.9, \underline{min(1, 0.7)}); \{X/a\}\rangle & \rightarrow_{SIS2} \\
\langle \underline{\&_{\mathtt{P}}(0.9, 0.7)}; \{X/a\}\rangle & \rightarrow_{SIS1} \\
\langle \underline{0.9 * 0.7}; \{X/a\}\rangle & \rightarrow_{SIS2} \\
\langle 0.63; \{X/a\}\rangle &
\end{array}
$$

Going back now to Example 3, we can rebuild the interpretive phase of Derivation D_1^* in terms of small interpretive steps, thus generating the following interpretive derivation D_2^*. Firstly, by applying a $\rightarrow_{SIS1}$ step on the L-expression $\&_{\mathtt{P}}(0.9, \underline{@(0.8, 0.7)})$, it becomes $\&_{\mathtt{P}}(0.9, \&_{\mathtt{G}}(\vee_{\mathtt{L}}(0.8, 0.6), 0.7))$, and from here, the interpretive derivation evolves exactly as derivation D_2 we have just done above.

At this moment, it is mandatory to meditate on cost measures regarding derivations D_1, D_1^*, D_2 and D_2^*. First of all, note that the operational cost $\mathcal{O}_c$ of all them coincides, which is quite natural. However, whereas $\mathcal{I}_c(D_1) = 3 > 2 = \mathcal{I}_c(D_1^*)$, we have now that $\mathcal{I}_c(D_2) = 7 < 8 = \mathcal{I}_c(D_2^*)$. This apparent contradiction might confuse us when trying to decide which program rule ($\mathcal{R}_1$ or $\mathcal{R}_1^*$) is "better". The use of Definition 5 in derivations D_2 and D_2^* is the key point to solve our

problem, as we are going to see. In Example 3 we justified that by simply counting the number of interpretive steps performed in Definition 3 might produce abnormal results, since the evaluation of connectives with different complexities were (wrongly) measured with the same computational cost. Fortunately, the notion of small interpretive step makes visible in the proper derivation all the connectives and primitive operators appearing in the (possibly recursively nested) definitions of any connective appearing in any derivation state. As we have seen, in D_2 we have expanded in three $\rightarrow_{SIS1}$ steps the definitions of three connectives, i.e. $\vee_{\mathtt{L}}, \&_{\mathtt{G}}$ and $\&_{\mathtt{P}}$, and we have applied four $\rightarrow_{SIS2}$ steps to solve four primitive operators, that is, $+$, min (twice) and $*$. The same computational effort as been performed in D_2^*, but also one more $\rightarrow_{SIS1}$ step was applied to accomplish with the expansion of the extra connective @. This justifies why $\mathcal{I}_c(D_2) = 7 < 8 = \mathcal{I}_c(D_2^*)$ and contradicts the wrong measures of Example 3: the interpretive effort developed in derivations D_1 and D_2 (both using the program rule $\mathcal{R}_1$), is slightly lower than the one performed in derivations D_1^* and D_2^* (which used rule $\mathcal{R}_1^*$), and not the contrary.

The accuracy of our new way for measuring and performing interpretive computations seems to be crucial when comparing the execution behaviour of programs obtained by transformation techniques such as the fold/unfold framework we describe in [5,3]. In this sense, instead of measuring the absolute cost of derivations performed in a program, we are more interested in the relative gains/lost of efficiency produced on transformed programs. For instance, by applying the so-called "aggregation operation" described in [3] we can transform rule $\mathcal{R}_1$ into $\mathcal{R}_1^*$ and, in order to proceed with alternative transformations (fold,unfold, et...) if the resulting program degenerates w.r.t. the original one (as occurs in this case), we need an appropriate cost measure as the one proposed here to help us for taken decisions. In [7] we faced the same problem in a different way which produced less accurate results. The idea was to redefine $\mathcal{I}_c$ in terms of the "weights" (that is, the number of primitive operators involved in the definition of) the connectives evaluated in each interpretive step of a given derivation. Although the method represents an step beyond the trivial, wrong way of simple counting steps without taken into account the number of primitive operators effectively evaluated in a derivation (as our new approach captures by means of $\rightarrow_{SIS2}$), it was not capable for distinguishing the number of calls to intermediate connectives (as our more clever technique do via $\rightarrow_{SIS1}$ steps).

This fact has capital importance for discovering drastic situations which can appear in degenerated transformation sequences such as the generation of highly nested definitions of aggregators. For instance, assume the following sequence of connective definitions: $@_{100}(x_1, x_2) \triangleq @_{99}(x_1, x_2), @_{99}(x_1, x_2) \triangleq @_{98}(x_1, x_2), \ldots,$ and finally $@_1(x_1, x_2) \triangleq x_1 * x_2$. When trying to solve two expression of the form $@_{99}(0.9, 0.8)$ and $@_1(0.9, 0.8)$, cost measures based on number of interpretive steps ([6]) and weights of interpretive steps ([7]) would assign 1 unit of interpretive cost to both derivations. Fortunately, our new approach is able to clearly distinguish between both cases, since the number of $\rightarrow_{SIS1}$ steps performed in each one is rather different (100 and 1, respectively).

4 Conclusions and Future Work

In this paper we were concerned with a pending task related with the formalization of the procedural semantics of multi-adjoint logic programming. Firstly, we have clarified the notion of small interpretive step by extending our preliminary definition of interpretive step in [6]. Then, we have showed how to accurately estimate the computational effort developed during the interpretive phase of derivations involving states with connectives whose definitions also invoke other connectives. The evaluation of this last kind of connectives consumes computational resources at execution time which are observed by our technique in a much more explicit way than the method proposed in [7] (based on weights of connectives in concordance with their complexities). In the near future we plan to implement the notion of small interpretive step into the FLOPER environment [12], as well as to take advantage of this concept to formally prove the efficiency of the fuzzy fold/unfold techniques [5,3] we are developing in our research group.

References

1. Baldwin, J.F., Martin, T.P., Pilsworth, B.W.: Fril–Fuzzy and Evidential Reasoning in Artificial Intelligence. John Wiley & Sons, Inc., Chichester (1995)
2. Guadarrama, S., Muñoz, S., Vaucheret, C.: Fuzzy Prolog: A new approach using soft constraints propagation. Fuzzy Sets and Systems 144(1), 127–150 (2004)
3. Guerrero, J.A., Moreno, G.: Optimizing fuzzy logic programs by unfolding, aggregation and folding. Electronic Notes in Theoretical Computer Science 219, 19–34 (2008)
4. Ishizuka, M., Kanai, N.: Prolog-ELF Incorporating Fuzzy Logic. In: Joshi, A.K. (ed.) Proceedings of the 9th Int. Joint Conference on Artificial Intelligence, IJCAI 1985, pp. 701–703. Morgan Kaufmann, San Francisco (1985)
5. Julián, P., Moreno, G., Penabad, J.: On Fuzzy Unfolding. A Multi-adjoint Approach. Fuzzy Sets and Systems 154, 16–33 (2005)
6. Julián, P., Moreno, G., Penabad, J.: Operational/Interpretive Unfolding of Multi-adjoint Logic Programs. Journal of Universal Computer Science 12(11), 1679–1699 (2006)
7. Julián, P., Moreno, G., Penabad, J.: Measuring the interpretive cost in fuzzy logic computations. In: Masulli, F., et al. (eds.) WILF 2007. LNCS (LNAI), vol. 4578, pp. 28–36. Springer, Heidelberg (2007)
8. Li, D., Liu, D.: A fuzzy Prolog database system. John Wiley & Sons, Inc., Chichester (1990)
9. Lloyd, J.W.: Foundations of Logic Programming. Springer, Berlin (1987)
10. Medina, J., Ojeda-Aciego, M., Vojtáš, P.: Multi-adjoint logic programming with continuous semantics. In: Eiter, T., Faber, W., Truszczyński, M. (eds.) LPNMR 2001. LNCS (LNAI), vol. 2173, pp. 351–364. Springer, Heidelberg (2001)
11. Medina, J., Ojeda-Aciego, M., Vojtáš, P.: Similarity-based Unification: a multi-adjoint approach. Fuzzy Sets and Systems 146, 43–62 (2004)
12. Morcillo, P.J., Moreno, G.: Programming with Fuzzy Logic Rules by using the FLOPER Tool. In: Bassiliades, N., et al. (eds.) RuleML 2008. LNCS, vol. 5321, pp. 119–126. Springer, Heidelberg (2008)
13. Vojtáš, P.: Fuzzy Logic Programming. Fuzzy Sets and Systems 124(1), 361–370 (2001)

Nestings of T–Conorms

Javier Martín, Gaspar Mayor, and Jaume Monreal

University of the Balearic Islands
Cra. Valldemossa km 7.5, 07122 Palma (Mallorca), Spain
{javier.martin,gmayor,jaume.monreal}@uib.es

Abstract. This paper deals with a new method to construct a new t–conorm from two previous ones, by what we call a nesting procedure. After some general considerations about this method, we study those t–conorms obtained by nesting appropriate t–conorms in the basic maximum, drastic and Lukasiewicz t–conorms.

Keywords: binary operation, nesting, ordinal sum, t–conorm.

1 Introduction

It is important to have at hand as many methods as possible for constructing aggregation operators, to be used in a wide range of applications. It is clear that the most famous method is the ordinal sum construction, which goes back to Climescu (1946) and Clifford (1954) for semigroups and can be applied to t-norms and t–conorms (see for example [1]), copulas [2], or generalized to aggregation operators [3]. Recently, the ordinal sum construction has been considered as a special case of an orthogonal grid construction [4].

In this paper we study a procedure for constructing t–conorms by a nesting method that includes the ordinal sum construction as well. This nesting procedure was first investigated for finitely-valued t–conorms [5,6], and here it is analyzed for ordinary t–conorms. Of course, equivalent results to those obtained for t–conorms can be also stated for t–norms, due to the duality that relates them.

Despite the theoretical character of this work, interesting families of t-conorms can be obtained by this method, which can be appropriated in specific situations.

The paper is organized as follows. In Section 2, we recall some basic definitions and properties of t–conorms. In Section 3, after introducing our procedure, we give a result that characterizes those nestings which are t–conorms, and then, based on this result, we study in Section 4 nestings in the basic background t–conorms.

2 Preliminaires

Consider the unit interval $I = [0, 1]$ equipped with the usual ordering. We begin to recall basic definitions, examples and some properties of t–conorms.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 52–59, 2009.

Definition 1. *A triangular conorm (t–conorm for short) on* $I = [0,1]$ *is a binary operation* $S : [0,1]^2 \to [0,1]$ *such that for all* $x, y, z \in I$ *the following axioms are satisfied:*

1. $S(x,y) = S(y,x)$ *(commutativity),*
2. $S\big(S(x,y),z\big) = S\big(x,S(y,z)\big)$ *(associativity),*
3. $S(x,y) \leqslant S(x',y')$ *whenever* $x \leqslant x'$, $y \leqslant y'$ *(monotonicity),*
4. $S(x,0) = x$ *(boundary condition).*

Example 1. We can consider as basic t–conorms the *maximum*,

$$S_M(x,y) = max(x,y); \tag{1}$$

the *drastic* t–conorm,

$$S_D(x,y) = \begin{cases} x & \text{if } y = 0, \\ y & \text{if } x = 0, \\ 1 & \text{otherwise;} \end{cases} \tag{2}$$

and the *bounded sum* or *Lukasiewicz* t–conorm,

$$S_L(x,y) = min(x+y, 1). \tag{3}$$

Well known properties of t–conorms are the following.

Proposition 1. *Let* S *be a t–conorm on* $I = [0,1]$*. Then we have:*

1. $S \leqslant S_D$*. Thus,* S_D *is the largest t–conorm.*
2. $S(x,y) \geqslant max(x,y)$, $\forall x, y \in I$*. That is,* S_M *is the smallest t–conorm.*
3. $S(x,1) = 1$, $\forall x \in I$*. In other words,* 1 *is an* annihilator*.*

Ordinal sums can be defined for a finite number or even a countable collection of t–conorms. Here we recall the well known ordinal sum theorem, first established in [7], applied to two t–conorms.

Proposition 2. *Let* S_1 *and* S_2 *be two t–conorms on* $I = [0,1]$*, and let* $a \in (0,1)$*. Consider the binary operation* S *defined on* I *as follows:*

$$S(x,y) = \begin{cases} aS_1(\frac{x}{a}, \frac{y}{a}) & if\ (x,y) \in [0,a]^2, \\ a + (1-a)S_2(\frac{x-a}{1-a}, \frac{y-a}{1-a}) & if\ (x,y) \in [a,1]^2, \\ \max(x,y) & otherwise. \end{cases}$$

Then, S *is a t–conorm on* I *which is called the* ordinal sum *of the* summands S_1 *and* S_2*. We will denote it by* $S = \langle S_1, S_2 \rangle$*.*

3 Nesting of T–Conorms

In this section we define a nesting procedure for constructing t–conorms. It has been already applied to the case of finite t–conorms [5,6]; here we apply it to t–conorms defined on the unit interval.

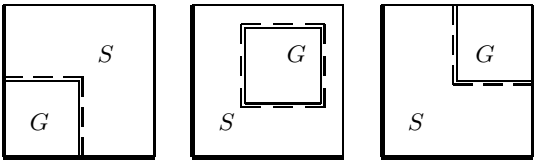


Fig. 1. The three different ways to nest

Definition 2. *Given a t–conorm S, real numbers $a, b \in [0,1]$, $a < b$, and a "t–conorm"*[1] *G on $[a,b]$, we define a binary operation $[G,S]$ on $[0,1]$ as follows:*

$$[G,S](x,y) = \begin{cases} G(x,y) & \text{if } a \leqslant x, y \leqslant a, \\ S(x,y) & \text{otherwise.} \end{cases} \tag{4}$$

We say that $[G,S]$ is the nesting *of G in S (a and b fixed).*

Depending on the values of a and b, it can be distinguished three cases, depicted in Fig. 1:

- $a = 0$ (and $b \neq 1$),
- $0 < a < b < 1$,
- $b = 1$ (and $a \neq 0$).

For any S, a and G , $[G,S]$ is commutative, non-decreasing in each place, with 0 and 1 as neutral and annihilator elements respectively. We are interested in obtaining by this method a new t–conorm.

Proposition 3. *Given S, a and G as above, the nesting $[G,S]$ is associative (is a t–conorm) if and only if the following conditions hold, when applicable:*

$$S\big(G(x,y),z\big) = S\big(S(x,y),z\big), \quad \forall x,y,z : a \leqslant x, y \leqslant b < z, \tag{5}$$

$$S(x,y) = \max(x,y), \quad \forall x,y : x < a \leqslant y < b. \tag{6}$$

Proof. Let us assume that $[G,S]$ is a t–conorm. First we observe that the associativity of $[G,S]$ and S implies condition (5). From monotonicity of $[G,S]$ and boundary conditions of G and S we can write

$$[G,S](0,y) \leqslant [G,S](x,y) \leqslant [G,S](a,y)$$

when $x < a \leqslant y \leqslant b$, and

$$y = S(0,y) \leqslant S(x,y) \leqslant G(a,y) = y.$$

So, $S(x,y) = \max(x,y) = y$ and (6) is satisfied.

[1] A "t–conorm" on $[a,b]$ is a binary operation which is commutative, associative and monotone, with a as neutral element. A t–conorm means here a "t–conorm" on the unit interval $[0,1]$. If S is a t–conorm and $0 \leqslant a < b \leqslant 1$ the we denote by $S_{a,b}$ the "t–conorm" on $[a,b]$ defined by $S_{a,b}(x,y) = a + (b-a)S(\frac{x}{b}, \frac{y}{b})$, or simply S_b if $a = 0$.

Reciprocally, monotonicity of $[G, S]$ follows from (6) and associativity, from both (5) and (6). □

Remarks

1. From this proposition we can see that if S is cancellative for some $z > b$, then $G(x, y) = S(x, y)$ for all $x, y \in [a, b]$ and so $[G, S] = S$.
2. In the case $a = 0$, a situation of interest is when $S(b, x) = \max(b, x)$ for all $x \in [0, 1]$. Under this hypothesis condition (5) is trivially satisfied for all "t–conorm" G and all b, $0 < b < 1$, and $[G, S]$ is just the ordinal sum $\langle S_1, S_2 \rangle$ where S_1 and S_2 are the t–conorms defined by

$$S_1(x, y) = \frac{1}{b} G(bx, by), \tag{7}$$

$$S_2(x, y) = \frac{1}{1-b}\big(S(b + (1-b)x, b + (1-b)y) - b\big). \tag{8}$$

   Following with this case $a = 0$, it is also worth to observe that if $[G, S]$ is a t–conorm then it is non-Archimedean[2] (a is a non trivial idempotent of $[G, S]$). Reciprocally, if S is a non-Archimedean t–conorm with a as non-trivial idempotent, then S is the nesting $S = [G, S]$, where G is the "t–conorm" on $[0, a]$ defined by $G(x, y) = aS(\frac{x}{a}, \frac{y}{a})$. In particular, any ordinal sum $\langle S_1, S_2 \rangle$ is a nesting.
   Thus the class of non-Archimedean t–conorms is equal to the class of nestings satisfying condition (5).
   Note that $[G, S]$ is left–continuous, and it is continuous if and only if it is the ordinal sum $\langle S_1, S_2 \rangle$ where S_1 and S_2 are the (continuous) t–conorms defined in (7) and (8).
3. If $0 < a < b < 1$, then condition (6) says that the restriction of $[G, S]$ to $[0, b]^2$ is a "t–conorm"; that is, $[[G, S]|_{[0,b]^2}, S]$ is a nesting of the first type. And for the case $b = 1$, it follows from the same condition that the resulting t–conorm is also an ordinal sum and hence it can be interpreted as a nesting of the first type.

4 Nestings in the Three Basic T–Conorms

We apply now the construction defined in Section 3 to the case of maximum, drastic and Łukasiewicz t–conorms; we also consider the iteration of this process. The study is limited to the nestings of the case $a = 0$: the other two cases can be reduced to it, as we have seen in point 3 of the previous remark.

4.1 Nestings in the Maximum T–Conorm

Proposition 4. *The nesting $[G, S_M]$ is a t–conorm for all "t–conorm" G and all b, $0 < b < 1$.*

Proof. As said in the remark of the previous section, point 2, a nesting in the maximum is an ordinal sum, that is, a t–conorm. □

[2] Here we consider Archimedean a t–conorm without non-trivial idempotent elements.

4.2 Nestings in the Drastic T–Conorm

Proposition 5. *The nesting* $[G, S_D]$ *is a t–conorm for all "t–conorm"* G *and all* b, $0 < b < 1$.

Proof. Only condition (5) must be verified: if $a \leqslant x, y \leqslant b < z$, then

$$S_D\big(G(x,y),z\big) = S_D\big(S_D(x,y),z\big) = 1.$$

□

Remarks

1. Observe that $[G, S_D]$ is a non-Archimedean t–conorm which is not an ordinal sum.
2. Taking into account nestings in maximum and drastic t–conorm, $[G, S]$ can be seen as a median:
$$[G,S] = \mathrm{med}\big(S, [G,S_M], [G,S_D]\big)$$
3. Note also the possibility of (up)-iteration of the two previous nesting cases, obtaining at each step new t–conorms:
$$[G,S_M],\ \big[[G,S_M]_c, S_M\big],\ \ldots \qquad [G,S_D],\ \big[[G,S_D]_c, S_D\big],\ \ldots$$
where $c \in (0,1)$ and $[G,S_M]_c$, for example, indicates the rescaling of the t–conorm $[G,S_M]$ to the interval $[0,c]$ (see footnote 1), obtaining in this way the first iteration. Obviously this process can be repeated indefinitely.

4.3 Nestings in the Łukasiewicz T–Conorm

Proposition 6. *The nesting* $[G, S_L]$ *is a t–conorm if and only if the following conditions hold:*

i) $b \geqslant \frac{1}{2}$,
ii) $G(x,y) = x + y$ *if* $x + y < 1 - b$.

Proof. First we observe that condition (5) can be rewritten in our case in this way:

$$\min\big(G(x,y) + z, 1\big) = \min(x + y + z, 1), \quad \forall x, y, z : x, y \leqslant b < z. \tag{9}$$

Let us suppose that this condition is satisfied; thus, $G(x,y) = x + y$ whenever $x + y < 1 - z$. This means $b \geqslant \frac{1}{2}$: if not, taking $z > b$ such that $b < 1 - z$ we would have $G(x,y) = x + y$ for all $b \leqslant x + y < 1 - z$, which is not possible.

Now, the condition on G can be given in the form:

$$G(x,y) = x + y, \quad \forall x, y : x + y < 1 - b,$$

and conditions *i)*, *ii)* are satisfied.

Reciprocally, given $b \geqslant \frac{1}{2}$ and G satisfying the above condition, it can be proved that (9) is fulfilled and $[G, S_L]$ is a t–conorm. □

Remarks

1. Note that from the second condition of the above proposition we can deduce $G(x,y) \geqslant 1-a$ if $x+y \geqslant 1-a$.
2. It is interesting also to observe that similar conditions as those of the Proposition 6 can be obtained for continuous nilpotent t–conorms. More precisely, if f is the normalized additive generator of a t–conorm S of this type, then the conditions are reformulated in this way:
   i) $b \geqslant f^{-1}(\frac{1}{2})$,
   ii) $G(x,y) = S(x,y)$ if $y \leqslant f^{-1}\big(1-f(b)-f(x)\big)$.

Example 2. For $b = \frac{1}{2}$, the nesting $[S_{L_{\frac{1}{2}}}, S_L]$ is the left-continuous t–conorm defined by:

$$[S_{L_{\frac{1}{2}}}, S_L](x,y) = \begin{cases} \min(x+y, \frac{1}{2}) & \text{if } 0 \leqslant x,y \leqslant \frac{1}{2}, \\ \min(x+y, 1) & \text{otherwise.} \end{cases} \tag{10}$$

The next result states the conditions for an iterated nesting in Łukasiewicz t–conorm to be also a t–conorm.

Proposition 7. *Let G a "t–conorm" on $[0,b]$, $[G, S_L]$ a t–conorm and c a real number such that $b < c < 1$. Then $\big[[G,S_L]_c, S_L\big]$ is a t–conorm if and only if the following conditions hold:*

i) $c \geqslant \frac{1}{1+b}$,
ii) $G(x,y) = x+y$ *if* $x+y < \frac{1-c}{c}$.

Corollary 1. *Given $0<b,c<1$ and a "t–conorm" G on $[0,b]$, then $\big[[G,S_L]_c, S_L\big]$ is a t–conorm if and only if it fulfils:*

i) $b \geqslant \frac{1}{2}$, $c \geqslant \frac{1}{1+b}$,
ii) $G(x,y) = x+y$ *if* $x+y < \max(1-b, \frac{1-c}{c})$.

Example 3.

i) If we take $b = \frac{1}{2}$, then $\big[[G,S_L]_c, S_L\big]$ is a t–conorm for all $c \geqslant \frac{2}{3}$ with $G(x,y) = \min(x+y, \frac{1}{2})$ (i.e. $G = S_{L_{\frac{1}{2}}}$).
ii) There is no G such that $\big[[G,S_L]_{\frac{1}{2}}, S_L\big]$ is a t–conorm.

As said before, this process of nesting can be up-iterated. In the case of nesting in Łukasiewicz t–conorm, we obtain a binary operation which can be denoted by $S_{L_{b_0,b_1,\ldots,b_{n-1},b_n}}$, with $0 = b_0 < b_1 < \ldots < b_{n-1} < b_n = 1$, or $S_{L_{b_1,\ldots,b_{n-1}}}$:

$$S_{L_{b_1,\ldots,b_{n-1}}} = \Big[\big[\ldots[[S_{L_{b_1}}, S_L]_{b_2}, S_L]_{b_3}\ldots\big]_{b_{n-1}}, S_L\Big],$$

which can be expressed as follows:

$$S_{L_{b_1,\ldots,b_{n-1}}}(x,y) = \min(x+y, b_k), \quad \text{if } b_{k-1} < \max(x,y) \leqslant b_k.$$

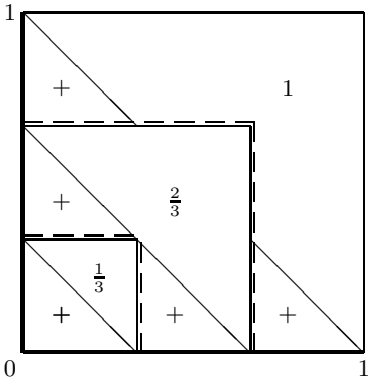


Fig. 2. The t–conorm $S_{L_3} = S_{L_{\frac{1}{3},\frac{2}{3}}} = \left[[S_{L_{\frac{1}{2}}}, S_L]_{\frac{2}{3}}, S_L\right]$

Proposition 8. *The binary operation* $S_{L_{b_1,\ldots,b_{n-1}}}$ *is a t–conorm if and only if* $b_i - b_{i-1} \leqslant b_1$ *for all* $i = 1, \ldots, n-1$.

Special t–conorms of this type are $S_{L_{\frac{1}{n},\ldots,\frac{n-1}{n}}}$, which will be denoted by S_{L_n} for the sake of brevity.

Example 4. For $n = 3$, we have the t–conorm S_{L_3} in Fig. 2.

An interesting expression for S_{L_n} is

$$S_{L_n}(x, y) = \min\left(S_L(x, y), S_{D_n}(x, y)\right).$$

It is the minimum of two t–conorms: the Łukasiewicz one and the t–conorm S_{D_n}, or $S_{D_{\frac{1}{n},\ldots,\frac{n-1}{n}}}$, which is the result of the up-iteration process of nesting in drastic t–conorm:

$$S_{D_n}(x, y) = \begin{cases} x & \text{if } y = 0, \\ y & \text{if } x = 0, \\ \frac{k}{n} & \text{if } x, y \neq 0 \text{ and } \frac{k-1}{n} < \max(x, y) \leqslant \frac{k}{n}. \end{cases}$$

or, in short,

$$S_{D_n}(x, y) = \frac{\lceil n \max(x, y) \rceil}{n},$$

where $\lceil \cdot \rceil$ is the ceiling function, which maps a real number to the next larger integer. In this way, S_{L_n} can be expressed as (compare with (3)):

$$S_{L_n}(x, y) = \min\left(x + y, \frac{\lceil n \max(x, y) \rceil}{n}\right).$$

We finish with an example where these t–conorms are applied.

Example 5. Aspirants to pass an examination carried out two attempts. We can classify them with the t–conorm S_{D_n} at several levels $L_i = \frac{i}{n}$, $i = 1, \ldots, n$, in this way: if the highest result is between L_{i-1} and L_i, we put them at level L_i, regardless of lowest one.

But when the sum of both scores of an aspirant does not reach the level L_i, we can penalize him by using the t–conorm S_{L_n}, which assigns precisely the value of this sum in this case.

5 Conclusions

A nesting method to construct binary operations from t–conorms has been introduced, which includes as a particular case the ordinal sum. The general conditions for this method to give as a result also t–conorms have been established and the nestings in the three basic t–conorms have been specially considered.

Acknowledgments. The authors acknowledge the support of the Govern Balear grant PCTIB2005GC1-07 and the Spanish DGI grant MTM2006-08322.

References

1. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer, The Netherlands (2000)
2. Nelsen, R.: An Introduction to Copulas, 2nd edn. Lecture Notes in Statistics, vol. 139. Springer, New York (2006)
3. De Baets, B., Mesiar, R.: Ordinal sums of aggregation operators. In: Bouchon–Meunier, B., Gutiérrez Ríos, J., Magdalena, L., Yager, R.R. (eds.) Technologies for Constructing Intelligent Systems: Tools, pp. 137–149. Physica–Verlag, Heidelberg (2002)
4. De Baets, B., De Meyer, H.: Orthogonal Grid Constructions of Copulas. IEEE Trans. on Fuzzy Systems 15, 1053–1062 (2007)
5. Martín, J., Mayor, G., Monreal, J.: Additive generation of some classes of finitely-valued t–conorms. In: Proceedings of IPMU 2008, 12th International Conference on Information Processing and Management of Uncertainty in Knowledge–Based Systems, Málaga, Spain, pp. 1326–1331 (2008)
6. Martín, J., Mayor, G., Monreal, J.: The problem of the additive generation of finitely-valued t–conorms. Mathware & Soft Computing 16, 17–27 (2009)
7. Schweizer, B., Sklar, A.: Associative functions and abstract semigroups. Publ. Math. Debrecen 10, 69–81 (1963)

On Coherence and Consistence in Fuzzy Answer Set Semantics for Residuated Logic Programs*

Nicolás Madrid and Manuel Ojeda-Aciego

Dept. Matemática Aplicada. Univ. Málaga, Spain
{nmadrid,aciego}@uma.es

Abstract. In this work we recall the first steps towards the definition of an answer set semantics for residuated logic programs with negation, and concentrate on the development of relationships between the notions of coherence and consistence of an interpretation.

1 Introduction

An answer set semantics has been recently introduced in the framework of residuated logic programming in response of the need of developing new reasoning paradigms for knowledge representation and non-monotonic deduction problems.

One of the most important areas on which this kind of research can be applied is in the development of the *semantic web*. Mainly due to its vast nature, reasoning with current technologies has been deliberately chosen to be monotonic. Although this choice seems to be the right one when dealing with the semantic web as a whole, the benefits of non-monotonic reasoning become apparent in a local sense: for instance, when a small number of agents communicate they need to consider only their own knowledge bases as all the relevant knowledge, regardless the big amount of information out there. In this case, non-monotonic reasoning is advantageous over monotonic reasoning, i.e. one can retract previous inferences on the discovery of new knowledge by one of the agents, or one can safely assume everything that is not known to the agents as false.

Originally, answer sets semantics was intended to deal with non-monotonic reasoning, in that it provides a method to handle negation in logic programming. Moreover, two kind of negations, one strong negation and one default negation, were allowed in the programs. The use of these two types of negation is advocated in many contexts of interest, in particular in [10] their use is justified in relation to web rules. Moreover, the overall framework of answer set programming has important links with description logics, as stated in [2,5]. Specifically, [2] have proposed a combination of logic programming under the answer set semantics with some description logics in order to build rules on top of ontologies and, to a limited extent, build ontologies on top of rules; on its turn, [5] introduces a language that unifies both answer set programming

* Partially supported by the Spanish Science Ministry grant TIN06-15455-C03-01 and by Junta de Andalucía grant P06-FQM-02049.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 60–67, 2009.

and expressive description logics as an alternative for intuitive non-monotonic reasoning with possibly infinite knowledge.

It is convenient to note that stable models, from which the answer set semantics arose, were initially aimed at formalizing the use of negation in logic programming as negation-as-failure and, thus, are closely related to reasoning under uncertainty. For instance, the closed world assumption for a given predicate P allows for extracting negative knowledge about P from the absence of positive information about it.

The ideal environment for developing a theory of management of uncertainty is fuzzy logic in any of its flavours. This is why we chose to introduce negations in a particular fuzzy logic programming paradigm, specifically the framework of residuated logic programs (which is negation-free) [1], and consider it as our target theory for a suitable generalization of answer set semantics.

In this paper, we start by recalling the basic definitions of stable model, answer set and coherent interpretation in the framework of residuated logic programs introduced in [7]. Then we initiate the analysis of the relationships between the notions of coherence of an interpretation with the more common notion of consistence.

2 On Fuzzy Answer Set Semantics for Residuated Programs

In this section we include the definitions needed to recall the answer set semantics for residuated logic programs with negation. Let us start with the definition of residuated lattice:

Definition 1. *A* residuated lattice *is a tuple* $(L, \leq, *, \leftarrow)$ *such that:*

1. $(L, \leq)$ *is a complete bounded lattice, with top and bottom elements 1 and 0.*
2. $(L, *, 1)$ *is a commutative monoid with unit element 1.*
3. $(*, \leftarrow)$ *forms an adjoint pair, i.e.* $z \leq (x \leftarrow y)$ *iff* $y * z \leq x \quad \forall x, y, z \in L$.

In residuated lattices one can interpret the operator $*$ like a conjunction and the operator $\leftarrow$ like an implication.

In the rest of the paper we will consider a residuated lattice enriched with two negation operators, $(L, \leq, *, \leftarrow, \sim, \neg)$. The two negations will modelize the notions of strong negation $\sim$ and default negation $\neg$ often used in logic programming. As usual, a negation operator, over L, is any decreasing mapping $n\colon L \to L$ satisfying $n(0) = 1$ and $n(1) = 0$.

The difference between strong and default negation in our context is essentially semantical, and relates to the method used to infer the truth value of one negated propositional symbol.

In order to introduce our logic programs, we will assume a set Π of propositional symbols. If $p \in \Pi$, then both p and $\sim p$ are called *literals*. We will denote arbitrary literals with the symbol ℓ (possible subscripted), and the set of all literals as Lit.

Definition 2. *Given a residuated lattice with negations* $(L, \leq, *, \leftarrow, \sim, \neg)$*, a* general residuated logic program $\mathbb{P}$ *is a set of weighted rules of the form*

$$\langle \ell \leftarrow \ell_1 * \cdots * \ell_m * \neg \ell_{m+1} * \cdots * \neg \ell_n; \quad \vartheta \rangle$$

where ϑ *is an element of* L *and* $\ell, \ell_1, \ldots, \ell_n$ *are literals.*

Rules will be frequently denoted as $\langle \ell \leftarrow \mathcal{B}; \vartheta \rangle$. As usual, the formula $\mathcal{B}$ is called the *body* of the rule whereas ℓ is called its *head*. We consider *facts* as rules with empty body, which are interpreted as a rule $\langle \ell \leftarrow 1; \quad \vartheta \rangle$.

Definition 3. *A fuzzy* L-interpretation *is a mapping* $I\colon Lit \to L$*; note that the domain of the interpretation can be lifted to any rule by homomorphic extension.*

We say that I satisfies *a rule* $\langle \ell \leftarrow \mathcal{B}; \quad \vartheta \rangle$ *if and only if* $I(\mathcal{B}) * \vartheta \leq I(\ell)$ *or, equivalently,* $\vartheta \leq I(\ell \leftarrow \mathcal{B})$*.*

Finally, I *is a* model *of* $\mathbb{P}$ *if it satisfies all rules (and facts) in* $\mathbb{P}$*.*

A general residuated logic program $\mathbb{P}$ is said to be:

- *positive* if it does not contain negation operators.
- *normal* if it does not contain strong negation.
- *extended* if it does not contain default negation.

2.1 Extended Logic Programs and Coherence

In this section, we concentrate on strong negation and, therefore, we will consider extended residuated logic programs.

Note that, as our interpretations are defined on the set of literals, every extended program has a least model which can be obtained, for instance, by iterating the immediate consequence operator, see [1]. However, one has to take into account the interaction between opposite literals. For example, in the classical case we reject the inconsistent models, i.e. p and $\sim p$ cannot be true at the same time. The advantage of working in a fuzzy framework is that one can allow that two opposite literals, such as p and $\sim p$, live together ... under some requirements.

Our approach will be based on a generalization of the concept of consistency which we have called *coherence*, to distinguish it from other existing definitions of consistency in a fuzzy setting.

Definition 4. *A fuzzy* L*-interpretation* I *over* Lit *is* coherent *if the inequality* $I(\sim p) \leq \sim I(p)$ *holds for every propositional symbol* p*.*

Now, a natural question arises: why the definition above provides an acceptable generalization? There are three main reasons: firstly, it is easy to implement, since it only depends on the negation operator (whereas other definitions use both a t-norm and a negation); secondly, it allows to handle missing information (i.e. I such that $I(\ell) = 0$ for all $\ell \in Lit$ is always coherent); thirdly, our notion of coherence coincides with consistency in the classical framework (it is not difficult to check this).

We will also apply hereafter the term "coherent" to refer to a logic program, as stated by the following definition.

Definition 5. *Let $\mathbb{P}$ be an extended residuated logic program, we say that $\mathbb{P}$ is* coherent *if its least model is coherent.*

In Section 3 we will introduce some results about coherence, in particular that an extended program is coherent if and only if it has *some* coherent model, and we will compare it with other approaches to the generalization of consistency to a fuzzy framework.

Example 1. Consider the following extended residuated logic program

$$\mathbb{P} = \{\langle p \leftarrow; 1\rangle, \langle \sim p \leftarrow; 0.3\rangle\}$$

over the unit interval and strong negation $\sim x = 1 - x$:

This program is not coherent because its unique minimal model $M = \{(p, 1), (\sim p, 0.3)\}$ is not a coherent interpretation, since $0.3 = M(\sim p) > \sim M(p) = 0$.

2.2 Fuzzy Answer Sets

Once the concept of coherence has been presented, we can introduce the notion of *fuzzy answer set* for extended logic programs. Such a set is a fuzzy set of literals, similarly to the classical case, the difference is that in our framework it will be considered a particular case of fuzzy L-interpretation.

Definition 6. *Let $\mathbb{P}$ be a coherent extended residuated logic program; the* fuzzy answer set *of $\mathbb{P}$ is its least coherent model of $\mathbb{P}$.*

Our aim in this section is to adapt the approach given in [3,4] to the *general* residuated logic programs defined above.

Let us consider a general residuated logic program $\mathbb{P}$ together with a fuzzy L-interpretation I. To begin with, we will construct a new normal program $\mathbb{P}_I$ by substituting each rule in $\mathbb{P}$ of the form

$$\langle \ell \leftarrow \ell_1 * \cdots * \ell_m * \neg \ell_{m+1} * \cdots * \neg \ell_n; \quad \vartheta \rangle$$

by the rule[1]

$$\langle \ell \leftarrow \ell_1 * \cdots * \ell_m; \quad \neg I(\ell_{m+1}) * \cdots * \neg I(\ell_n) * \vartheta \rangle$$

Notice that the new program $\mathbb{P}_I$ is extended, that is, does not contain default negation; in fact, the construction closely resembles that of a reduct in the classical case, this is why we introduce the definition below.

Definition 7. *The program $\mathbb{P}_I$ is called the* reduct *of $\mathbb{P}$ wrt the interpretation I.*

It is not difficult to prove that every model M of the program $\mathbb{P}$ is a model of the reduct $\mathbb{P}_M$.

[1] Note the overloaded use of the negation symbol, as a syntactic function in the formulas and as the algebraic negation in the truth-values.

Remark 1. As a result, note that given two fuzzy L-interpretations I and J, then the reducts $\mathbb{P}_I$ and $\mathbb{P}_J$ have the same rules, and might only differ in the values of the weights. By the monotonicity properties of $*$ and $\neg$, we have that if $I \leq J$ then the weight of a rule in $\mathbb{P}_I$ is greater or equal than its weight in $\mathbb{P}_J$.

Now we are ready to introduce our notion of (fuzzy) answer set for general residuated logic program.

Definition 8. *Let $\mathbb{P}$ be a general residuated logic program and let I be a coherent fuzzy L-interpretation; I is said to be an* answer set[2] *of $\mathbb{P}$ iff I is a minimal model of $\mathbb{P}_I$.*

Theorem 1. *Any answer set of $\mathbb{P}$ is a minimal model of $\mathbb{P}$.*

Obviously, this approach is a conservative extension of the classical approach. In the following example we use a simple normal logic program with just one rule in order to clarify the definition of answer set.

Example 2. Consider the program $\langle p \leftarrow \neg q; \quad \vartheta \rangle$. Given a fuzzy L-interpretation $I\colon \Pi \to L$, the reduct $\mathbb{P}_I$ is the rule (actually, the fact) $\langle p; \quad \vartheta * \neg I(q) \rangle$ for which the least model is $M(p) = \vartheta * \neg I(q)$, and $M(q) = 0$. As a result, I is an answer set of $\mathbb{P}$ if and only if $I(p) = \vartheta * \neg I(0) = \vartheta * 1 = \vartheta$ and $I(q) = 0$.

3 On Coherence and Consistence

Let us start this section by introducing the usual extension of the concept of consistent interpretation to the fuzzy case, which needs both a t-norm and a negation operator.

Definition 9. *Let $*$ be a t-norm and $\sim$ a negation operator. We say that an interpretation $I\colon Lit \to L$ on the set of literals is* α-consistent *if for all propositional symbol p we have that $I(p) * I(\sim p) \leq \alpha$.*

Note that, by the adjoint condition, $I(p) * I(\sim p) \leq \alpha$ iff $I(\sim p) \leq \alpha \leftarrow I(p)$. In other words, α-consistence provides an upper bound to the value of $I(\sim p)$ in terms of $I(p)$ and the parameter α. On its turn, recall that a coherent interpretation (Definition 4) directly provides such an upper bound, namely $\sim I(p)$, which depends only on the operator intended to interpret the strong negation.

Obviously, in a classical context, both terms are equivalent as stated in the proposition below:

Proposition 1. *In classical logic, an interpretation is coherent if and only if it is α-consistent for all $\alpha \in [0, 1)$.*

Example 3. Let us study the set of coherent interpretations associated to two extreme cases of negation. Firstly, for the least negation operator

[2] For normal residuated logic programs, this definition reduces to that of stable set.

$$\sim(x) = \begin{cases} 1 \text{ if } x = 0 \\ 0 \text{ if } x > 0 \end{cases}$$

we have that a coherent interpretation cannot assign a positive value to a propositional symbol and its negation,i.e if $I(p) > 0$ then $I(\sim p) = 0$.

Now, consider the greatest negation operator:

$$\sim(x) = \begin{cases} 1 \text{ if } x < 1 \\ 0 \text{ if } x = 1 \end{cases}$$

For this negation, the coherence condition states that if $I(p) = 1$ then $I(\sim p) = 0$ or, alternatively, that any positive value of $\sim p$ (arbitrarily small), implies that p cannot be certainly true. □

As shown in the previous example, although in a fuzzy context both definitions differ in general, there exist some relations between them. For instance, given $*$ and $\sim$, consider the value

$$\alpha^*_\sim = \sup\{x * \sim x \mid x \in L\}$$

which will be called the *consistence bound*. Let us see some examples on the unit interval.

Example 4. Consider the negation given by $\sim x = 1 - x$

1. For Gödel t-norm, $\min(x, y)$, the consistence bound is 0.5.
2. For product t-norm, $x \cdot y$, the consistence bound is 0.25.
3. For Łukasiewicz t-norm, $\max(0, x + y - 1)$, the consistence bound is 0. □

Proposition 2. *Let $*$ be a t-norm and $\sim$ a negation operator, then any coherent interpretation is $\alpha^*_\sim$-consistent.*

Remark 2. Note that, in the above example, any coherent interpretation is 0-consistent wrt Łukasiewicz t-norm, which requires the strongest type of consistence; however, in a fuzzy context this does not mean that either p or $\sim p$ should be evaluated as 0 since, for instance, if $I(p) = 0.5$ and $I(\sim p) = 0.5$ one still has $I(p) * I(\sim p) = 0$.

One of the main features of the notion of coherence is that it uniquely depends on the negation operator in use, contrariwise to the definition of α-consistence, which involves as well the underlying t-norm and the consistence level. Of course, Proposition 2 above helps the programmer to implement the intended behaviour regarding strong negation regarding the maximum common level that both a propositional symbol and its negation can have maintaining coherence.

Example 5. Assume that we are working with the standard negation operator $\sim x = 1 - x$. In order to find what is the maximum possible common value for $I(p)$ and $I(\sim p)$ in a coherent interpretation I firstly note that, the value of $I(\sim p)$ should reach its upper bound, that is $I(\sim p) = \sim I(p)$. By definition of the negation operator, this amounts to $I(\sim p) = 1 - I(p)$ and, as the values of

$I(p)$ and $I(\sim p)$ are assume to be the same, the previous equation leads that $I(p) = I(\sim p) = 0.5$.

As a result, independently of the underlying t-norm, the use of the standard negation operator and coherent interpretations does not allow that $I(p) = I(\sim p) > 0.5$.

Should we wish a bound different from 0.5, the solution would be to fix a different negation operator. A more restrictive bound is obtained by using $\sim x = 1 - \sqrt{x}$; In effect, by the same reasoning as with the standard negation, we are led to the equation $1 - \sqrt{x} = x$, whose positive solution is $x \approx 0.38$. On the other hand, a less restrictive one is obtained with the operator $\sim x = 1 - x^2$, in this case the equation to be solved is $1 - x^2 = x$, whose positive solution is $x \approx 0.62$. □

The previous example presented some negations which grant more or less restrictive consistence values; however, one would like to be able to construct the corresponding negation to a prescribed consistence value. In the unit interval, this is given by the following

Proposition 3. *Consider $\varepsilon \in [0, 1]$, then the following negation operator*

$$\sim x = \begin{cases} 1 - \frac{1-\varepsilon}{\varepsilon} x & \text{if} \quad x < \varepsilon \\ \frac{\varepsilon}{1-\varepsilon}(1 - x) & \text{if} \quad x \geq \varepsilon \end{cases}$$

satisfies that in all coherent interpretation

$$\sup\{\alpha \in [0, 1] \mid \exists p \text{ such that } I(p) = I(\sim p) = \alpha\} \leq \varepsilon.$$

Note that the operator given in the previous proposition is not the only one satisfying the statement, we have just provided a continuous negation operator with the intended behaviour.

In order to continue with some properties of the notion of coherence, take into account that an interpretation I assigns a truth degree to any negative literal $\sim p$ independently from the negation operator. This way, if we have two different negation operators ($\sim_1$ and $\sim_2$) we can talk about the coherence of I wrt any of these operators.

Proposition 4. *Let $\sim_1$ and $\sim_2$ be two negation operators such that $\sim_1 \leq \sim_2$, then any interpretation I that is coherent wrt $\sim_1$ is coherent wrt $\sim_2$.*

Another formulation of the previous property of coherence can be given, this time in terms of two interpretations and just one negation operator, as follows:

Proposition 5. *Let I and J be two interpretations satisfying $I \leq J$. If J is coherent, then I is coherent as well.*

Corollary 1. *If M is a fuzzy coherent model of $\mathbb{P}$, then any other model T such that $T \leq M$ is a coherent model.*

Corollary 2. *A extended residuated logic program is coherent if and only if it has at least one coherent model.*

4 Conclusions and Future Work

We have recalled the basic definitions of the answer set semantics of general residuated logic programs. We have concentrated on the relationships between the notions of coherence and consistence of an interpretation.

The notion of *consistence bound* has been introduced for a given t-norm $*$ and strong negation $\sim$, and its relationship with coherent interpretations has been presented. Then, coherence has been studied in terms of the ordering relation between interpretations.

A number of issues still have to be studied: for instance, the epistemological implications of the concept of coherence. We have only taken into account that the resulting fuzzy answer sets should be validated for coherence, as a consistency-related notion, and developed some of its initial properties. Future work, should go towards imbricating this notion with threshold computation which turns out to be an important issue for negation-as-failure. For instance, the absence of evidence of p could be interpreted that the value of p is at most a threshold value which cannot be detected by the sensors which provide our information.

Finally, it is important to further relate our approach with other existing approaches [6,9], and study their possible interactions, as well as studying the modifications needed in order to extend the answer set semantics to multi-adjoint logic programs [8].

References

1. Damásio, C.V., Pereira, L.M.: Monotonic and residuated logic programs. In: Benferhat, S., Besnard, P. (eds.) ECSQARU 2001. LNCS, vol. 2143, pp. 748–759. Springer, Heidelberg (2001)
2. Eiter, T., Lukasiewicz, T., Schindlauer, R., Tompits, H.: Combining answer set programming with description logics for the semantic web. In: Principles of Knowledge Representation and Reasoning (KR 2004), pp. 141–151 (2004)
3. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programming. In: Proc. of ICLP 1988, pp. 1070–1080 (1988)
4. Gelfond, M., Lifschitz, V.: Classical negation in logic programs and disjunctive databases. New Generation Computing 9, 365–385 (1991)
5. Heymans, S., Vermeir, D.: Integrating semantic web reasoning and answer set programming. Answer Set Programming, 195–209 (2003)
6. Lukasiewicz, T.: Fuzzy description logic programs under the answer set semantics for the semantic web. Fundamenta Informaticae 82(3), 289–310 (2008)
7. Madrid, N., Ojeda Aciego, M.: Towards an answer set semantics for residuated logic programs. In: IEEE/WIC/ACM Intl. Conf. on Web Intelligence and Intelligent Agent Technology, WI-IAT 2008, pp. 260–264 (2008)
8. Medina, J., Ojeda-Aciego, M., Vojtáš, P.: Similarity-based unification: a multi-adjoint approach. Fuzzy Sets and Systems 146(1), 43–62 (2004)
9. Van Nieuwenborgh, D., De Cock, M., Vermeir, D.: An introduction to fuzzy answer set programming. Ann. Math. Artif. Intell. 50(3-4), 363–388 (2007)
10. Wagner, G.: Web rules need two kinds of negation. In: Bry, F., Henze, N., Małuszyński, J. (eds.) PPSWR 2003. LNCS, vol. 2901, pp. 33–50. Springer, Heidelberg (2003)

Rough Set Approach to Rule Induction from Imprecise Decision Tables

Masahiro Inuiguchi

Graduate School of Engineering Science, Osaka University,
Toyonaka, Osaka 560-8531, Japan
inuiguti@sys.es.osaka-u.ac.jp

Abstract. In this paper, we investigate rule induction from imprecise decision tables. In the imprecise decision tables, decision attribute values are specified imprecisely. Under a definition of rough set with respect to imprecise decision tables, several rule induction schemes are considered. In each rule induction scheme, the conventional decision matrix method is extended to the case of imprecise decision tables.

1 Introduction

Rough set approach proposed by Pawlak [6] is known to be a useful tool for reasoning from data. It has been applied to various fields such as medicine, engineering, management and so on. In order to extend the applicability, rough sets have been generalized in various ways [2,3,8,9]. Some [5] of rough set approaches treat imprecise data in decision tables. Decision tables are composed of objects usually described by combinations of condition attribute values and a decision attribute value. The approaches to imprecision seem to be mainly toward condition attribute values. Indeed, precise decision attribute values are usually obtained.

However, in the real world, we come across cases when we only obtain data with imprecise decision attribute values. For example, evaluation of the economic situation would be difficult to tell precisely. Failure diagnosis of complex systems would start from the expert hunch or conjecture. The conjectured source of failure is a decision attribute value and it would be imprecise. Consider a forecast, it would be difficult to be exact and precise. Some tolerance would be necessary. Moreover, evaluations by humans are often imprecise. Even data with imprecise decision attribute values would be useful to induce rough knowledge or to find the condition attributes possibly to effect on the decision attribute value. It is much more informative than ignorance. Utilization and analyzing such data is valuable unless a sufficient number of precise data are available.

From this point of view, Inuiguchi and Li [4] have proposed a rough set approach to imprecise decision tables, i.e., decision tables with imprecise decision attribute values. They proposed an active approach so that different pieces of information can cooperate to obtain useful knowledge. Using the defined rough sets under an imprecise decision table, they investigated attribute reduction. However, rule induction has not yet been proposed.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 68–76, 2009.

Table 1. The imprecise decision table

Object	c_1	c_2	$\cdots$	c_m	d
u_1	$c_1(u_1)$	$c_2(u_1)$	$\cdots$	$c_m(u_1)$	$F(u_1)$
u_2	$c_1(u_2)$	$c_2(u_2)$	$\cdots$	$c_m(u_2)$	$F(u_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_n	$c_1(u_n)$	$c_2(u_n)$	$\cdots$	$c_m(u_n)$	$F(u_n)$

In this paper, we investigate rule induction under imprecise decision tables. We adopt the definition of rough sets proposed by Inuiguchi and Li [4]. As several attribute reduction schemes are considered in Inuiguchi and Li [4], several rule induction schemes are conceivable. For each rule induction scheme, we propose a decision matrix method.

This paper is organized as follows. In next section, imprecise decision tables are introduced. A few underlying object sets of rule induction are defined and their properties are described. In Section 3, based on the object sets defined in the previous section, rule induction schemes are proposed. Then decision matrix method in each rule induction scheme is given. In Section 4, concluding remarks are given.

2 Rough Sets under Imprecise Decision Tables

2.1 Imprecise Decision Tables

In this paper, we treat imprecise decision tables, i.e., decision tables with imprecise decision values shown in Table 1. An imprecise decision table is represented by a quadruple $\mathcal{I} = (U, C \cup \{d\}, F, V)$. U is a finite set of objects, $U = \{u_1, u_2, \ldots, u_n\}$. C is a finite set of condition attributes, $C = \{c_1, c_2, \ldots, c_m\}$. Each attribute c_i can be seen as a function from U to V_{c_i}, where V_{c_i} is the domain of condition attribute c_i. The function value $c_i(u_j)$ indicates the attribute value of u_j. d is a decision attribute whose value $d(u_j)$ is unique for each object u_j. F is a set-valued function from U to 2^{V_d}, where 2^{V_d} is a power set of V_d and V_d is the domain of decision attribute d. $F(u_j)$ indicates a set of possible decision attribute values of u_j. Finally, $V = \bigcup_{c \in C} V_c \cup V_d$.

In Table 1, decision attribute values are allowed to take set-values, while in the conventional decision tables, decision attribute values should be singletons, i.e., single values. Each set-value of decision attribute shows possible decision attribute values of the object. Such set-value may be obtained when decision attribute values are not specified precisely. For example, when our knowledge is not complete but partial or when we hesitate our evaluation, we may specify the decision value such as "not d_1" or "d_1 or d_2". In those cases, decision attribute values can be treated as imprecise values. Even imprecise decision values would be more useful than no information and a number of imprecise decision values may collaborate to obtain precise value and useful result. Imprecise decision

attribute values can be also regarded as values by conjecture. Then allowing imprecise decision attribute values enables us to analyze data by conjectures.

2.2 Rough Sets under Imprecise Decision Tables

Considering the imprecise nature of decision attribute values of imprecise decision tables, we define generalized decision values $\delta_P(u_j)$ and aggregated decision values $\hat{F}_P(u_j)$ under a given condition attribute set $P \subseteq C$ as follows [4]:

$$\delta_P(u_j) = \{F(u) \mid u \in U,\ c_i(u) = c_i(u_j),\ \forall c_i \in P\}, \tag{1}$$

$$\hat{F}_P(u_j) = \begin{cases} \bigcap \delta_P(u_j), \text{ if } \bigcap \delta_P(u_j) \neq \emptyset, \\ \bigcup \delta_P(u_j), \text{ otherwise.} \end{cases} \tag{2}$$

$\delta_P(u_j)$ collects imprecise decision values $F(u)$ of all objects u taking same condition attribute values with respect to $P \subseteq C$ as u_j takes. Since we assume the true decision attribute value of u_j is in $F(u_j)$ and the same decision attribute value would be assigned for all objects which share same condition attribute values, we may obtain a smaller possible range for decision attribute value of u_j by intersecting $F(u)$'s of all such objects u. However, if the given data is not totally consistent, the intersection can be empty. If the intersection is empty, some of $F(u)$ in the given table would be wrong or some condition attribute would be missing. In this case, the union would show the possible range. Based on these ideas, $\hat{F}_P(u_j)$ is defined. Taking union when the intersection is empty set is similar to Dubois and Prade's combination rule [1] in evidence theory.

In [4], lower and upper object sets are defined under imprecise decision tables. However, we use the following underlying object sets of rule induction:

$$Conf_P = \left\{u \in U \mid \bigcap \delta_P(u) = \emptyset\right\}, \tag{3}$$

$$P_*(X) = \{u \in U \mid \hat{F}_P(u) \subseteq X\}, \tag{4}$$

$$CS_P(X) = \{u \in U - Conf_P \mid \hat{F}_P(u) \supseteq X\}, \tag{5}$$

$$CR_P(X) = \{u \in Conf_P \mid \hat{F}_P(u) \cap X = \emptyset\}, \tag{6}$$

where $P \subseteq C$ and $X \subseteq V_d$. $Conf_P$ is a set of conflicting objects. $P_*(X)$ is a lower object set of X. If $u \in P_*(X)$, the decision attribute value of u is in X with no conflict with given data. In other words, if $u \in P_*(X)$, the decision attribute value of u is surely in X as far as the given decision table is correct. $CS_P(X)$ is the set of objects which consistently support the possible realizations of their decision attribute values in X. In other words, the possible realization of decision attribute value $x \in X$ for $u \in CS_P(X)$ is consistently confirmed by the given decision table. $CR_P(X)$ is the set of objects which consistently reject the possible realizations of their decision attribute values in X. In other words, no object in the given decision table supports the realization of decision attribute value $x \in X$ for $u \in CR_P(X)$.

Table 2. A table of failure conjectures by users

conjecture	(case,user)	f1	f2	f3	f4	f5	cause	$\hat{F}_C$
u_1	(P_1,E_1)	yes	yes	no	no	yes	{B, C}	{B, C}
u_2	(P_1,E_2)	yes	yes	no	no	yes	{A, B, C}	{B, C}
u_3	(P_2,E_1)	yes	yes	no	yes	yes	{A, B}	{A, B}
u_4	(P_3,E_2)	yes	yes	yes	yes	yes	{A}	{A, C}
u_5	(P_4,E_1)	yes	yes	no	no	no	{C}	{C}
u_6	(P_3,E_1)	yes	yes	yes	yes	yes	{C}	{A, C}
u_7	(P_5,E_2)	no	yes	yes	yes	yes	{A, C}	{A, C}
u_8	(P_6,E_2)	yes	no	no	yes	no	{B}	{B}

Example 1. Consider Table 2 showing conjectures u_1, u_2, ..., u_8 by two users E_1 and E_2 about failure causes of 6 cases P_1, P_2, ..., P_6 in a complex system from 5 functions f1, f2, ..., f5. We assume that users E_1 and E_2 have difference experiences so that their conjectures can be different even for the same case. There are three possible causes A, B and C. In this table, $U = \{u_i \mid i = 1, 2, \ldots, 8\}$, $C = \{\mathrm{f1}, \mathrm{f2}, \mathrm{f3}, \mathrm{f4}, \mathrm{f5}\}$, $V_a = \{\mathrm{yes}, \mathrm{no}\}$ for $a = \mathrm{f1}, \ldots, \mathrm{f5}$ and $V_{\mathrm{cause}} = \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$. Then $V = \{\mathrm{yes}, \mathrm{no}, \mathrm{A}, \mathrm{B}, \mathrm{C}\}$. The second column of Table 2 shows a pair (P_i,E_i) of case P_i and user E_i. The pair shows that the failure cause of P_i is conjectured by E_i and the result is shown in the column of "cause". For each conjecture u_i, $\hat{F}_C(u_i)$ is shown in the rightmost column of Table 2. The set of conflicting objects, some of lower object sets, consistently supporting object sets and consistently rejecting object sets are obtained as

$Conf_C = \{u_4, u_6\}$, $C_*(\{\mathrm{B}, \mathrm{C}\}) = \{u_1, u_2, u_5, u_8\}$, $CS_C(\{\mathrm{B}, \mathrm{C}\}) = \{u_1, u_2\}$,
$CR_C(\{\mathrm{B}, \mathrm{C}\}) = \emptyset$, $C_*(\{\mathrm{A}, \mathrm{C}\}) = \{u_4, u_5, u_6, u_7\}$, $CS_C(\{\mathrm{A}, \mathrm{C}\}) = \{u_4, u_6, u_7\}$,
$CR_C(\{\mathrm{A}, \mathrm{C}\}) = \{u_8\}$, $C_*(\{\mathrm{B}\}) = \{u_8\}$, $CS_C(\{\mathrm{B}\}) = \{u_1, u_2, u_3, u_8\}$
$CR_C(\{\mathrm{B}\}) = \{u_4, u_6\}$.

Before discussion about rule induction, we note that we have $\bigcap \delta_P(X) \subseteq \bigcap \delta_Q(X)$ but $\bigcup \delta_P(X) \supseteq \bigcup \delta_Q(X)$ for $P \subseteq Q \subseteq C$. Therefore, $P_*(X) \subseteq Q_*(X)$ does not always hold for $P \subseteq Q \subseteq C$. On the contrary, we always have $Conf_P \supseteq Conf_Q$ and $\delta_P(u) \subseteq \delta_Q(u)$ for $P \subseteq Q \subseteq C$ and $u \in U$. Moreover we have the following properties:

$$CS_P(X) = \bigcap_{x \in X} CS_P(\{x\}), \qquad CR_P(X) = \bigcup_{x \in X} CR_P(\{x\}). \tag{7}$$

The properties of $P_*(X)$ are shown in Inuiguchi and Li [4].

3 Rule Induction from Imprecise Decision Tables

3.1 Rule Induction Schemes

Given an imprecise decision table, we induce minimal rules. $\hat{F}_C(u_i)$ for $u_i \notin Conf_C$ is regarded as the certain range of decision attribute values which has no conflict

in the given imprecise decision table. On the contrary, $\hat{F}_C(u_i)$ for $u_i \in Conf_C$ is regarded as the possible range of decision attribute values which is supported by at least one object in the given imprecise decision table. Let $Cond(u)$ be a condition for an object. Corresponding to the two ranges described above, we consider the following rule induction schemes:

(a) Inducing certain range rules: for $u_i \in C_*(\hat{F}_C(u_i))$ such that $u_i \in U - Conf_C$, we induce minimal rules, "if $Cond(u)$ is satisfied then the certain range of $d(u)$ is $\hat{F}_C(u_i)$". If the premise of this rule is satisfied for an arbitrary object u, we may infer that any value in $\hat{F}_C(u_i)$ is highly conceivable for $d(u)$.
(b) Inducing possible range rules: for $u_i \in C_*(\hat{F}_C(u_i))$ such that $u_i \in Conf_C$, we induce minimal rules, "if $Cond(u)$ is satisfied then the possible range of $d(u)$ is $\hat{F}_C(u_i)$". If the premise of this rule is satisfied for an arbitrary object u, we may infer that any value in $\hat{F}_C(u_i)$ is possible for $d(u)$.

Moreover, from $CS_C(\{x\})$ and $CR_C(\{x\})$ for $x \in V_d$, we consider the following rule induction schemes, respectively:

(c) Inducing consistently supported value rules: for $x \in V_d$, we induce minimal rules, "if $Cond(u)$ is satisfied then $d(u)$ can take x". If the premise of this rule is satisfied for an arbitrary object u, we may infer that $d(u)$ can be x.
(d) Inducing consistently rejected value rules: for $x \in V_d$, we induce minimal rules, "if $Cond(u)$ is satisfied then $d(u)$ will not take x". If the premise of this rule is satisfied for an arbitrary object u, we may infer that $d(u)$ will not be x.

Schemes (a) and (b) are similar to schemes (c) and (d), respectively. However, schemes (c) and (d) are more active. For example, let $\hat{F}_C(u_i) = \{x_1, x_2, x_3\}$ for $u_i \in C_*(\hat{F}_C(u_i))$ such that $u \notin Conf_C$. In scheme (a), the minimal condition for x_1, x_2 and x_3 to be simultaneously conceivable is selected as the premises. On the contrary, in scheme (c), the minimal condition for one of x_1, x_2 and x_3 to be conceivable is selected as the premises. Similar explanation of the difference can be applied to schemes (b) and (d).

3.2 Decision Matrix Methods

In the classical rough set approach, all rules are induced by the decision matrix method [7]. In the method, all rules are obtained as prime implicants of a Boolean function obtained from a decision matrix.

In this subsection, we extend the decision matrix method [7] to the case of imprecise decision tables. To this end, we investigate a decision matrix associated with each rule induction scheme.

First let us discuss a decision matrix for scheme (a), i.e., certain range rules. Let $u_i \in C_*(\hat{F}_C(u_i))$ such that $u_i \in U - Conf_C$. For each $u_j \in U$, we define the following component:

$$D_{ij}^{\mathrm{cer}} = \begin{cases} \{(c_k, c_k(u_i)) \mid c_k(u_i) \neq c_k(u_j)\}, & \text{if } u_j \in Conf_C \text{ or } \hat{F}_C(u_j) \neq \hat{F}_C(u_i), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{8}$$

Regarding each element $(c_k, c_k(u_i))$ of D_{ij}^{cer} as a statement "$c_k(u) = c_k(u_i)$", all minimal conditions for $\hat{F}_C(u_i)$ to be the certain range are obtained as prime implicants of the following Boolean function:

$$f_i^{\mathrm{cer}} = \bigwedge_{u_j \in U} \bigvee D_{ij}^{\mathrm{cer}}, \tag{9}$$

where if D_{ij}^{cer} is an empty set, it is regarded as a tautology. Then all minimal rules inferring the certain range as $\hat{F}_C(u_i)$ are obtained by putting all prime implicants in the premises of rules.

Repeating this rule induction procedure for each $u_i \in C_*(\hat{F}_C(u_i))$ such that $u_i \in U - Conf_C$, we obtain a body of rules. If an unseen object u satisfies premises of several rules with different certain ranges, $\hat{F}_C(u_i)$, $i = 1, 2, \ldots, q$, then the values in $\bigcap_{i=1,2,\ldots,q} \hat{F}_C(u_i)$ are intuitively most conceivable if they exist. Otherwise, certain range cannot be estimated. The values in $\bigcup_{i=1,2,\ldots,q} \hat{F}_C(u_i)$ are possible.

Now let us discuss a decision matrix for scheme (b), i.e., possible range rules. Let $u_i \in C_*(\hat{F}_C(u_i))$ such that $u_i \in Conf_C$. For each $u_j \in U$, we define the following component:

$$D_{ij}^{\mathrm{pos}} = \begin{cases} \{(c_k, c_k(u_i)) \mid c_k(u_i) \neq c_k(u_j)\}, & \text{if } u_j \in U - Conf_C, F_C(u_j) \not\subseteq \hat{F}_C(u_i), \\ \{(c_k, c_k(u_i)) \mid c_k(u_i) \neq c_k(u_j)\}, & \text{if } u_j \in Conf_C, \hat{F}_C(u_j) \neq \hat{F}_C(u_i), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{10}$$

Regarding each element $(c_k, c_k(u_i))$ of D_{ij}^{pos} as a statement "$c_k(u) = c_k(u_i)$", all minimal conditions for $\hat{F}_C(u_i)$ to be the possible range are obtained as prime implicants of the following Boolean function:

$$f_i^{\mathrm{pos}} = \bigwedge_{u_j \in U} \bigvee D_{ij}^{\mathrm{pos}}, \tag{11}$$

where if D_{ij}^{pos} is an empty set, it is regarded as a tautology. Then all minimal rules inferring the possible range as $\hat{F}_C(u_i)$ are obtained by putting all prime implicants in the premises of rules.

Repeating this rule induction procedure for each $u_i \in C_*(\hat{F}_C(u_i))$ such that $u \in Conf_C$, we obtain a body of rules. If an unseen object u satisfies premises of several rules with different possible ranges, $\hat{F}_C(u_i)$, $i = 1, 2, \ldots, q$, then the estimated possible range of $d(u)$ becomes the union, $\bigcup_{i=1,2,\ldots,q} \hat{F}_C(u_i)$.

Let us discuss a decision matrix for scheme (c), i.e., consistently supported value rules. Let $x \in V_d$. For each $u_i \in CS_C(\{x\})$ and for each $u_j \in U - CS_C(\{x\})$, we define the following (i, j)-component of decision matrix $D^{\mathrm{cs}}(x)$:

$$D_{ij}^{\mathrm{cs}}(x) = \begin{cases} \{(c_k, c_k(u_i)) \mid c_k(u_i) \neq c_k(u_j)\}, & \text{if } u_j \in Conf_C \text{ or } x \notin \hat{F}_C(u_j), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{12}$$

Regarding each element $(c_k, c_k(u_i))$ of $D_{ij}^{\rm cs}(x)$ as a statement "$c_k(u) = c_k(u_i)$", all minimal conditions for $x \in V_d$ to be a conceivable value are obtained as prime implicants of the following Boolean function:

$$f^{\rm cs}(x) = \bigvee_{u_i \in CS_C(\{x\})} \bigwedge_{u_j \in U} \bigvee D_{ij}^{\rm cs}(x), \tag{13}$$

where if $D_{ij}^{\rm cs}(x)$ is an empty set, it is regarded as a tautology. Then all minimal rules inferring x as a conceivable value are obtained by putting all prime implicants in the premises of rules.

Finally, let us discuss a decision matrix for scheme (d), i.e., consistently rejected value rules. Let $x \in V_d$. For each $u_i \in CR_C(\{x\})$ and for each $u_j \in U - CR_C(\{x\})$, we define the following (i,j)-component of decision matrix $D^{\rm cr}(x)$:

$$D_{ij}^{\rm cr}(x) = \begin{cases} \{(c_k, c_k(u_i)) \mid c_k(u_i) \neq c_k(u_j)\}, & \text{if } x \in \hat{F}_C(u_j), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{14}$$

Regarding each element $(c_k, c_k(u_i))$ of $D_{ij}^{\rm cr}(x)$ as a statement "$c_k(u) = c_k(u_i)$", all minimal conditions for $x \in V_d$ to be a rejected value are obtained as prime implicants of the following Boolean function:

$$f^{\rm cr}(x) = \bigvee_{u_i \in CR_C(\{x\})} \bigwedge_{u_j \in U} \bigvee D_{ij}^{\rm cr}(x), \tag{15}$$

where if $D_{ij}^{\rm cr}(x)$ is an empty set, it is regarded as a tautology. Then all minimal rules inferring x as a rejected value are obtained by putting all prime implicants in the premises of rules.

Example 2. Let us apply the proposed rule induction schemes to Table 2. From u_1, let us induce certain range rules. $D_{1j}^{\rm cer}$ is obtained as in Table 3. Then the following certain range rule is obtained:

$$\text{if f4}(u) = \text{no and f5}(u) = \text{yes then the certain range is } \{\text{B}, \text{C}\}.$$

From u_4, let us induce possible range rules. $D_{44}^{\rm pos}$, $D_{46}^{\rm pos}$ and $D_{47}^{\rm pos}$ are empty and then omitted. The other $D_{4j}^{\rm pos}$, $j = 1, 2, 3, 5, 8$ are obtained as in Table 4. Then the following possible range rule is obtained:

$$\text{if f3}(u) = \text{yes then the possible range is } \{\text{A}, \text{C}\}.$$

Moreover, let us induce the consistently supported rules with respect to B. $D^{\rm cs}(\text{B})$ is obtained as in Table 5. Then the following consistently supported rules with respect to B are obtained:

$$\begin{array}{l} \text{if f3}(u) = \text{no and f5}(u) = \text{yes then } d(u) \text{ can be B}, \\ \text{if f4}(u) = \text{no and f5}(u) = \text{yes then } d(u) \text{ can be B}, \\ \text{if f3}(u) = \text{no and f4}(u) = \text{yes then } d(u) \text{ can be B}, \\ \text{if f2}(u) = \text{no then } d(u) \text{ can be B}, \\ \text{if f4}(u) = \text{yes and f5}(u) = \text{no then } d(u) \text{ can be B}. \end{array}$$

Table 3. D_{1j}^{cer}

	u_1, u_2	u_3	u_4, u_6	u_5	u_7	u_8
u_1	$\emptyset$	{(f4,no)}	{(f3,no),(f4,no)}	{(f5,yes)}	{(f1,yes),(f3,no),(f4,no)}	{(f2,yes),(f4,no)(f5,yes)}

Table 4. D_{4j}^{pos}

	u_1, u_2	u_3	u_5	u_8
u_4	{(f3,yes),(f4,yes)}	{(f3,yes)}	{(f3,yes),(f4,yes),(f5,yes)}	{(f2,yes),(f3,yes),(f5,yes)}

Table 5. $D^{\mathrm{cs}}(\mathrm{B})$

	u_1, u_2, u_3, u_8	u_4, u_6	u_5	u_7
u_1, u_2	{(f3,no),(f4,no)}	{(f5,yes)}	{(f3,no),(f4,no)}	{(f1,yes),(f3,no),(f4,no)}
u_3	{(f3,no)}	{(f4,yes),(f5,yes)}	{(f3,no)}	{(f1,yes),(f3,no)}
u_8	{(f2,no),(f3,no),(f5,no)}	{(f2,no),(f4,yes)}	{(f2,no),(f3,no),(f5,no)}	{(f1,yes),(f2,no),(f3,no),(f5,no)}

Finally, let us induce the consistently rejected rules with respect to B. $D^{\mathrm{cr}}(\mathrm{B})$ is omitted because its non-empty components are shown in (u_1, u_2), u_3 and u_8 columns of Table 4. The following consistently rejected rule with respect to B is obtained:

$$\text{if f3}(u) = \text{yes then } d(u) \text{ will not take B.}$$

4 Concluding Remarks

We have investigated rule induction from imprecise decision tables. Four induction schemes have been proposed. It is shown that rules in those induction schemes can be obtained by decision matrix methods. The applications to real world data as well as the other conceivable approaches to imprecise decision tables would be studied in near future.

Acknowledgments

This work was supported by JSPS KAKENHI No.17310098, No.18651078 and No.19300074.

References

1. Dubois, D., Prade, H.: Representation and combination of uncertainty with belief functions and possibility measures. Computational Intelligence 4, 244–264 (1988)
2. Greco, S., Matarazzo, B., Slowinski, R.: Rough set theory for multicriteria decision analysis. European Journal of Operational Research 129, 1–47 (2001)

3. Inuiguchi, M.: Generalization of rough sets and rule extraction. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 96–119. Springer, Heidelberg (2004)
4. Inuiguchi, M., Li, B.: Rough set approach to information tables with imprecise decisions. In: Chan, C.-C., Grzymala-Busse, J.W., Ziarko, W.P. (eds.) RSCTC 2008. LNCS (LNAI), vol. 5306, pp. 121–130. Springer, Heidelberg (2008)
5. Kryszkiewicz, M.: Rough set approach to incomplete information systems. Information Sciences 112, 39–49 (1998)
6. Pawlak, Z.: Rough sets. International Journal of Information and Computer Sciences 11, 341–356 (1982)
7. Shan, N., Ziarko, W.: Data-based acquisition and incremental modification of classification rules. Computational Intelligence 11, 357–370 (1995)
8. Słowiński, R., Vanderpooten, D.: A Generalized Definition of Rough Approximations Based on Similarity. IEEE Trans. Data and Knowledge Engin. 12(2), 331–336 (2000)
9. Yao, Y.Y., Lin, T.Y.: Generalization of Rough Sets Using Modal Logics. Intelligent Automation and Soft Computing 2(2), 103–120 (1996)

Uninorm Based Fuzzy Network for Tree Data Structures

Angelo Ciaramella[1], Witold Pedrycz[2], and Alfredo Petrosino[1]

[1] Dept. of Applied Sciences University of Naples "Parthenope",
Isola C4, Centro Direzionale, I-80143, Napoli, Italy
{ciaramella,petrosino}@uniparthenope.it
[2] Dept. of Electrical and Computer Engineering
University of Alberta, Edmonton AB, Canada
pedrycz@ee.ualberta.ca

Abstract. The aim of this study is to introduce a fuzzy model to process structured data. A structured organization of information is typically required by symbolic processing. Most connectionist models assume that data are organized in a form of relatively simple structures such as vectors or sequences. In this work, we propose a connectionist model that can directly process labeled trees. The model is based on a new category of logic connectives and logic neurons that use the concept of uninorms. Uninorms are a generalization of t-norms and t-conorms used for aggregating fuzzy sets. Using a back-propagation algorithm we optimize the parameters of the model (relations and membership functions). The learning issues are presented and some experimental results obtained for synthetic realistic data, are reported.

Keywords: Graphical Models, Trees, Fuzzy Logic Connectives, t-norms (t-conorms), Uninorms, Structured Data.

1 Introduction

Recently, structured data is becoming more and more important in data mining and data analysis. Structured domains are characterized by complex patterns which are usually represented as lists, trees and graphs of variable sizes and complexity. Data can be naturally represented by tree or graph structures in several application areas, including proteomics and molecular biology, image analysis, scene description, software engineering, natural language processing, XML document retrieval and others.

While neural networks are able to classify static information or temporal sequences, the current state of the art does not allow the efficient classification of structures of different size. The standard way to approach the classification of structured data by using a neural network is to encode the tree or graph in a fixed-size vector. In detail, it is encoded in a vector which is fed to a feedforward neural network for classification. The encoding process is usually defined *a priori* and does not depend on the classification task. The *a priori* definition of the encoding process has two main drawbacks

1. the relevance of different features of the graphs may change dramatically for different learning tasks.
2. each graph must get a different representation; this may result in vectorial representaions which are very difficult to classify.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 77–84, 2009.

To overcome the above difficulties, in [13] the authors propose to adapt the encoding process through an additional neural network which is trained to learn the best way to encode the graphs for the given classification task. To achieve this aim, they introduce a generalization of a recursive neuron. Moreover in [7] and [10] a Recursive Neural Network (RNN) for acyclic graphs was proposed. The RNN model can only process Directed Positional Acyclic Graphs (DPAGs). More recently, an extended model was proposed that can cope with non-positional acyclic graphs with labeled edges (i.e. DAGs) [2].

To solve the classification of labeled trees and handling encertainty, we propose a model based on fuzzy information. A fuzzy connectionist structure is developed from the input tree topology, by extending the capabilities of the RNN model. It is already demostrated [8] that it is possible to encode a nondeterministic fuzzy tree automata into a RNN. This encoding has been studied from a theoretical point of view, proving its stability. This theoretical study led to the model design reported in the present paper. Specifically, In the structure we propose the nodes are connected by using a relation and the composition of information is obtained by using an uninorm-based connective.

The paper is organized as follows. In Section 2 the notation about the graph adopted in the paper is introduced. In Section 3, we describe the uninorm-based generalization of norms and in Section 4 we introduce the Fuzzy Recursive Neural Model. In Section 5, we present several experimental results.

2 Graph and Tree Notation

A labeled graph (or graph) $\mathbf{G}$ is a quadruple ($\mathbf{N}$, $\mathbf{E}$, λ, ϵ), where $\mathbf{N}$ is the set of nodes (or vertices), and $\mathbf{E}$ is the set of edges between nodes, i.e. $\mathbf{E} \subseteq \{(u, v)|u, v \in \mathbf{N}\}$. Nodes and edges constitute the *skeleton* of the graph. The last two items, that associate vectors of real numbers of dimension respectively $\Re^{l_N}$ and $\Re^{l_E}$ to each node and edge, are respectively a node-labeling function $\lambda : \mathbf{N} \rightarrow \Re^{l_N}$ and an *edge-labeling* function $\epsilon : \mathbf{E} \rightarrow \Re^{l_E}$. Node labels are represented by $\mathbf{l}_n$ and edge labels by $\mathbf{l}_{uv}$. If the graph is directed, an edge (u, v) is an ordered pair of nodes, where u is the father and v its child. If the graph is undirected, the ordering between u and v in (u, v) is not defined, i.e. $(u, v) = (v, u)$. A graph is called acyclic if there is no path, i.e. a sequence of connected edges, that starts and ends at the same node. Combining some properties it is possible to specify various graph categories: Directed Positional Acyclic Graphs (DPAGs), Directed Acyclic Graphs (DAGs) and other graph-oriented structures. A tree $\mathbf{T}$ could be defined as an acyclic connected graph, represented by a quadruple ($\mathbf{N}$, $\mathbf{E}$, λ, ϵ), where each node has a set of zero or more children nodes and at most one parent node.

3 Uninorms

Triangular norms (t-norms) and the corresponding t-conorms play a fundamental role in several branches of mathematics [11], e.g., in probabilistic metric spaces, the theory of generalized measures, game theory, and fuzzy logic. The semantics of logic operators (logic connectives) in fuzzy sets is enormously rich. Some of the most recent conceptual

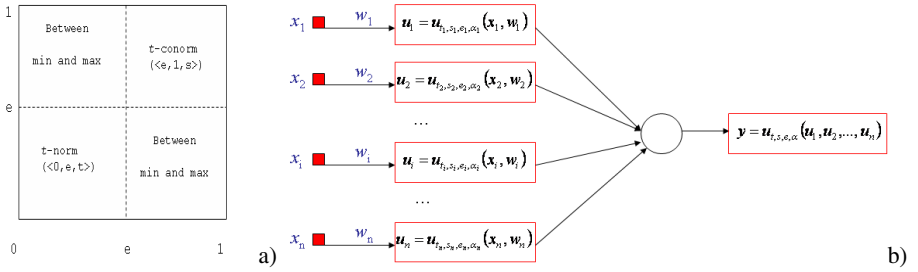


Fig. 1. Uninorm based structures : a) uninorm with neutral element e; b) unineuron

developments along this line involve uninorms [12,6,9,14], nullnorms [1] and ordinal sums [15] of t-norms, just to name a few of them.

We start our discussion about uninorms by recalling that both a neutral element 1 of a t-norm and the neutral element 0 of a t-conorm are boundary points of the unit interval. A *uninorm* is a binary operation $u : [0,1]^2 \to [0,1]$ which satisfies the properties of *Commutativity*, *Monotonicity*, *Associativity* and it has a *neutral element* $e \in [0,1]$ (see Figure 1a). Noticeably, we allow the values of the identity element e to vary in-between 0 and 1. As a result of this, we can implement switching between pure AND and OR properties of the logic operators occurring in this construct. In this study we confine ourselves to the following family of constructs that seem to be highly interpretative and thus intuitively appealing:

Let t be a t-norm, s be a t-conorm and $e \in [0,1]$. In the spirit of the construction of Ordinal Sums the following operation $\mathbf{u}_{t,s,e,\alpha} : [0,1]^2 \to [0,1]$ ($\alpha = \min$ or $\alpha = \max$) make $[0,1]$ into fully ordered semigroups with neutral element e:

$$\mathbf{u}_{t,s,e,\alpha}(x,y) = \begin{cases} e \cdot t(\frac{x}{e}, \frac{y}{e}) & \text{if } (x,y) \in [0,e]^2 \\ e + (1-e) \cdot s(\frac{x-e}{1-e}, \frac{y-e}{1-e}) & \text{if } (x,y) \in]e,1]^2 \\ \alpha(x,y) & \text{otherwise} \end{cases} \quad (1)$$

Obviously, $\mathbf{u}_{t,s,e,\min}$ is a conjunctive, and $\mathbf{u}_{t,s,e,\max}$ is a disjunctive uninorm.

Interestingly, we observe that the two intermediate regions deliver some flexibility to the specific realization of the uninorm [12].

3.1 Uninorm-Based Logic Neuron

The previous studies carried out in the realm of logic-based neurocomputing, we can distinguish between two general categories of neurons that are OR and AND neurons. We already proposed in literature extensions of these operators [5,4,12].

Let $\mathbf{x}$ be a vector in the unit hypercube, $\mathbf{x} \in [0,1]^n$ and y denote an element in $[0,1]$. Formally speaking, the underlying logic processing is governed by a composition between the individual inputs and the corresponding connections (weights) $\mathbf{w} \in [0,1]^n$. In detail, $L_1 : (x_i, w_i) \to [0,1]$ followed by some overall logic aggregation L_2 giving rise to the output y, that is

$$y = L_2[L_1(x_1,w_1), L_1(x_2,w_2), \ldots, L_1(x_n,w_n)] \quad (2)$$

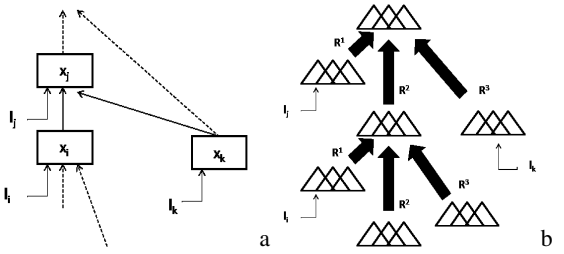


Fig. 2. Node connection: a) connection between nodes; b) fuzzyfication of the nodes

In the OR neuron, the semantics of L_1 is concerned with any realization of the AND operator (t-norm), $L_1 =$ AND. L_2 concerns the realization of the OR operator (s-norm), $L_2 =$ OR. On the contrary, in the AND neuron we have these two operations reversed, that is $L_1 =$ OR and $L_2 =$ AND.

By using the uninorms, that admit more flexibility into the overall aggregation, we lead to the expression (see Figure 1b)

$$y = u[u_1(x_1, w_1, e_1, \alpha_1), u_2(x_2, w_2, e_2, \alpha_2), \ldots, u_n(x_n, w_n, e_n, \alpha_n), e] \qquad (3)$$

where $u_i = \mathbf{u}_{t_i,s_i,e_i,\alpha_i}$ is the i-th uninorm with parameters t_i, s_i, e_i and α_i to be estimated by the optimization process. Moreover, $u = \mathbf{u}_{t,s,e,\alpha}$ is the uninorm that permits to obtain the overall composition [5].

4 Uninorm Based Fuzzy Network

In this Section we introduce the proposed Uninorm Based Fuzzy Network (UFN) model. Our framework for tree processing must implement a function φ that computes an output $\varphi(\mathbf{T})$ for each tree $\mathbf{T}$. Each node, that can be considered as a state, is described by M real attributes $\mathbf{x}_n$ where the dimension M is a predefined parameter. The state $\mathbf{x}_n$ of the n-th node is fuzzy: K membership functions are used to describe the node information. We denote with $\mu(\mathbf{x}_n)$ this fuzzyfied state. In Figure 2a and 2b we show a possible connection between the nodes $\mathbf{x}_j$, $\mathbf{x}_i$ and $\mathbf{x}_k$ with the corresponding fuzzy sets obtained from the fuzzyfication. The M attributes of the node are also fuzzified. As show in Figure 3 if we consider the n-th state then the M attributes $\mathbf{l}_n^1 \ldots \mathbf{l}_n^M$ are fuzzified obtaining the K fuzzy sets $\mu^1(\mathbf{l}_n^i) \ldots \mu^K(\mathbf{l}_n^i)$ for each attribute i. Now we stress that the membership functions $\mu(\mathbf{l}_n) = [\mu_1(\mathbf{l}_n) \ldots \mu_K(\mathbf{l}_n)]$ of the n-th node can be generated by uninorm-based neurons (as shown in Figure 3a). For this reason the composition of the k-th fuzzy set is

$$\mu_k(\mathbf{l}_n) = \bigvee_{i=1}^{M} \left(R_{att}(\mu^k(\mathbf{l}_n^i))_i^k t \mu^k(\mathbf{l}_n^i) \right) \qquad (4)$$

where $R_{att}(\mu^k(\mathbf{l}_n^i))_i^k$ is the weight (relation) between the k-th fuzzy set of the i-th attribute and the k-th membership function of $\mu(\mathbf{l}_n)$, t is a t-norm and $\bigvee$ is an s-norm depending from the uni-neuron definition. To compose the fuzzy state $\mu(\mathbf{x}_n)$

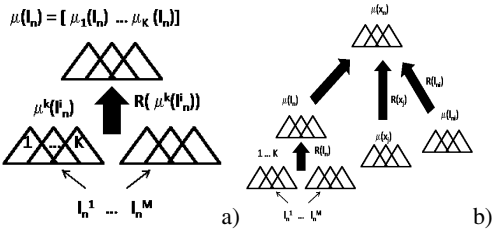


Fig. 3. Composition of the fuzzy sets for the node labels: a) attribute composition; b) node composition

we define a fuzzy relation between the k-th membership $\mu_k(\mathbf{l}_n)$ and the j-th fuzzy set $\mu_j(\mathbf{x}_n)$. We denote this relation as $R_{label}(\mu(\mathbf{l}_n))^j_k$. To simplify our derivation, in the following we consider $k = j$ (diagonal relation) and we denote the relation as $R_{label}(\mu(\mathbf{l}_n))^j$.

In the same way we can define the membership functions of the edge labels. If we consider $\mathbf{l}_{ni}$ to be the label between the states n and i, respectively, then we denote as $\mu_k(\mathbf{l}_{ni})$ the membership function of the k-th fuzzy set. In this case the weights of the relation become $R_{edge}(\mu(\mathbf{l}_{ni}))^j_k$ (also in this case for $k = j$, $R_{edge}(\mu(\mathbf{l}_{ni}))^j$).

Further relations are defined between the states. We denote as $\mu_k(\mathbf{x}_i)$ the k-th fuzzy set of the i-th state and as $R_{state}(\mu(\mathbf{x}_{ni}))^j_k$ (for $k = j$, $R_{state}(\mu(\mathbf{x}_{ni}))^j$) the relation between the k-th fuzzy set of the i-th state and the j-th fuzzy set of the n-th state.

Finally we obtain that the j-th fuzzy set of the n-th state can be estimate as the union of these relations

$\mu_j(\mathbf{x}_n) =$

$$\left(\bigvee_{i=1}^{|\mathbf{ch}[n]|} \left(R_{state}(\mu(\mathbf{x}_{ni}))^j t \mu_j(\mathbf{x}_i) \right) \right) \vee \left(\bigvee_{i=1}^{|\mathbf{ch}[n]|} \left(R_{edge}(\mu(\mathbf{l}_{ni}))^j t \mu_j(\mathbf{l}_{ni}) \right) \right) \vee \quad (5)$$

$$\left(\bigvee_{i=1}^{K} \left(R_{label}(\mu(\mathbf{l}_n))^j t \mu_j(\mathbf{l}_n) \right) \right). \quad (6)$$

To classify the graphs a defuzzyfication stage is added at the root state (or *super-source* $\mathbf{x}_0$). The defuzzyfication is obtained with a weighted sum of the memberships of the *super-source*

$$o_i = \sum_{k=1}^{K} w_k \mu_k(\mathbf{x}_0)) \quad (7)$$

where $\mu_k(\mathbf{x}_0)$ is the k-th membership of the *super-source*. The weights w_k are determined during the learning phase. The *super-source* is the only supervised node, i.e. with target $\mathbf{t}_i$ is assigned to it.

4.1 Optimization Process

In the learning process, the parameters to be estimated are the membership functions of both labels and edges, and the relations. To estimate those parameters we use a back-propagation algorithm. The output o_i can be calculated starting from the leaves and proceedings towards up to the super-source node, whereas the error is back-propagated from the root to the leaves of the UFN model. In our experiments, the optimization is obtained by minimizing a sum-of-squares error between the output of the model $o_i(\mathbf{T})$ and the target of the graphs t_i

$$E(\mathbf{T}) = \frac{1}{2}\sum_{i=1}^{N}(o_i(\mathbf{T}) - t_i)^2 \tag{8}$$

where n is the number of patterns in the training set. The error is successively back-propagated to each node. In the following, we show how we can obtain the gradient of this error w.r.t. the parameters to learn.

The partial derivative of the global error with respect to the weights w_k of the defuzzyfication is

$$\frac{\partial E(\mathbf{T})}{\partial w_k} = \frac{\partial E(\mathbf{T})}{\partial \mu_k(\mathbf{x}_0)}\frac{\partial \mu_k(\mathbf{x}_0)}{\partial w_k} = (o^i - t^i)\mu_k(\mathbf{x}_0) \tag{9}$$

To update the other parameters we need to calculate the gradient of the error w.r.t. this parameters and successively back-propagate the error from the top to the leaves. For example, let us consider the computation of the gradient at node 1 that is a child of the root. We consider the weight $R_{label}(\mu(\mathbf{l}_1))^j$ between the fuzzy set $\mu_j(\mathbf{l}_1)$ and the fuzzy set $\mu_j(\mathbf{x}_1)$. In this case the derivative is

$$\frac{\partial E(\mathbf{T})}{\partial R_{label}(\mu(\mathbf{l}_1))^j} = \frac{\partial E(\mathbf{T})}{\partial \mu(\mathbf{x}_1)^j}\frac{\partial \mu(\mathbf{l}_1)^j}{\partial R_{label}(\mu(\mathbf{l}_1))^j} = (o^i - t^i)w_j\frac{\partial \mu(\mathbf{x}_1)^j}{\partial R_{label}(\mu(\mathbf{l}_1))^j} \tag{10}$$

The second derivative in this expression depends on the definition of the uni-norms. Membership functions are Gaussian in our case. To learn the parameters of this memberships we apply the same back-propragation approach. If we consider the following j-th Gaussian membership of the i-th attribute

$$\mu^j(\mathbf{l}_1^i) = \exp\left(\frac{(\mathbf{l}_1^i - \mathbf{m}_i^j)^2}{(\sigma_i^j)^2}\right) \tag{11}$$

where $\mathbf{m}_i^j$ and σ_i^j are the mean and standard deviation, respectively, then the partial derivatives becomes

$$\frac{\partial E(\mathbf{T})}{\partial \mathbf{m}_i^j} == (o^i - t^i)w_j R_{label}(\mu(\mathbf{l}_1))^j\frac{\mathbf{l}_1^i - \mathbf{m}_i^j}{(\sigma_i^j)^2}\mu^j(\mathbf{l}_1^i) \tag{12}$$

and

$$\frac{\partial E(\mathbf{T})}{\partial \sigma_i^j} == (o^i - t^i)w_j R_{label}(\mu(\mathbf{l}_1))^j\frac{(\mathbf{l}_1^i - \mathbf{m}_i^j)^2}{(\sigma_i^j)^3}\mu^j(\mathbf{l}_1^i) \tag{13}$$

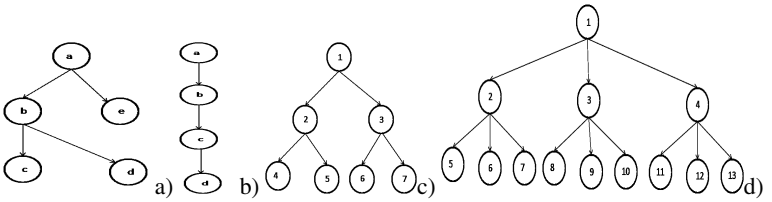


Fig. 4. Example of graphs: a) 4 nodes graph; b) 4 nodes graph; c) binary; d) ternary

5 Experimental Results

In this Section we investigate the performance of the proposed model to classify labeled trees. In a first experiment we analyze the capability of the model to discriminate two classes of trees. Examples of these trees are shown in Figures 4a and 4b, respectively. For both the graphs each node contains three attributes that identify a bi-dimensional position in the range $[0, 10]$ and a random attribute in the $[0, 1]$ interval. The objective of this experiment is to test the capability of the model to discriminate the trees using both label and positional information.

We created a data set of 500 trees (75% for the training and 25% for the test). The fuzzyfication is obtained by adopting 3 fuzzy Gaussian memberships and the composition of the uninorms is obtained by adopting the algebraic product as t-norm and the algebraic sum as t-conorm. We tested the model varying the neutral element e. We remark that the optimal performance were obtained varying e in the range $[0, 0.6]$, leading to 100% of perfect classification for training and test sets. For a neutral element higher than 0.6, in fact, the model is instable and a low rate of classification is achieved (near 70%).

In the second experiment we consider the same tree topology with the addition of edge labels. The UFN structure and parameters were set as in the previous experiment. The best performance were obtained varying e in the range $[0, 0.5]$, leading to 100% of perfect classification on training and test sets. This confirm also in this case that the uninorm composition must be OR-dominant.

In the third experiment we consider other two classes of trees. In this case the aim was to discriminate between binary and ternary trees. The parameters of the model are as in the previous experiments. In Figure 4c-d we show the topology of this trees. Each node contains 3 attributes that are randomly chosen in the interval $[0, 1]$. We generate 500 trees where the attributes are fixed but we randomly delete a node. With this experiment we wish to prove the capability of UFN to learn typologies of trees from corrupted patterns. Also in this case the best performance is obtained for OR-ness connections. We note, indeed, that varying e in the range $[0, 0.5]$ the performance is close to 100% of perfect classification on both training and test set. Moreover, also varying the attributes of the single nodes by adding a Gaussian random noise in the $[0, 0.1]$ range, the performance are comparable with that previous obtained.

6 Conclusions

In this work, a fuzzy recursive model based on uninorms to process structured data with a no flat representation has been reported. The model is based on a new category

of logic connectives and logic neurons based on the concept of uninorms. Uninorms are a generalization of t-norms and t-conorms used for composing fuzzy sets. To optimize the parameters of the model we use an encoding network and a back-propagation algorithm. The learning issues are reported to classify classes of labeled trees. Reported experiments show the capability of the model to approach structured data as in the case of tRNA structures, Region Adjacent Graph (RAG), HTML or XML page classification, and so on. Ongoing work is the generalization of the model for more complex structures described by labeled graphs and the validation of the model from a theoretical and practical point of view.

References

1. Calvo, T., Mesiar, R.: Continuous Generate Associative Aggregation Operators. Fuzzy Sets and Systems 126, 191–197 (2002)
2. Bianchini, M., Maggini, M., Martinelli, E., Sarti, L., Scarselli, F.: Recursive Neural Networks for Processing Graphs with Labelled Edges: Theory and Applications. Neural Networks 18, 1040–1050 (2005)
3. Ciaramella, A., Pedrycz, W., Tagliaferri, R.: The Genetic Development of Ordinal Sums. Fuzzy Sets and Systems 151, 303–325 (2005)
4. Ciaramella, A., Pedrycz, W., Tagliaferri, R.: OR/AND Neurons for Fuzzy Set Connectives Using Ordinal Sums and Genetic Algorithms. In: Bloch, I., Petrosino, A., Tettamanzi, A.G.B. (eds.) WILF 2005. LNCS (LNAI), vol. 3849, pp. 188–194. Springer, Heidelberg (2006)
5. Ciaramella, A., Pedrycz, W., Tagliaferri, R.: The Genetic Development of Uninorms Based Neurons. In: Masulli, F., Mitra, S., Pasi, G. (eds.) WILF 2007. LNCS, vol. 4578, pp. 69–76. Springer, Heidelberg (2007)
6. Fodor, J.C., Yager, R.R., Rybalov, A.: Structure of Uninorms. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 5, 411–427 (1997)
7. Frasconi, P., Gori, M., Sperduti, A.: A General Framework for Adaptive Processing of Data Structures. IEEE Transaction on Neural Networks 9(5), 768–786 (1998)
8. Gori, M., Petrosino, A.: Encoding Nondeterministic Fuzzy Tree Automata Into Recursive Neural Networks. IEEE Transactions on Neural Networks 15(6), 1435–1449 (2004)
9. Hirota, K., Pedrycz, W.: OR/AND Neuron in Modeling Fuzzy Set Connectives. IEEE Transaction on Fuzzy Systems 2, 151–161 (1994)
10. Kuchler, A., Goller, C.: Inductive Learning in Symbolic Domains Using Structure-Driven Recurrent Neural Networks. In: Görz, G., Hölldobler, S. (eds.) KI 1996. LNCS, vol. 1137, pp. 183–197. Springer, Heidelberg (1996)
11. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, Dordrecht (2001)
12. Pedrycz, W.: Logic-Based Fuzzy Neurocomputing with Unineurons. IEEE Transaction on Fuzzy Systems 14(6), 860–873 (2006)
13. Sperduti, A., Starita, A.: Supervised Neural Networks for the Classification of Structures. IEEE Transaction on Neural Networks 8(3) (1997)
14. Yager, R.R.: Uninorms in Fuzzy Systems Modeling. Fuzzy Sets and Systems 122, 167–175 (2001)
15. Zuckerman, M.-M.: Ordinal Sum-Sets. Proceedings of Americal Mathematical Society 35(1), 242–248

A Note on the Conditional Expectation of IF-Observables

Veronika Valenčáková

Matej Bel University, Tajovského 40, SK-974 01 Banská Bystrica, Slovakia
valencak@fpv.umb.sk

Abstract. The contribution deals with the probability theory on Intuitionistic Fuzzy Sets (IF-sets, [1]). It can be considered as a generalization of the classical probability theory on σ-algebras of sets. From the large field of operations with IF-sets we use Gödel connectives [5]. We focus on the conditional expectation: the aim is to formulate the definition of the conditional expectation of an IF-observable and prove the version of martingale convergence theorem.

Keywords: IF-event, conditional expectation, martingale convergence theorem.

1 Introduction

The classical probability theory has three basic terms: random variable, probability measure and expectation. The probability theory on IF-sets is a generalization of the probability theory on fuzzy sets. But in fuzzy probability theory there is no concept corresponding to random variables in Kolmogorov probability theory or to observables in the quantum structure probability theory. In IF-probability theory we have the corresponding notions to the probability (state), random variable (observable) and expectation in Kolmogorov probability theory. We follow the results in [3,6,7,10] and build up the conditional expectation.

Like in classical probability space we work with random event, also here we work with an IF-event. Let us consider a measurable space $(\Omega, \mathcal{S})$, i.e. a nonempty set Ω together with the σ-algebra $\mathcal{S}$ of its subsets. An IF-event is a couple of $\mathcal{S}$-measurable functions $A = (\mu_A, \nu_A)$ defined on a nonempty set Ω and with values in the unit interval, such that $\mu_A(x) + \nu_A(x) \leq 1$ holds for any $x \in \Omega$. (Recall that a function f is $\mathcal{S}$-measurable if $f^{-1}((-\infty, x)) \in \mathcal{S}$ for all $x \in R$.)

Denote the family of all IF-events by

$$\mathcal{F} = \{A = (\mu_A, \nu_A); \mu_A, \nu_A : \Omega \to [0,1]; \mu_A, \nu_A \text{ are } \mathcal{S}\text{-measurable}; \mu_A + \nu_A \leq 1_\Omega\}.$$

The ordering on $\mathcal{F}$ is defined by $A \leq B \iff \mu_A \leq \mu_B,\ \nu_A \geq \nu_B$, the largest element of $\mathcal{F}$ is the couple $(1_\Omega, 0_\Omega)$ and the smallest element is $(0_\Omega, 1_\Omega)$. Symbol 1_Ω denotes the function $f(x) \equiv 1$ on Ω (similarly for 0_Ω). By the notation $A_n \nearrow A$ we mean $\mu_{A_n} \nearrow \mu_A,\ \nu_{A_n} \searrow \nu_A$ (similarly for $A_n \searrow A$).

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 85–92, 2009.

Gödel operations [5] are for any $A, B \in \mathcal{F}$ defined by

$$A \vee B = (\mu_A \vee \mu_B,\ \nu_A \wedge \nu_B)\ ,$$
$$A \wedge B = (\mu_A \wedge \mu_B,\ \nu_A \vee \nu_B)\ .$$

Definition 1. *M-state is a mapping* $m : \mathcal{F} \to [0,1]$ *which satisfies*

1. $m((1_\Omega, 0_\Omega)) = 1,\ m((0_\Omega, 1_\Omega)) = 0,$
2. $m(A) + m(B) = m(A \vee B) + m(A \wedge B)$ *for any* $A, B \in \mathcal{F}$,
3. *if* $A_n \nearrow A,\ B_n \searrow B$ *then* $m(A_n) \nearrow m(A),\ m(B_n) \searrow m(B)$.

Definition 2. *M-observable is a mapping* $x : \mathcal{B}(R) \to \mathcal{F}$ *satisfying conditions*

1. $x(R) = (1_\Omega, 0_\Omega),\ x(\emptyset) = (0_\Omega, 1_\Omega),$
2. $x(C \cup D) = x(C) \vee x(D),\ x(C \cap D) = x(C) \wedge x(D)$ *for any* $C, D \in \mathcal{B}(R)$,
3. *if* $C, C_n, D, D_n \in \mathcal{B}(R)$ $(n = 1, 2, \dots),\ C_n \nearrow C,\ D_n \searrow D$ *then* $x(C_n) \nearrow x(C),\ x(D_n) \searrow x(D)$.

The composite mapping $m \circ x$ denoted by $m_x : \mathcal{B}(R) \to [0,1]$ is a probability measure (see [8, Proposition 1.4]).

To be able to define the indefinite integral we need to define the complement (denoted by $A^* = (\mu_{A^*}, \nu_{A^*})$) to an IF-event $A = (\mu_A, \nu_A) \in \mathcal{F}$. It is natural to require the following equalities to be satisfied: $A \wedge A^* = (0_\Omega, 1_\Omega), A \vee A^* = (1_\Omega, 0_\Omega)$. But that is possible only if we consider the crisp IF-events, i.e. the functions μ_A, ν_A are identical to the characteristic functions $\mu_A = \chi_A,\ \nu_A = \chi_{\bar{A}}$. In this case $A^* = (1 - \chi_A, \chi_A)$ will be the complement to $A = (\chi_A, 1 - \chi_A)$.

We will denote by $\mathcal{A}$ the family of all crisp IF-events $A = (\chi_A, 1 - \chi_A)$ in $\mathcal{F}$.

2 Conditional Expectation

We introduce the definition and some properties of the indefinite integral first. The motivation comes from the classical probability theory: $\int_C f\ d\mu = \int_X f\chi_C\ d\mu$.

Definition 3. *Let* $A \in \mathcal{A}$ *and* $x : \mathcal{B}(R) \to \mathcal{F}$ *be an M-observable. We define* $x_A : \mathcal{B}(R) \to \mathcal{F}$ *for any* $C \in \mathcal{B}(R)$ *by*

$$x_A(C) = \begin{cases} A \wedge x(C), & \text{if } 0 \notin C; \\ (A \wedge x(C)) \vee A^*, & \text{if } 0 \in C. \end{cases}$$

x_A is an M-observable (see Lemma 2 in Appendix).

Proposition 1. *If* x *is an integrable M-observable (i.e.* $\int_R t\ dm_x(t)$ *exists) then* x_A *is an integrable M-observable too.*

Proof. Let x be an integrable M-observable. We will prove the existence of integrals

$$\int_{(-\infty,0)} t\ dm_{x_A}(t), \qquad \int_{[0,\infty)} t\ dm_{x_A}(t)\ .$$

We define the mapping $\lambda : \mathcal{B}(R) \to R$ by $\lambda(C) = m_{x_A}(C \setminus \{0\}) = m(A \wedge x(C \setminus \{0\}))$. Then λ is a measure and

$$\lambda(C) = m(A \wedge x(C \setminus \{0\})) \leq m(A \wedge x(C)) \leq m(x(C)) = m_x(C) \ .$$

Then

$$0 \leq \int\limits_{[0,\infty)} t \ dm_{x_A}(t) = \int\limits_{\{0\}} t \ dm_{x_A}(t) + \int\limits_{(0,\infty)} t \ dm_{x_A}(t) =$$
$$= \int\limits_{(0,\infty)} t \ d\lambda(t) \leq \int\limits_{(0,\infty)} t \ dm_x(t) < \infty \ .$$

For $t < 0$ is clear that

$$0 > \int\limits_{(-\infty,0)} t \ dm_{x_A}(t) = \int\limits_{(-\infty,0)} t \ d\lambda(t) \geq \int\limits_{(-\infty,0)} t \ dm_x(t) > -\infty \ .$$

□

Definition 4. *Let $x : \mathcal{B}(R) \to \mathcal{F}$ be an integrable M-observable, $A \in \mathcal{A}$. The indefinite integral $\nu : \mathcal{A} \to R$ is defined by*

$$\nu(A) = E(x_A) = \int\limits_R t \ dm_{x_A}(t) \ .$$

Proposition 2. *Let x be an integrable M-observable and $y : \mathcal{B}(R) \to \mathcal{A}$ be an M-observable. Define the mapping $\kappa : \mathcal{B}(R) \to R$ by the formula $\kappa(C) = \nu(y(C)) = E(x_{y(C)})$. Then κ is a finite generalized measure.*

Proof. First we prove σ-additivity. Let $C = \bigcup\limits_{i=1}^{\infty} C_i$ where $C_i \in \mathcal{B}(R)$, $C_i \cap C_j = \emptyset$ for $i \neq j$. Then $y(C) = \bigvee\limits_{i=1}^{\infty} y(C_i)$ and by Lemma 3 (see Appendix) we get

$$\kappa(C) = \nu(y(C)) = \nu\left(\bigvee_{i=1}^{\infty} y(C_i)\right) = \sum_{i=1}^{\infty} \nu(y(C_i)) = \sum_{i=1}^{\infty} \kappa(C_i) \ .$$

Since $\lambda(C) = m(y(C) \wedge x(C \setminus \{0\})) \leq m_x(C)$, then

$$|\kappa(C)| = \left| \int\limits_R t \ d\lambda(t) \right| \leq \int\limits_R |t| \ d\lambda(t) \leq \int\limits_R |t| \ dm_x(t) < \infty. \qquad \square$$

Now we are able to define the conditional expectation of an integrable M-observable.

Proposition 3. *Let x, y be M-observables where x is an integrable M-observable and $y(C) \in \mathcal{A}$ for all $C \in \mathcal{B}(R)$. Then there exists a Borel function $g : R \to R$ such that*

$$\nu(y(C)) = \int_C g \, dm_y \tag{1}$$

for any $C \in \mathcal{B}(R)$.

Proof. Consider measures $\kappa = \nu \circ y$, m_y in probability space $(R, \mathcal{B}(R))$. Then

$$\kappa(C) = \nu(y(C)) = \int_R t \, dm_{x_{y(C)}}(t) \ .$$

We prove $\kappa \ll m_y$. Assume $m_y(C) = m(y(C)) = 0$ and $D \in \mathcal{B}(R)$ be arbitrary. We have

$$m_{x_{y(C)}}(D) = \begin{cases} m(y(C) \wedge x(D)), & 0 \notin D; \\ m([y(C) \wedge x(D)] \vee y^*(C)), & 0 \in D. \end{cases}$$

Evidently $0 \leq m(x_{y(C)}(D)) = m(y(C) \wedge x(D)) \leq m(y(C)) = 0$ and also

$$\begin{aligned} m(x_{y(C)}(D)) &= m([y(C) \wedge x(D)] \vee y^*(C)) = m([y(C) \vee y^*(C)] \wedge [x(D) \vee y^*(C)]) = \\ &= m((1_\Omega, 0_\Omega) \wedge [y^*(C) \vee x(D)]) = m(y^*(C) \vee x(D)) \geq m(y^*(C)) = 1 \ . \end{aligned}$$

But m is an M-state, i.e. the values are from $[0, 1]$. Therefore $m([y(C) \wedge x(D)] \vee y^*(C)) = 1$. We can see that $m_{x_{y(C)}}$ is a Dirac measure δ_0:

$$m_{x_{y(C)}}(D) = \delta_0(D) = \begin{cases} 0, & 0 \notin D; \\ 1, & 0 \in D. \end{cases}$$

Therefore

$$\kappa(C) = \int_R t \, dm_{x_{y(C)}}(t) = \int_{R \setminus \{0\}} t \, d\delta_0(t) + \int_{\{0\}} t \, d\delta_0(t) = 0 + 0\delta_0(\{0\}) = 0 \ .$$

The existence of an integrable function $g : R \to R$ (measurable on $\mathcal{B}(R)$) satisfying (1) for any $C \in \mathcal{B}(R)$ follows from the Radon-Nikodym theorem. □

Definition 5. *Let x, y be M-observables where x is an integrable M-observable and $y(C) \in \mathcal{A}$ for all $C \in \mathcal{B}(R)$. Let $C \in \mathcal{B}(R)$ be arbitrary. Conditional expectation $E(x|y)$ is a Borel function $g : R \to R$ which for every $C \in \mathcal{B}(R)$ satisfies*

$$E(x_{y(C)}) = \nu(y(C)) = \int_C g \, dm_y \ .$$

Remark 1. If g_1, g_2 satisfy the previous definition then $g_1 = g_2$ m_y-almost everywhere.

Now we are going to prove the martingale convergence theorem. For the properties of the joint M-observable see Definition 6 in Appendix. We recall that to each pair of M-observables x, y there exists their joint M-observable [7].

Proposition 4. *Let x, y be M-observables where x is an integrable M-observable and $y(C) \in \mathcal{A}$. Let h be the joint M-observable for M-observables x and y. Let ξ, η be random variables defined by $\xi(u,v) = u,\ \eta(u,v) = v$ in probability space $(\Omega, \mathcal{S}, P)$ where $\Omega = R^2, \mathcal{S} = \mathcal{B}(R^2), P = m \circ h$. Let P_η be the probability distribution of random variable η. Then $P_\eta = m_y$ and $E(x|y) = E(\eta|\xi)$ m_y-almost everywhere.*

Proof. From the definition of P_η follows

$$P_\eta(C) = P(\eta^{-1}(C)) = m \circ h \circ \eta^{-1}(C) = m(h(R \times C)) = m(x(R) \wedge y(C)) = m_y(C) \ .$$

Further $\int\limits_{\eta^{-1}(C)} \xi \ dP = \int\limits_{R \times C} u \ dm \circ h(u,v)$. The indefinite integral $\nu(y(C))$ can be expressed as

$$\nu(y(C)) = E(x_{y(C)}) = \int\limits_R u \ dm_{x_{y(C)}}(u) = \int\limits_R u \ d\lambda(u) \ ,$$

where $\lambda(D) = m(y(C) \wedge x(D)) = m \circ h(D \times C)$ for fixed $C \in \mathcal{B}(R)$.

We will prove the equality

$$\int\limits_{R \times C} g(u) \ dm \circ h(u,v) = \int\limits_R g \ d\lambda \tag{2}$$

for an arbitrary Borel function $g : R \to R$. Put $g = \chi_G$, $G \in \mathcal{B}(R)$. Then

$$\int\limits_{R \times C} \chi_G(u) \ dm \circ h(u,v) = \int\limits_{R \times R} \chi_{R \times C}(u,v) \chi_{G \times R}(u,v) \ dm \circ h(u,v) =$$
$$= \int\limits_{R \times R} \chi_{G \times C} \ dm \circ h = m(h(G \times C)) = m(x(G) \wedge y(C)) = \lambda(G) = \int\limits_R \chi_G \ d\lambda \ .$$

Equality (2) for an arbitrary Borel function g can be simply proved by using the linearity and continuity of integrals. Now put $g(u) = u$. Then

$$\int\limits_C E(x|y) \ dm_y = E(x_{y(C)}) = \int\limits_R u \ d\lambda(u) = \int\limits_{R \times C} g(u) \ dm \circ h(u,v) =$$
$$= \int\limits_{\eta^{-1}(C)} \xi \ dP = \int\limits_C E(\xi|\eta) \ dP_\eta = \int\limits_C E(\xi|\eta) \ dm_y$$

holds for any $C \in \mathcal{B}(R)$. Therefore $E(x|y) = E(\xi|\eta)$ m_y-almost everywhere. □

Proposition 5. *Let $g_n : R \to R$ be Borel functions such that $g_n^{-1}(\mathcal{B}(R)) \subset g_{n+1}^{-1}(\mathcal{B}(R))$ for each $n \in N$ and $\bigcup\limits_{n=1}^{\infty} g_n^{-1}(\mathcal{B}(R)) = \mathcal{B}(R)$. Let $y : \mathcal{B}(R) \to \mathcal{F}$ be an M-observable and let $y_n = y \circ g_n^{-1}$, $m_y \ll m_{y_n}$ for each $n \in N$. Let $x : \mathcal{B}(R) \to \mathcal{F}$ be an integrable M-observable. Then $E(x|y_n) \to E(x|y)$ m_y-almost everywhere.*

Proof. Consider the probability space $(R^2, \mathcal{B}(R^2), P = m \circ h)$ and let ξ, η be random variables defined by $\xi(u,v) = u,\ \eta(u,v) = v$. Put $\eta_n = g_n \circ \eta$ and $S_n = \eta^{-1}(g_n^{-1}(\mathcal{B}(R))) = \eta_n^{-1}(\mathcal{B}(R))$. Then $S_n \subset S_{n+1}$ for each $n \in N$ and moreover

$$\bigcup_{n=1}^{\infty} S_n = \bigcup_{n=1}^{\infty} \eta^{-1}(g_n^{-1}(\mathcal{B}(R))) = \eta^{-1}\left(\bigcup_{n=1}^{\infty} g_n^{-1}(\mathcal{B}(R))\right) = \eta^{-1}(\mathcal{B}(R)) = S \ .$$

Then from the martingale convergence theorem [9] follows $E(\xi|\eta_n) \to E(\xi|\eta)$ P_η-almost everywhere. Following Proposition 4 we have: $P_\eta = m_y,\ E(\xi|\eta) = E(x|y)$. Since for any $C \in \mathcal{B}(R)$

$$P_{\eta_n}(C) = P(\eta^{-1}(g_n^{-1}(C))) = m(h(R \times g_n^{-1}(C))) = m(x(R) \wedge y(g_n^{-1}(C))) = m_{y_n}(C),$$

then $P_{\eta_n} = m_{y_n}$. Furthermore

$$\int_C E(\xi|\eta_n)\ dP_{\eta_n} = \int_{\eta_n^{-1}(C)} \xi\ dP = \int_{\eta^{-1}(g_n^{-1}(C))} \xi\ dP = \int_{g_n^{-1}(C)} E(\xi|\eta)\ dP_\eta =$$
$$= \int_{g_n^{-1}(C)} E(x|y)\ dm_y = \int_{y(g_n^{-1}(C))} x\ dm = \int_C E(x|y_n)\ dm_{y_n} \ .$$

Since $P_{\eta_n} = m_{y_n}$ then $E(\xi|\eta_n) = E(x|y_n)$ m_{y_n}-almost everywhere. But $m_y \ll m_{y_n}$, therefore the equality $E(\xi|\eta_n) = E(x|y_n)$ holds m_y-almost everywhere. Finally $E(x|y_n) = E(\xi|\eta_n) \to E(\xi|\eta) = E(x|y)$ m_y-almost everywhere. □

Acknowledgements

The paper was supported by Grant VEGA 1/0539/08.

References

1. Atanassov, K.: Intuitionistic Fuzzy Sets: Theory and Applications. Physica Verlag, New York (1999)
2. Atanassov, K., Riečan, B.: On two operations over intuitionistic fuzzy sets. Journal of Applied Mathematics, Statistics and Informatics 2, 145–148 (2006)
3. Ćunderlíková-Lendelová, K., Riečan, B.: Intuitionistic fuzzy probability theory. In: Edited volume on Intuitionistic Fuzzy Sets: Recent Advances of the series Studies in Fuzziness and Soft Computing. Springer, Heidelberg (accepted)
4. Grzegorzewski, P., Mrówka, E.: Probability of intuistionistic fuzzy events. In: Grzegorzewski, P., et al. (eds.) Soft Methods in Probability, Statistics and Data Analysis, pp. 105–115. Physica Verlag, New York (2002)
5. Krachounov, M.: Intuitionistic probability and intuitionistic fuzzy sets. In: El-Darzi, E., et al. (eds.) First Intern. Workshop on IFS, Generalized Nets and Knowledge Engineering, pp. 18–24. Univ. of Westminster, London (2006)
6. Riečan, B.: General form of M-probabilities on IF-events. In: Magdalena, L., et al. (eds.) Proceedings of IPMU 2008 [CD-ROM]. Torremolinos (Málaga), pp. 1675–1677 (2008)

7. Riečan, B.: M-probability theory on IF events. In: Štepnička, M., et al. (eds.) New Dimensions in Fuzzy Logic and Related Technologies, vol. I, pp. 227–230. Universitas Ostraviensis, Ostrava (2007)
8. Riečan, B.: Probability theory on IF events. In: Aguzzoli, S., Ciabattoni, A., Gerla, B., Manara, C., Marra, V., et al. (eds.) ManyVal 2006. LNCS, vol. 4460, pp. 290–308. Springer, Heidelberg (2007)
9. Riečan, B., Neubrunn, T.: Integral, Measure, and Ordering. Kluwer Academic Publishers, Ister Science, Dordrecht, Bratislava (1997)
10. Mazureková, P., Valenčáková, V.: Conditional M-probability. In: Ruan, D., et al. (eds.) Computational Intelligence in Decision and Control. Proceedings of the 8th International FLINS Conference, Madrid, Spain, September 21-24, 2008, pp. 331–336. World Scientific, Madrid (2008)

Appendix

Lemma 1. *If* $A_i \in \mathcal{F},\ i = 1, 2, \ldots$ *and* $A_i \wedge A_j = (0_\Omega, 1_\Omega)$ *for* $i \neq j$ *then* $m\left(\bigvee_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i)$.

Proof. The proof is obvious. □

Lemma 2. x_A *is an M-observable.*

Proof. Evidently x_A satisfies the condition (1) of an M-observable.
(2) If $0 \notin C$ and $0 \notin D$ then $0 \notin C \cup D,\ 0 \notin C \cap D$ and we have

$$\begin{aligned} x_A(C \cup D) &= A \wedge x(C \cup D) = A \wedge (x(C) \vee x(D)) = \\ &= (A \wedge x(C)) \vee (A \wedge x(D)) = x_A(C) \vee x_A(D)\ , \\ x_A(C \cap D) &= A \wedge x(C \cap D) = A \wedge (x(C) \wedge x(D)) = \\ &= (A \wedge x(C)) \wedge (A \wedge x(D)) = x_A(C) \wedge x_A(D)\ . \end{aligned}$$

If $0 \in C$ and $0 \notin D$ then $0 \in C \cup D,\ 0 \notin C \cap D$ and we get

$$\begin{aligned} x_A(C \cup D) &= (A \wedge x(C \cup D)) \vee A^* = [(A \wedge x(C)) \vee (A \wedge x(D))] \vee A^* = \\ &= [(A \wedge x(C)) \vee A^*] \vee (A \wedge x(D)) = x_A(C) \vee x_A(D)\ , \\ x_A(C) \wedge x_A(D) &= [(A \wedge x(C)) \vee A^*] \wedge (A \wedge x(D)) = \\ &= [(A \wedge x(C)) \wedge (A \wedge x(D))] \vee [A^* \wedge (A \wedge x(D))] = \\ &= [A \wedge (x(C) \wedge x(D))] \vee [(A^* \wedge A) \wedge x(D)] = \\ &= (A \wedge x(C \cap D)) \vee ((0_\Omega, 1_\Omega) \wedge x(D)) = \\ &= (A \wedge x(C \cap D)) \vee (0_\Omega, 1_\Omega) = A \wedge x(C \cap D) = x_A(C \cap D)\ . \end{aligned}$$

Similarly we would prove the property (2) for $0 \notin C$ and $0 \in D$.
Finally if $0 \in C$ and $0 \in D$ then $0 \in C \cup D,\ 0 \in C \cap D$ and

$$\begin{aligned} x_A(C \cup D) &= (A \wedge x(C \cup D)) \vee A^* = [(A \wedge x(C)) \vee (A \wedge x(D))] \vee A^* = \\ &= [(A \wedge x(C)) \vee A^*] \vee [(A \wedge x(D)) \vee A^*] = x_A(C) \vee x_A(D)\ , \\ x_A(C \cap D) &= (A \wedge x(C \cap D)) \vee A^* = [A \wedge (x(C) \wedge x(D))] \vee A^* = \\ &= [(A \wedge x(C)) \wedge (A \wedge x(D))] \vee A^* = \\ &= [(A \wedge x(C)) \vee A^*] \wedge [(A \wedge x(D)) \vee A^*] = x_A(C) \wedge x_A(D)\ . \end{aligned}$$

(3) Let $C \subset D$. If $0 \in C$ then $0 \in D$ and we have

$$x_A(C) = (A \wedge x(C)) \vee A^* \leq (A \wedge x(D)) \vee A^* = x_A(D) \ .$$

If $0 \notin C$ and $0 \in D$ then $x_A(C) = A \wedge x(C) \leq (A \wedge x(D)) \vee A^* = x_A(D)$.
If $0 \notin C$ and $0 \notin D$ then $x_A(C) = A \wedge x(C) \leq A \wedge x(D) = x_A(D)$.

Let $C_n \nearrow C$ and $0 \notin C$. Then $0 \notin C_n$ for each n and we get

$$x_A(C_n) = A \wedge x(C_n) \nearrow A \wedge x(C) = x_A(C) \ .$$

Let $C_n \nearrow C$ and $0 \in C$. Then there exists n_0 such that $0 \in C_n$ for every $n \geq n_0$ and we have

$$x_A(C_n) = (A \wedge x(C_n)) \vee A^* \nearrow (A \wedge x(C)) \vee A^* = x_A(C). \qquad \square$$

Lemma 3. *Let x be an integrable M-observable. Then ν is σ-additive in the following sense: if $A_i \in \mathcal{A},\ (i = 1, 2, \dots),\ A_i \wedge A_j = (0_\Omega, 1_\Omega)$ for $i \neq j$ then*

$$\nu\left(\bigvee_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \nu(A_i) \ .$$

Proof. If $A_i \in \mathcal{A},\ (i = 1, 2, \dots),\ A_i \wedge A_j = (0_\Omega, 1_\Omega)$ for $i \neq j$, it is evident that

$$A = \bigvee_{i=1}^{\infty} A_i = \left(\bigvee_{i=1}^{\infty} \chi_{A_i}, \bigwedge_{i=1}^{\infty} (1 - \chi_{A_i})\right) = (\chi_{\cup A_i}, 1 - \chi_{\cup A_i}) \in \mathcal{A} \ .$$

When we put $\lambda(C) = m_{x_A}(C \setminus \{0\})$, $\lambda_i(C) = m_{x_{A_i}}(C \setminus \{0\})$ for any $C \in \mathcal{B}(R)$ then we can write $\nu(A) = \int\limits_R t \ d\lambda(t)$, $\nu(A_i) = \int\limits_R t \ d\lambda_i(t)$. Further $\lambda(C) = m\left(A \wedge x(C \setminus \{0\})\right) = m\left(\left(\bigvee_{i=1}^{\infty} A_i\right) \wedge x(C \setminus \{0\})\right) = m\left(\bigvee_{i=1}^{\infty} A_i \wedge x(C \setminus \{0\})\right)$. Since $(A_i \wedge x(C \setminus \{0\})) \wedge (A_j \wedge x(C \setminus \{0\})) = (0_\Omega, 1_\Omega)$ for $i \neq j$ then by Lemma 1

$$\lambda(C) = m\left(\bigvee_{i=1}^{\infty} A_i \wedge x(C \setminus \{0\})\right) = \sum_{i=1}^{\infty} m(A_i \wedge x(C \setminus \{0\})) = \sum_{i=1}^{\infty} \lambda_i(C) \ .$$

Therefore

$$\nu(A) = \int_R t \ d\lambda(t) = \sum_{i=1}^{\infty} \int_R t \ d\lambda_i(t) = \sum_{i=1}^{\infty} \nu(A_i). \qquad \square$$

Definition 6 ([7]). *Let x, y be M-observables. Joint M-observable is a mapping $h : \mathcal{B}(R^2) \to \mathcal{F}$ satisfying*

1. *$h(R^2) = (1_\Omega, 0_\Omega),\ h(\emptyset) = (0_\Omega, 1_\Omega)$,*
2. *$h(C \cup D) = h(C) \vee h(D),\ h(C \cap D) = h(C) \wedge h(D)$ for any $C, D \in \mathcal{B}(R^2)$,*
3. *if $C, C_n, D, D_n \in \mathcal{B}(R^2),\ C_n \nearrow C,\ D_n \searrow D$ then $h(C_n) \nearrow x(C),\ h(D_n) \searrow x(D)$,*
4. *$h(C \times D) = x(C) \wedge y(D)$ for any $C, D \in \mathcal{B}(R)$.*

A Survey on the Algebras of the So–Called Intuitionistic Fuzzy Sets (IFS)*

Gianpiero Cattaneo and Davide Ciucci

Dipartimento Di Informatica, Sistemistica e Comunicazione
Università di Milano – Bicocca, Viale Sarca 336 – U14, I–20126 Milano - Italy
{cattang,ciucci}@disco.unimib.it

Abstract. Some relevant algebraic structures involved by the so–called Intuitionistic Fuzzy Sets (IFS) are discussed, with a wide description of their relevant properties especially from the point of view of the algebraic semantic of a logical system. Algebraic comparison with analogous structures involving usual Fuzzy Sets are discussed.

1 Introduction

Presently, there is a wide interest in the scientific community of computer science about the so–called Intuitionistic Fuzzy Set (IFS) theory, introduced for the first time by Atanassov in his seminal papers [3], [1]. ¿From the foundational point of view IFS gave rise to an interesting terminological debate about the improper use of the term "intuitionistic" adopted by Atanassov in defining these kind of objects, especially related to some algebraic properties of the "negation" operation. The source of this debate can be dated back to the paper [15, p. 183] in which it is stressed that the Atanassov negation cannot be considered as a correct algebraic version of the intuitionist negation. This debate has continued in [6], [7], and successively in [19], [22] with an answer by Atanassov in [2] and finally in [10]. Another drawback of the original Atanassov exposition is that the involved operations are discussed in a "uniform" list, without a systematic organization of them inside an explicitly introduced algebraic structure. In the present paper, we adopt this last kind of exposition, unifying a scattered collection of information about this argument which can be found fragmented in [6], [7], [9], [10], [11].

2 The Logic–Algebraic Approach to Fuzzy Sets

Since IFS are nothing else than pairs of fuzzy sets (f, g) under the *orthogonality* condition $f + g \leq 1$, in order to make a deep discussion about IFS algebras it is necessary a brief exposition about the corresponding algebraic structures of fuzzy sets.

* This work has been supported by MIUR\PRIN project "Mathematical aspects and forthcoming applications of automata and formal languages".

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 93–100, 2009.

Let us recall that a fuzzy set (FS) on a universe X is mathematically realized by a mapping $f : X \mapsto [0,1]$, whose collection will be denoted by $\mathcal{F}(X)$ in the sequel. For any such mapping f one can introduce the *necessity* domain $\Delta_1(f) := \{x \in X : f(x) = 1\}$ with corresponding *contingency* domain $\Delta_c(f) := \{x \in X : f(x) \neq 1\} = \Delta_1(f)^c$ and the *possibility* domain $\Delta_p(f) := \{x \in X : f(x) \neq 0\}$ with corresponding *impossibility* domain $\Delta_0(f) := \{x \in X : f(x) = 0\} = \Delta_p(f)^c$. As particular subclasses of fuzzy sets we take into account *constant* fuzzy sets and *crisp* sets. The former are defined for any $k \in [0,1]$ as the mapping $\mathbf{k} : X \to [0,1]$ assigning to any point x of the universe the real value $\mathbf{k}(x) = k$. The latter consist of Boolean valued fuzzy sets $\chi : X \mapsto \{0,1\}$ which can be bijectively expressed as characteristic functionals χ_A of some subset A of X.

The first, immediate, algebraic structure one introduces on FS is the one of distributive lattice $\langle \mathcal{F}(X), \wedge, \vee, \mathbf{0}, \mathbf{1} \rangle$ with respect to the binary operations of lattice meet and join defined pointwise as $(f \wedge g)(x) := \min\{f(x), g(x)\}$ and $(f \vee g)(x) := \max\{f(x), g(x)\}$. This distributive (atomic complete) lattice is bounded by the least element $\mathbf{0} = \chi_\emptyset$ and the greatest element $\mathbf{1} = \chi_X$. Moreover, the induced partial order relation ($f \leq g$ iff $f = f \wedge g$, or equivalently $g = f \vee g$) is the pointwise one: $f \leq g$ iff $\forall x \in X,\ f(x) \leq g(x)$. These binary operations can be considered as the versions of the logical connectives of conjunction And and disjunction Or, respectively, in a context of the algebraic semantic of some (many–valued) logical systems. As to a possible algebraic realization of the logical Not connective the unary operation usually taken into account is the correspondence $f \to \neg f$ pointwise defined as follows: $(\neg f)(x) := 1 - f(x)$. This is the algebraic version of a *Kleene* negation since it satisfies the following Kleene conditions of an orthocomplementation (see [14]):

$(ocK1)$	$\neg\neg f = f$	double negation law
$(ocK2)$	$\neg(f \vee g) = \neg f \wedge \neg g$	first de Morgan law
$(ocK3)$	$f \wedge \neg f \leq \mathbf{1/2} \leq g \vee \neg g$	Kleene condition

Note that $\mathbf{1} = \neg\mathbf{0}$. Of course, neither the *noncontradiction law* $\forall f \in \mathcal{F}(X)$, $f \wedge \neg f = \mathbf{0}$ (for instance, $\mathbf{1/2} \wedge \neg\mathbf{1/2} = \mathbf{1/2} \neq \mathbf{0}$) nor the *excluded middle law* $\forall f \in \mathcal{F}(X)$, $f \vee \neg f = \mathbf{1}$ (for instance, $\mathbf{1/2} \vee \neg\mathbf{1/2} = \mathbf{1/2} \neq \mathbf{1}$) hold. As discussed in [5], [10] these results exclude the Kleene negation as a candidate for an intuitionistic negation since, for instance, intuitionistic logic accepts the noncontradiction principle which is not verified by the Kleene negation and reject the strong double negation law in the direction $\neg\neg f \leq f$, here accepted.

On the basis of the Kleene lattice $\langle \mathcal{F}(X), \wedge, \vee, \neg, \mathbf{0} \rangle$ another extension of Boolean negation can be taken into account: the *Brouwer* negation associating with any fuzzy set f the new fuzzy set $\sim f := \chi_{\Delta_0(f)}$, i.e., the crisp set of the impossibility domain of f explicitly defined by the pointwise rule

$$(\sim f)(x) = \begin{cases} 1 & \text{if } f(x) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

This is a real intuitionistic negation (see [10]) since it satisfies the conditions

$$\begin{array}{ll} (ocB1) & \sim f \leq \sim\sim f \qquad \text{weak double negation law} \\ (ocB2) & \sim (f \vee g) = \sim f \wedge \sim g \qquad \text{first de Morgan law} \\ (ocB3) & f \wedge \sim f = \mathbf{0} \qquad \text{noncontradiction law} \end{array}$$

Let us stress that the dual principle $\sim\sim f \leq f$ does not hold (for instance $\sim\sim \mathbf{1/2} = \mathbf{1}$), in agreement with the Heyting analysis of intuitionistic negation which rejects this condition [24]. Similarly, also the principle of excluded middle fails, $\forall f \in \mathcal{F}(X)$, $f \vee \sim f = \mathbf{1}$, according to the intuitionistic approach to logic. Further, let us stress that under (ocB1) the condition (ocB2) is equivalent to the intuitionistic contraposition law: $f \leq g$ implies $\sim g \leq \sim f$, but in general the second de Morgan law cannot be obtained as an equivalent formulation of (ocB2). However, in the present concrete case of fuzzy sets, it is possible to directly prove that the further second de Morgan law ($ocB4$) holds: $\sim (f \wedge g) = \sim f \vee \sim g$ which under (ocB1)–(ocB3) is equivalent (see [26]) to the *Stone condition*: (S) $\sim f \vee \sim\sim f = \mathbf{1}$. In this way, we have obtained a structure $\langle \mathcal{F}(X), \wedge, \vee, \neg, \sim, \mathbf{0} \rangle$, algebraic semantic of a fuzzy logic with the standard connectives of conjunction And ($\wedge$) and disjunction Or ($\vee$), equipped with a Kleene (also *Zadeh*) negation ($\neg$) and a *Brouwer* ($\sim$) one. This algebraic structure has been called (Stone) Brouwer–Zadeh (BZ^S) lattice (see [15] for the definition and properties) since the interconnection rule *(in)* $\neg \sim f = \sim\sim f$ between the two negations holds.

On the basis of a BZ structure, making use of the Kleene and Brouwer negation, two other connectives can be introduced which can be considered as Many Valued realizations of modal connectives: the *necessity* and the *possibility* of f which are the crisp sets of the necessity and the possibility domains of f:

$$\Box f := \sim \neg f = \chi_{\Delta_1} \quad \text{and} \quad \Diamond f := \neg \sim f = \chi_{\Delta_p} \tag{2.2}$$

The role of these two unary operators as algebraic realizations of the connectives of necessity and possibility of a S5 modal–like system based on a Kleene lattice (instead of on a Boolean one) is described in [4, 9]. Moreover, these modal operators satisfies the further distrubitivity properties $\nu(f) \vee \nu(g) = \nu(f \vee g)$ and $\mu(f) \wedge \mu(g) = \mu(f \wedge g)$ of a $MDS5$ algebra [12].

As final structure about fuzzy sets we mention the one of Heyting–Wajsberg (HW) algebra introduced and discussed in [4, 13]. This HW algebra is a semantical characterization of a logical system equipped with two implication connectives. In the fuzzy set environment these two connectives are the *Łukasiewicz implication* $\rightarrow_L$ and the *Gödel implication* $\rightarrow_G$ defined as

$$(f_1 \rightarrow_L f_2)(x) := \min\{1, 1 - f_1(x) + f_2(x)\}$$

$$(f_1 \rightarrow_G f_2)(x) := \begin{cases} 1 & f_1(x) \leq f_2(x) \\ f_2(x) & \text{otherwise} \end{cases}$$

On the basis of the system $\langle \mathcal{F}(X), \rightarrow_L, \rightarrow_G, \mathbf{0} \rangle$ all the previously discussed connectives of fuzzy set theory can be derived since it is easy to verify that

$\neg f = f \to_L \mathbf{0}$ and $\sim f := f \to_G \mathbf{0}$, i.e., the Kleene and the Brouwer negations are obtained by the Łukasiewicz and the Gödel implication connectives in the usual way. Moreover, the meet and the join are obtained by the Łukasiewicz implication as $f \wedge g = \neg((\neg f \to_L \neg g) \to_L \neg g)$ and $f \vee g = (f \to_L g) \to_L g$. In this way, the BZ structure of fuzzy sets can be induced as a peculiar substructure of the corresponding HW algebra. We remark that the subsystem $\langle \mathcal{F}(X), \to_L, \mathbf{0}\rangle$ satisfies the axioms of Wajsberg algebra (and hence also of the equivalent Chang's MV algebra), an algebraic model of the Łukasiewicz many valued logic. On the other hand, the substructure $\langle \mathcal{F}(X), \wedge, \vee, \to_G, \neg, \mathbf{0}\rangle$ satisfies the axioms of a symmetric Heyting algebra (and also of Gödel algebra [23]) formalized in [26] as an algebraic model of intuitionistic logic.

3 The Logic–Algebraic Approach to Intuitionistic Fuzzy Sets

Let us consider the same set of objects X as universe of the fuzzy set system $\mathcal{F}(X)$ discussed in the previous section 2. An *intuitionistic fuzzy set (IFS)* on X is any pair of fuzzy sets $A = \langle f_A, g_A\rangle \in \mathcal{F}(X) \times \mathcal{F}(X)$ under the condition $f_A + g_A \leq \mathbf{1}$, or more formally such that $f_A \leq \neg g_A$ or equivalently $g_A \leq \neg f_A$, which defines a binary relation of *orthogonality* $f_A \perp g_A$ on $\mathcal{F}(X)$ according to [14]. For this reason we prefer to speak of pair of fuzzy sets in an orthogonality relation, or *orthopair of fuzzy sets* (OFS).

Let $\mathcal{IF}(X)$ be the collection of IFS on the universe X. Then we can introduce on it a distributive lattice structure $\langle \mathcal{IF}(X), \cap, \cup, (\mathbf{0}, \mathbf{1}), (\mathbf{1}, \mathbf{0})\rangle$ where the meet and join are defined as

$$\langle f_A, g_A\rangle \cap \langle f_B, g_B\rangle = \langle f_A \wedge f_B, g_A \vee g_B\rangle \tag{3.1a}$$

$$\langle f_A, g_A\rangle \cup \langle f_B, g_B\rangle = \langle f_A \vee f_B, g_A \wedge g_B\rangle \tag{3.1b}$$

and whose induced partial order relation is: $\langle f_A, g_A\rangle \subseteq \langle f_B, g_B\rangle$ iff $\forall x \in X$, $f_A(x) \leq f_B(x)$ and $g_B(x) \leq g_A(x)$, with respect to which $\mathbb{O} := (\mathbf{0}, \mathbf{1})$ and $\mathbb{I} := (\mathbf{1}, \mathbf{0})$ are the least and the greatest element, respectively.

Regarding the negation and implication connectives we have the following situation. The unary operation defined for any arbitrary IFS $\langle f_A, g_A\rangle$ by $-\langle f_A, g_A\rangle = \langle g_A, f_A\rangle$ is a de Morgan complementation, i.e., properties (ocK1) and (ocK2) are satisfied by any pair of IFSs. The Kleene condition (ocK3) is not valid and so $\mathcal{IF}(X)$ with the operation $-$ are examples of de Morgan algebras which are not Kleene. Furthermore, in [18] it has been proved that any de Morgan negation on IFS, whatever be its concrete definition, *cannot* satisfy the Kleene condition (ocK3). As a consequence, no Łukasiewicz implication, i.e., an implication $\Rightarrow_L$ which satisfies the Wajsberg axioms, can be defined on IFS since the induced standard negation $\langle f_A, g_A\rangle \Rightarrow_L \langle \mathbf{0}, \mathbf{1}\rangle$ should be necessarily of Kleene type.

On the contrary, an intuitionistic implication can be introduced on $\mathcal{IF}(X)$. This operation was firstly proposed in [17, p. 64] relatively to the unit interval

$[0,1]$ and then extended to $\mathcal{F}(X)$ in [7]. Let $A = \langle f_A, g_A\rangle$ and $B = \langle f_B, g_B\rangle$ be two IFSs, then for x ranging on X let us define:

$$\big(\langle f_A, g_A\rangle \Rightarrow \langle f_B, g_B\rangle\big)(x) := \begin{cases} (1,0) & \text{if } f_A(x) \le f_B(x) \\ & \text{and } g_A(x) \ge g_B(x) \\ (1-g_B(x), g_B(x)) & \text{if } f_A(x) \le f_B(x) \\ & \text{and } g_A(x) < g_B(x) \\ (f_B(x), 0) & \text{if } f_A(x) > f_B(x) \\ & \text{and } g_A(x) \ge g_B(x) \\ (f_B(x), g_B(x)) & \text{if } f_A(x) > f_B(x) \\ & \text{and } g_A(x) < g_B(x) \end{cases} \tag{3.2}$$

The Brouwer negation induced by the implication connective $\Rightarrow$ in the usual manner $\sim \langle f_A, g_A\rangle = \langle f_A, g_A\rangle \Rightarrow \langle \mathbf{0}, \mathbf{1}\rangle$ is pointwise defined (and compare for an analogy with (2.1)) for all $x \in X$ as:

$$\sim \langle f_A, g_A\rangle (x) = \begin{cases} (1,0) & \text{if} \quad g_A(x) = 1 \\ (0,1) & \text{if} \quad g_A(x) \ne 1 \end{cases} \tag{3.3}$$

We underline that the Brouwer negation, besides (ocB1)–(ocB3), satisfies also the Stone (S) condition (or equivalently (ocB4)). Thus, the structure $\langle \mathcal{IF}(X), \cap, \cup, \Rightarrow_G, \langle \mathbf{0}, \mathbf{1}\rangle\rangle$ is a Gödel algebra. Further, if we add the de Morgan negation on IFS $-$, we obtain the structure $\langle \mathcal{IF}(X), \cap, \cup, \Rightarrow_G, -, \langle \mathbf{0}, \mathbf{1}\rangle\rangle$ which is a symmetric Heyting algebra.
By the combination of these two non standard negations we obtain, similarly to the equations (2.2) of the fuzzy sets case, the modal operators of necessity $\Box A := \sim -A$ and possibility $\Diamond A := - \sim A$ pointwise defined as:

$$\Box \langle f_A, g_A\rangle (x) = \begin{cases} (1,0) & \text{if} \quad f_A(x) = 1 \\ (0,1) & \text{if} \quad f_A(x) \ne 1 \end{cases} \tag{3.4a}$$

$$\Diamond \langle f_A, g_A\rangle (x) = \begin{cases} (0,1) & \text{if} \quad f_A(x) = 1 \\ (1,0) & \text{if} \quad f_A(x) \ne 1 \end{cases} \tag{3.4b}$$

which also in the case of IFS satisfy a S5 modal-like behavior but on a De Morgan lattice (differently from the pure fuzzy case in which the basic structure is of Kleene lattice). Finally, another possibility to define a negation is to consider the unary operation introduced in [15] as

$$\approx \langle f_A, g_A\rangle = \langle g_A, \neg g_A\rangle \tag{3.5}$$

which is a weaker form of Brouwer negation; indeed, it satisfies only properties (ocB1) and (ocB2) but not the non contradiction law (ocB3). However, if we use this negation and the de Morgan one to define the modal operators by the usual compositions, one obtains the following necessity and possibility on IFS introduced in [1]: $\blacksquare\langle f_A, g_A\rangle = \langle f_A, \neg f_A\rangle$ and $\blacklozenge\langle f_A, g_A\rangle = \langle \neg g_A, g_A\rangle$ and also they show a S5 modal behavior on a de Morgan lattice. The peculiarity of these

modal operators with respect to the above equations (3.4) is that in general they are not $\{(0,1),(1,0)\}$ valued.

4 The Logic–Algebraic Approach to ICS

In this subsection, we investigate the particular subclass $\mathcal{IC}(X) \subseteq \mathcal{IF}(X)$ of all ortho–pairs of characteristic functions $\langle \chi_{A_1}, \chi_{A_0} \rangle$ of subsets A_1, A_0 of X with $\chi_{A_1} + \chi_{A_0} \leq \mathbf{1}$, that is $\mathcal{IC}(X)$ is the collection of all orthogonal crisp sets (ICS) on the universe X. Trivially $\chi_{A_1} \perp \chi_{A_0}$ iff $A_1 \cap A_0 = \emptyset$ that is, we can identify the ICS $\langle \chi_{A_1}, \chi_{A_0} \rangle$ with the pair $\langle A_1, A_0 \rangle$ of disjoint subsets of X. The subset A_1 (resp., A_0) is the *certainty* (resp., *impossibility*) *domain* of the involved ICS $\langle A_1, A_0 \rangle$. This terminology is strictly linked to the fact that an interesting subcase of ICS is represented by rough sets [28], where given a set A, A_1 is the lower approximation of A and A_0 the exterior region (i.e., whose complement is the upper approximation of A as its *possibility domain* $A_p = A_0^c$).

The lattice operations induced by (3.1) in this particular case assume the form $\langle A_1, A_0 \rangle \cap \langle B_1, B_0 \rangle = \langle A_1 \cap B_1, A_0 \cup B_0 \rangle$ and $\langle A_1, A_0 \rangle \cup \langle B_1, B_0 \rangle = \langle A_1 \cup B_1, A_0 \cap B_0 \rangle$. The complementation induced by the IFS de Morgan one $-\langle A_1, A_0 \rangle = \langle A_0, A_1 \rangle$ is now a Kleene complementation, that is in this particular subcase also property (ocK3) is satisfied. Moreover, it is possible to define a Łukasiewicz implication as $\langle A_1, A_0 \rangle \Rightarrow_L \langle B_1, B_0 \rangle = \langle (A_1^c \cap B_0^c) \cup A_0 \cup B_1, A_1 \cap B_0 \rangle$ whose induced negation is just the above introduced Kleene negation $\langle A_1, A_0 \rangle \Rightarrow_L \langle \mathbf{0}, \mathbf{1} \rangle = -\langle A_1, A_0 \rangle$.

The restriction to the present case of the intuitionistic implication (3.2) produces the ICS implication:

$$\langle A_1, A_0 \rangle \Rightarrow_G \langle B_1, B_0 \rangle = \langle (A_1^c \cap B_0^c) \cup A_0 \cup B_1,\ A_0^c \cap B_0 \rangle \tag{4.1}$$

We remark that these implication connectives coincide with the implications introduced in the context of rough set theory by Pagliani [27, 9]. Moreover, it is possible to show that the structure $\langle \mathcal{IC}(X), \Rightarrow_L, \Rightarrow_G, (\emptyset, X) \rangle$ is a HW algebra. The complementation $\sim$ induced by (3.3) (or equivalently obtained from (4.1)) in the present case assumes the form $\sim \langle A_1, A_0 \rangle = \langle A_0, (A_0)^c \rangle$, which is also equal to the rescrition to $\mathcal{IC}(X)$ of equation (3.5). The modal operators induced by the two negations are the necessity $\Box(A_1, A_0) = (A_1, A_1^c)$ and possibility $\Diamond(A_1, A_0) = (A_0^c, A_0)$. The structure $\langle \mathcal{IC}(X), \cap, \cup, -, \Diamond, \Box, (X, \emptyset), (\emptyset, X) \rangle$ is a DD_5 algebra according to [12]. Finally, any ICS satisfies the *three–value condition* (3) $a \vee \sim a = a \vee -a$, property which allows one to characterize ICS as different from FS and IFS, which, in general, do not satisfy it.

Remark 4.1. The concept of ICS has been considered also in [15] and later on in [16]. It has also been studied in an equivalent form in [21], [25] and [30]. Indeed, in these last papers pairs of ordinary subsets of the universe X of the kind $\langle A_1, A_p \rangle$, under the condition $A_1 \subseteq A_p$ are considered. The mapping $\langle A_1, A_0 \rangle \to \langle A_1, (A_0)^c \rangle$ institute a one-to-one and onto correspondence which allows one to identify the two approaches.

Remark 4.2. ICS are equivalent to shadowed sets, a different approach to vagueness proposed by Pedrycz [29]. A *shadowed set* on a universe X is any mapping $s : X \to \{0, (0,1), 1\}$. Once introduced for any ICS $\langle A_1, A_0\rangle$ its *uncertainty domain* $A_u = X \setminus (A_1 \cup A_0)$, then the mapping $\langle A_1, A_0\rangle \to \chi_{A_1} + (0,1) \cdot \chi_{A_u}$ is a one-to-one and onto correspondence which allows one to identify ICS and shadowed sets. For a more detailed investigation about the theoretical approach to shadowed sets see [8, 11].

5 Conclusions

We gave an algebraic framework to FS, IFS and ICS investigating in particular, their behavior with respect to Łukasiewicz and intuitionistic logic. We can conclude that there is at least a property which enable to differentiate these three way to address vagueness from the algebraic standpoint. In particular, when considering the two ortho–complementations $\sim$ and $\neg$, ICS satisfy the Kleene (K), Stone (S) and three–value condition (3), whereas FS satisfy (K) and (S) but not (3), and finally IFS only (S). We note that this is not the only one algebraic approach to IFS (or equivalent structures, see for instance [20]) and a comparison will be presented in a forthcoming work.

References

[1] Atanassov, K.T.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
[2] Atanassov, K.T.: Answer to D. Dubois, S. Gottwald, P. Hajek, J. Kacprzyk and H. Prade's paper terminological difficulties in fuzzy set theory - the case of Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 156, 496–499 (2005)
[3] Atanassov, K.T., Stoeva, S.: Intuitionistic fuzzy sets. In: Polish Symp. on Interval & Fuzzy Mathematics (Poznan), pp. 23–26 (August 1983)
[4] Cattaneo, G., Ciucci, D.: Heyting Wajsberg algebras as an abstract environment linking fuzzy and rough sets. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS, vol. 2475, pp. 77–84. Springer, Heidelberg (2002)
[5] Cattaneo, G., Ciucci, D.: An algebraic approach to shadowed sets. Electronic Notes in Theoretical Computer Science 82(4), 64–75 (2003); Proceedings of International Workshop on Rough Sets in Knowledge Discovery and Soft Computing, April 12–13, 2003, Warsaw
[6] Cattaneo, G., Ciucci, D.: Generalized negations and intuitionistic fuzzy sets. A criticism to a widely used terminology. In: Proceedings of International Conference in Fuzzy Logic and Technology (EUSFLAT 2003) (Zittau), University of Applied Sciences of Zittau–Goerlitz, pp. 147–152 (2003)
[7] Cattaneo, G., Ciucci, D.: Intuitionistic fuzzy sets or orthopair fuzzy sets? In: Proceedings of International Conference in Fuzzy Logic and Technology (EUSFLAT 2003) (Zittau), University of Applied Sciences of Zittau–Goerlitz, pp. 153–158 (2003)
[8] Cattaneo, G., Ciucci, D.: Shadowed sets and related algebraic structures. Fundamenta Informaticae 55, 255–284 (2003)
[9] Cattaneo, G., Ciucci, D.: Algebraic structures for rough sets. In: Peters, J.F., Skowron, A., Dubois, D., Grzymała-Busse, J.W., Inuiguchi, M., Polkowski, L. (eds.) Transactions on Rough Sets II. LNCS, vol. 3135, pp. 218–264. Springer, Heidelberg (2004)

[10] Cattaneo, G., Ciucci, D.: Basic intuitionistic principles in fuzzy set theories and its extensions (a terminological debate on Atanassov IFS). Fuzzy sets and Systems 157, 3198–3219 (2006)
[11] Cattaneo, G., Ciucci, D.: Theoretical aspects of shadowed sets. In: Pedrycz, W., Skowron, A., Kreinovich, V. (eds.) Handbook of Granular Computing, pp. 603–627. John Wiley & Sons, Chichester (2008)
[12] Cattaneo, G., Ciucci, D., Dubois, D.: Algebraic models of deviant modal logics based on Kleene algebras (preprint) (2008)
[13] Cattaneo, G., Ciucci, D., Giuntini, R., Konig, M.: Algebraic structures related to many valued logical systems. part I: Heyting Wajsberg algebras. Fundamenta Informaticae 63(4), 331–355 (2004)
[14] Cattaneo, G., Marino, G.: Non-usual orthocomplementations on partially ordered sets and fuzziness. Fuzzy Sets and Systems 25, 107–123 (1988)
[15] Cattaneo, G., Nisticò, G.: Brouwer-Zadeh posets and three valued Łukasiewicz posets. Fuzzy Sets and Systems 33, 165–190 (1989)
[16] Coker, D.: A note on intuitionistic sets and intuitionistic points. Turkish Journal of Mathematics 20, 343–351 (1996)
[17] Cornelis, C., Deschrijver, G., Kerre, E.: Implication in intuitionistic fuzzy and interval–valued fuzzy set theory: construction, classification, application. International Journal of Approximate Reasoning 35, 55–95 (2004)
[18] Deschrijver, G., Cornelis, C., Kerre, E.: Triangle and square: a comparison. In: IPMU 2004, pp. 1389–1395 (2004)
[19] Dubois, D., Gottwald, S., Hajek, P., Kacprzyk, J., Prade, H.: Terminological difficulties in fuzzy set theory - the case of Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 156, 485–491 (2005)
[20] Van Gasse, B., Cornelis, C., Deschrijver, G., Kerre, E.E.: Triangle algebras: A formal logic approach to interval–valued residuated lattices. Fuzzy Sets and Systems 159, 1042–1060 (2008)
[21] Gentilhomme, M.Y.: Les ensembles flous en linguistique. Cahiers de linguistique theoretique et applique, Bucarest 47, 47–65 (1968)
[22] Grzegorzewski, P., Mrowka, E.: Some notes on (Atanassov) Intuitionistic fuzzy sets. Fuzzy Sets and Systems 156, 492–495 (2005)
[23] Hájek, P.: Metamathematics of fuzzy logic. Kluwer, Dordrecht (1998)
[24] Heyting, A.: Intuitionism: an introduction, 2nd edn. North Holland, Amsterdam (1966) (first edn. 1956)
[25] Iwinski, T.: Algebras for rough sets. Bulletin of the Polish Academy of Sciences, series: Mathematics 35, 673–683 (1987)
[26] Monteiro, A.: Sur les algèbres de Heyting symétriques. Portugaliae Mathematica 39, 1–237 (1980)
[27] Pagliani, P.: Rough Sets and Nelson Algebras. Fundamenta Informaticae 27(2,3), 205–219 (1996)
[28] Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Sciences 177, 3–27 (2007)
[29] Pedrycz, W.: Shadowed sets: Representing and processing fuzzy sets. IEEE Transaction on Systems, Man and Cybernetics - PART B: Cybernetics 28(1), 103–109 (1998)
[30] Yao, Y., Li, X.: Comparison of rough-set and interval-set models for uncertain reasoning. Fundamenta Informaticae 27, 289–298 (1997)

General Form of Probabilities on IF-Sets

Lavinia Ciungu[1,2] and Beloslav Riečan[3,4]

[1] Polytechnical University of Bucharest, Splaiul Independentei 113, Bucharest, Romania
[2] State University of New York - Buffalo, 244 Mathematics Building, Buffalo NY, 14260-2900, USA
[3] Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovakia
[4] Matematický Ústav SAV, Štefánikova 49, 814 73 Bratislava, Slovakia
lavinia_ciungu@math.pub.ro, riecan@fpv.umb.sk

Abstract. The paper has two aims. First, a review of various definitions of probabilities on Atanassov IF-sets, and corresponding representation theorems. Secondly, a new representation theorem is proved for so-called φ-probabilities including a large variety of special cases.

Keywords: IF-events, probability.

1 IF-Events

According to Atanassov ([1]), an IF-set is a mapping

$$A = (\mu_A, \nu_A),$$

defined on an non-empty set Ω to $[0,1]^2$ (i.e. $\mu_A : \Omega \to [0,1], \nu_A : \Omega \to [0,1]$), such that

$$\mu_A(x) + \nu_A(x) \leq 1$$

for any $x \in \Omega$. An IF-set $A = (\mu_A, \nu_A)$ is called an IF-event, if $\mu_A, \nu_A : \Omega \to [0,1]$ are measurable mappings with respect to the given σ-algebra $\mathcal{S}$ of subsets of Ω. Recall that μ_A is called a membership function, ν_A the non-membership function. Therefore it is natural to define

$$A = (\mu_A, \nu_A) \leq B = (\mu_B, \nu_B)$$

if and only if

$$\mu_A \leq \mu_A, \nu_A \geq \nu_B.$$

Denote by $\mathcal{F}$ the family of all IF-events. With respect to the preceding definition it is easy to see that $(0_\Omega, 1_\Omega)$ is the smallest element of $\mathcal{F}$, $(1_\Omega, 0_\Omega)$ is the greatest element of $\mathcal{F}$.

Probability of an IF-event A was defined first constructively (see [7,6]) as a compact interval, lately axiomatically (see [11,12,13]) as a mapping

$$\mathcal{P} : \mathcal{F} \to \mathcal{J}$$

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 101–107, 2009.

where $\mathcal{J}$ is the family of all compact subintervals $[c,d]$ of [0,1] in the real line. Since

$$\mathcal{P}(A) = [\mathcal{P}^\flat(A), \mathcal{P}^\sharp(A)],$$

we obtain two real functions

$$\mathcal{P}^\flat, \mathcal{P}^\sharp : \mathcal{F} \to [0,1].$$

These functions will be called states (analogously as in quantum structures [5]). Evidently any solution of the problem of IF - states on $\mathcal{F}$ leads naturally to a solution of the problem of IF - probabilities.

Similarly as in the classical or quantum case resp., IF - state $m : \mathcal{F} \to [0,1]$ can be defined as a normalized, additive and continuous mapping. The first condition is clear:

$$(1) \quad m((0_\Omega, 1_\Omega)) = 0, m((1_\Omega, 0_\Omega)) = 1.$$

Similarly continuity is determine uniquely:

$$(2) \quad A_n \nearrow A \Longrightarrow m(A_n) \nearrow m(A).$$

Here $A_n \nearrow A$ is equivalent with two convergences of real functions:

$$\mu_{A_n} \nearrow \mu_A, \nu_{A_n} \searrow \nu_A.$$

Of course, the additivity cannot be determined uniquely. First it was defined by the help of Lukasiewicz connectives.

2 Lukasiewicz States

If $f, g : \Omega \to [0,1]$ are two fuzzy sets. then the Lukasiewicz connectives are

$$f \oplus_L g - \min(f+g, 1), f \odot_L g - \max(f+g=1, 0).$$

Therefore for $A, B \in \mathcal{F}$ we define

$$A \oplus_L B = (\mu_A \oplus_L \mu_B, \nu_A \odot_L \nu_B),$$

$$A \odot_L B = (\mu_A \odot_L \mu_B, \nu_A \oplus_L \nu_B).$$

Hence L-additivity means the implication

$$(3) \quad A \odot_L B = (0_\Omega, 1_\Omega) \Longrightarrow m(A \oplus_L B) = m(A) + m(B).$$

A mapping $m : \mathcal{F} \to [0,1]$ is called L-state if the conditions (1), (2) and (3) hold. The first important result was the representation theorem ([13]):

Theorem 1. *Let $m : \mathcal{F} \to [0,1]$ be an L-state and there exists a probability measure $P : \mathcal{S} \to [0,1]$ and a function $f : [0,1]^2 \to [[0,1]$ such that*

$$(*) \qquad m(A) = f(\int_\Omega \mu_A dP, \int_\Omega \nu_A dP).$$

Then f is linear, hence there exists $\alpha \in [0,1]$ such that

$$m(A) = (1-\alpha)\int_\Omega \mu_A dP + \alpha(1 - \int_\Omega \nu_A dP)$$

for any $A \in \mathcal{F}$.

Recently L. Ciungu proved that the assumption (*) can be omitted (see [3]).

Theorem 2. *Let $m : \mathcal{F} \to [0,1]$ be an L-state. Then there are probability measures $P, Q : \mathcal{S} \to [0,1]$ and $\alpha \in [0,1]$ such that*

$$m(A) = (1-\alpha)\int_\Omega \mu_A dP + \alpha(1 - \int_\Omega \nu_A dQ)$$

for any $A \in \mathcal{F}$.

The second important result is in the following theorem ([14,15,4]).

Theorem 3. *To any $\mathcal{F}$ with an L-state $m : \mathcal{F} \to [0,1]$ there exists an MV algebra M with a state $\mu : M \to [0,1]$ such that $(\mathcal{F}, m)$ and (M, μ) are isomorphic.*

Theorem 3 gives possibility to use well developed probability theory on MV algebras (see [16]). Of course, later there appeared alternative definitions using alternative connectives and corresponding states.

1. M-probability with

$$A \oplus_M B = (\max(\mu_A, \mu_B), \min(nu_A, \nu_B))$$
$$A \odot_M B = (\min(\mu_A, \mu_B), \max(nu_A, \nu_B)).$$

(see [8,15,4]).

2. Q-probability with

$$A \oplus_Q B = (\sqrt{\mu_A^2 + \mu_B^2}, 1 - \sqrt{(1-\nu_A)^2 + (1-\nu_B)^2}),$$
$$A \odot_Q B = (\min(\mu_A + \mu_B, 1), \max(\nu_A + \nu_B - 1, 0)).$$

(see [2]).

3. P-probability with

$$A \oplus_P B = (\mu_A + \mu_B - \mu_A\mu_B, \nu_A\nu_B),$$
$$A \odot_P B = (\mu_A\mu_B, \nu_A + \nu_B - \nu_A\nu_B).$$

(see [2]).

The aim of this paper is to prove a representation theorem for so-called φ-states. Here $\varphi : [0,1] \to [0,1]$ is an increasing bijection such that $\varphi(u) \leq u (u \in [0,1])$. Following [9] we shall consider the following pair of connectives

$$A \oplus_\varphi B = (\varphi^{-1}(\min(\varphi(\mu_A) + \varphi(\mu_B), 1)), 1 - \varphi^{-1}(\min(\varphi(1-\nu_A) + \varphi(1-\nu_B), 1))),$$
$$A \odot B = (\max(\mu_A + \mu_B - 1, 0), \min(\nu_A + \nu_B, 1)).$$

3 φ -States

φ-state is a mapping $m : \mathcal{F} \to [0,1]$ such that

$$(i) \quad m((0_\Omega, 1_\Omega)) = 0, m((1_\Omega, 0_\Omega)) = 1;$$

$$(ii) \quad A \odot B = (0_\Omega, 1_\Omega) \Longrightarrow m(A \oplus_\varphi B) = m(A) + m(B);$$

$$(iii) \quad A_n \nearrow A \Longrightarrow m(A_n) \nearrow m(A).$$

Evidently any L-state is a φ-state, where $\varphi(u) = u, u \in [0,1]$. Similarly any Q-state is a φ-state, where $\varphi(u) = u^2$. Moreover, in [9] the Yager state was considered where $\varphi(u) = u^n (n \in N$;for $n = 1$ we obtain the Lukasiewicz state, for $n = 2$ the Yager state). In [9] also the cases $\varphi(u) = 2^u$ and $\varphi(u) = logu$ were considered.

In [10] Renčová introduced the notion of the strong additivity using the operation

$$A \odot_\varphi B = ((\varphi(\mu_A) + \varphi(\mu_B) - 1) \vee 0, (\varphi(1 - \nu_A) + \varphi(1 - \nu_B)) \wedge 1).$$

The state m is strongly additive, if

$$A \odot_\varphi B = (0_\Omega, 1_\Omega) \Longrightarrow m(A \oplus_\varphi B) = m(A) + m(B).$$

Since $\varphi(u) \leq u$ for any $u \in [0,1]$, it is not difficult to show that

$$A \odot B = (0_\Omega, 1_\Omega) \Longrightarrow A \odot_\varphi B = (0_\Omega, 1_\Omega).$$

Therefore any strongly φ-additive state is φ-additive. Renčová using Theorem 1 proved the representation theorem for strongly φ-additive states:

$$m(A) = (1 - \alpha) \int_\Omega \varphi(\mu_A) dP + \alpha \int_\Omega \varphi(1 - \nu_A) dP.$$

In the following theorem the same result will be proved for arbitrary φ-state, of course again with the following additional assumption:

$$(**) \quad m(A) = f(\int_\Omega \varphi(\mu_A) dP, \int_\Omega \varphi(1 - \nu_A) dQ),$$

where $P, Q : \mathcal{S} \to [0,1]$ are some probability measures and $f : [0,1]^2 \to [0,1]$ is a continuous function.

Theorem 4. *To any φ-state m satisfying the condition (**) there exists $\alpha \in [0,1]$ such that*

$$m(A) = (1 - \alpha) \int_\Omega \varphi(\mu_A) dP + \alpha \int_\Omega \varphi(1 - \nu_A) dQ$$

for any $A \in \mathcal{F}$.

Proof. First we see that

$$0 = m((0_\Omega, 1_\Omega)) = f(\int_\Omega \varphi(0_\Omega)dP, \int_\Omega \varphi(1 - 1_\Omega)dQ) = f(0,0)$$

hence $f(0,0) = 0$. Similarly

$$1 = m((1_\Omega, 0_\Omega)) = f(\int_\Omega \varphi(1_\Omega)dP, \int_\Omega \varphi(1 - 0_\Omega)dQ) = f(1,1),$$

hence $f(1,1) = 1$. Let $A \odot B = (0_\Omega, 1_\Omega)$. It means $\mu_A + \mu_B \leq 1, \nu_A + \nu_B \geq 1$, hence $(1 - \nu_A) + (1 - \nu_B) \leq 1$. Therefore

$$\varphi(\mu_A) + \varphi(\mu_B) \leq 1, \varphi(1 - \nu_A) + \varphi(1 - \nu_B) \leq 1,$$

$$A \oplus_\varphi B = (\varphi^{-1}(\varphi(\mu_A) + \varphi(\mu_B)), 1 - \varphi^{-1}(\varphi(1 - \nu_A) + \varphi(1 - \nu_B))).$$

Put

$$\int_\Omega \varphi(\mu_A)dP = u_1, \int_\Omega \varphi(1 - \nu_A)dQ = u_2,$$

$$\int_\Omega \varphi(\mu_B)dP = v_1, \int_\Omega \varphi(1 - \nu_B)dQ = u_2.$$

Then

$$m(A) = f(u_1, u_2), m(B) = f(v_1, v_2),$$

$$m(A \oplus_\varphi B) = f(\int_\Omega \varphi(\varphi^{-1}(\varphi(\mu_A) + \varphi(\mu_B)))dP, \int_\Omega \varphi(1 - 1 + \varphi^{-1}(\varphi(1 - \nu_A) + \varphi(1 - \nu_B)))dQ) =$$

$$= f(\int_\Omega \varphi(\mu_A)dP + \int_\Omega \varphi(\mu_B)dP, \int_\Omega \varphi(1 - \nu_A)dQ + \int_\Omega \varphi(1 - \nu_B)dQ) =$$

$$= f(u_1 + v_1, u_2 + v_2),$$

hence we have obtained the identity

$$(1) \quad f(u_1 + v_1, u_2 + v_2) = f(u_1, u_2) + f(v_1 + v_2).$$

Putting $A = B$ and using induction we obtain

$$(2) \quad f(kx) = kf(x)$$

for any $k \in N$ such that $kx \in [0,1]^2$. Let $\frac{p}{q} \in Q$ with $\frac{p}{q}x \in [0,1]^2$. Then

$$f(x) = f(\frac{1}{q}x) + ...f(\frac{1}{q}x) = qf(\frac{1}{q}x),$$

hence

$$f(\frac{1}{q}x) = \frac{1}{q}f(x),$$

and

$$(3) \quad \frac{p}{q}f(x) = f(\frac{p}{q}x).$$

Since f is continuous, we obtain the identity

$$(4) \quad f(ax) = af(x), a \in [0,1], x \in [0,1]^2, ax \in [0,1]^2.$$

We have seen that

$$f(0,0) = 0, f(1,1) = 1.$$

Put $f(0,1) = \alpha$. By (1) we obtain

$$1 = f(1,1) = f(0,1) + f(1,0) = \alpha + f(1,0),$$

hence

$$f(1,0) = 1 - \alpha.$$

Finally

$$m(A) = f(u_1, u_2) = f(u_1, 0) + f(0, u_2) =$$
$$= u_1 f(1,0) + u_2 f(0,1) = (1-\alpha)u_1 + \alpha u_2 =$$
$$= (1-\alpha) \int_\Omega \varphi(\mu_A) dP + \alpha \int_\Omega \varphi(1 - \nu_A) dQ.$$

□

Acknowledgements

The paper was supported by Grant VEGA 1/0539/08, and Grant APVV LPP-0046-06.

References

1. Atanassov, K.: Intuitionistic Fuzzy Sets: Theory and Applications. Physica Verlag, New York (1999)
2. Atanassov, K.: Riečan, B: On two new types of probability theory on IF-events. Notes on IFS (2007)
3. Ciungu, L.: Riečan, B.: General for of probabilities on IF-sets. Information Science (submitted)
4. Čunderlíková, K., Riečan, B.: Intuitionistic fuzzy probability theory. Edited volume on Intuitionistic Fuzzy Sets: recent Advances of the series Studies on Fuzziness and Soft Computing. Springer, Heidelberg (2008)
5. Dvurečenskij, A., Pulmannová, S.: New Trends in Quantum Structures. Kluwer, Dordrecht (2000)
6. Gerstenkorn, T., Manko, J.: Probabilities of intuitionistic fuzzy events. In: Hryniewicz, O., et al. (eds.) Issues in Intelligent Systems: Paradigms, pp. 63–68. EXIT, Warszawa (2005)
7. Grzegorzewski, P., Mrowka, E.: Probability of intuitionistic fuzzy events. In: Grzegorzewski, P., et al. (eds.) Soft Methods in Probability, Statistics and data Analysis, pp. 105–115. Physica Verlag, New York (2002)
8. Krachounov, M.: Intuitionistic probability and intuitionistic fuzzy sets. In: El-Darzi, et al. (eds.) First Intern. Workshop on IFS, pp. 714–717. Univ. Of Westminster, London (2006)

9. Renčová, M.: On the φ-probability and φ-observables. Fuzzy Sets and Systems (submitted)
10. Renčová, M.: General form of strongly additive φ-probability. In: Magdalenca, L., et al. (eds.) Proc. IPMU 2008, pp. 1671–1674 (2008)
11. Riečan, B.: A descriptive definition of the probability on intuitionistic fuzzy sets. In: Wagenecht, M., Hampet, R. (eds.) Proc. EUSFLAT 2003, Zittau-Goerlitz Univ. Appl. Sci, Dordrecht, pp. 263–266 (2000)
12. Riečan, B.: Representation of probabilities on IFS events. In: López-Diaz, M., et al. (eds.) Advances in Soft Computing, Soft Methodology and Random Information Systems, pp. 243–246. Springer, Berlin (2004)
13. Riečan, B.: On a problem of Radko Mesiar: general form of IF - probabilities. Fuzzy sets and Systems 157, 1485–1490 (2006)
14. Riečan, B.: On the entropy of IF dynamical systems. In: Proc. IWIFSGN 2005, Warszawa, September 16, 2005, pp. 328–336 (2005)
15. Riečan, B.: Probability theory on IF events. In: Aguzzoli, S., Ciabattoni, A., Gerla, B., Manara, C., Marra, V. (eds.) ManyVal 2006. LNCS, vol. 4460, pp. 290–308. Springer, Heidelberg (2007)
16. Riečan, B.: - Mundici, D.: Probability on MV-algebras. In: Pap, E. (ed.) Hanbook on Measure Theory, pp. 869–909. Elsevier, Amsterdam (2002)
17. Riečan, B., Neubrunn, T.: Integral, Measure, and Ordering. Kluwer, Dordrecht (1997)

On the E-Probability on IF-Events

Magdaléna Renčová

Matej Bel University, Tajovského 40, SK-974 01 Banská Bystrica, Slovakia
rencova@fpv.umb.sk

Abstract. Following [2] some properties of E-probability and E-states are studied. The existence of the joint observable and the central limit theorem are proved.

Keywords: probability, state, representation theorem, IF-events.

1 Introduction

Although there are different opinions about IF-events, the following definitions are accepted generally ([1], [5]). Let $(\Omega, \mathcal{S})$ be a measurable space. By an IF-event ([5]) we mean any pair

$$\mathbf{A} = (\mu_A, \nu_A)$$

of $\mathcal{S}-$measurable functions, such that $\mu_A, \nu_A : \Omega \to [0,1]$ and $\mu_A + \nu_A \leq 1$.

The function μ_A is the membership function and the function ν_A is the non-membership function. The family $\mathcal{F}$ of all IF-events is ordered in the following way:

$$\mathbf{A} \leq \mathbf{B} \Leftrightarrow \mu_A \leq \mu_B, \nu_A \geq \nu_B.$$

Evidently the notion of IF-event is a natural generalization of the notion of a fuzzy event. Hence we want to define probability on IF-events generalizing probability on fuzzy events. And actually, two constructions were proposed independently by Gregorzewski [5] and Gerstenkorn [4], both based on the Łukasiewicz operations. Operations on $[0,1]^2$ (not necessarily Łukasiewicz operations) can be naturally extended to IF-events in the following way

$$\mathbf{A} \oplus \mathbf{B} = (\mu_A \oplus \mu_B, \nu_A \odot \nu_B),$$

$$\mathbf{A} \odot \mathbf{B} = (\mu_A \odot \mu_B, \nu_A \oplus \nu_B),$$

where $\mathbf{A} = (\mu_A, \nu_A)$ and $\mathbf{B} = (\mu_A, \nu_B)$.

If $\mu : \Omega \to [0,1]$ is a fuzzy set, then $(\mu, 1-\mu)$ is an IF set corresponding to this fuzzy set. Similarly as in the classical case, in the fuzzy case and in the quantum case, a probability (or state) has been introduced as a mapping $m : \mathcal{F} \to [0,1]$ being continuous, additive and satisfying some boundary conditions. Here the main difference is the additivity which is now of the following form

$$m(\mathbf{A}) + m(\mathbf{B}) = m(\mathbf{A} \oplus \mathbf{B}) + m(\mathbf{A} \odot \mathbf{B}).$$

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 108–115, 2009.

The obtained IF-probability is very successful (see [11]). First, any IF-probability space can be included into an MV-algebra probability space, hence the well developed MV-algebra probability theory can be applied directly. Second, there exists a general representation theorem for IF-probability. If $(\Omega, \mathcal{S}, P)$ is a probability space, then to any Łukasiewicz state $m : \mathcal{F} \to [0,1]$ there exists $\alpha \in [0,1]$ such that

$$m(\mathbf{A}) = (1-\alpha)\int\limits_{\Omega} \mu_A dP + \alpha(1 - \int\limits_{\Omega} \nu_A dP)$$

for any $\mathbf{A} \in \mathcal{F}$ (see [2]).

Throughout this paper we consider the following operations with IF-events

$$\mathbf{A} \oplus_E \mathbf{B} = (\mu_A, \nu_A) \oplus_E (\mu_B, \nu_B) =$$

$$= ((\log_a(a^{\mu_A} + a^{\mu_B} - 1)) \wedge 1; 1 - (\log_a(a^{1-\nu_A} + a^{1-\nu_B} - 1)) \wedge 1)$$

$$A \odot B = (\mu_A, \nu_A) \odot (\mu_B, \nu_B) = ((\mu_A + \mu_B - 1) \vee 0; (\nu_A + \nu_B) \wedge 1)$$

Remark 1. This is a special case of operations studied in [8], where $\varphi(u) = \frac{a^u - 1}{a-1}$, $a > 1, a \in R$ is fixed for each $u \in [0,1]$.

We are not able to embed the family $\mathcal{F}$ with these operations into an MV-algebra. Of course, we are able to prove probability representation theorems, to construct the joint observable and prove such fundamental theorems as central limit theorem or laws of large numbers.

2 E-Probability and E-Observables

Definition 1. *Let $\mathcal{F}$ be the family of all IF-events, $\mathcal{J}$ be the family of all compact subintervals of the unit interval $[0,1]$. E-probability is any mapping $\mathcal{P} : \mathcal{F} \to \mathcal{J}$ satisfying the following conditions:*

(i) $\mathcal{P}((\mathbf{1}, \mathbf{0})) = [1,1], \mathcal{P}((\mathbf{0}, \mathbf{1})) = [0,0]$;
(ii) $\mathbf{A} \odot \mathbf{B} = (0,1) \Rightarrow \mathcal{P}(\mathbf{A} \oplus_E \mathbf{B}) = \mathcal{P}(\mathbf{A}) + \mathcal{P}(\mathbf{B})$;
(iii) $\mathbf{A_n} \nearrow \mathbf{A} \Rightarrow \mathcal{P}(\mathbf{A_n}) \nearrow \mathcal{P}(\mathbf{A})$. (Here $[a_n, b_n] \nearrow [a,b]$, if $a_n \nearrow a, b_n \nearrow b$.)

Definition 2. *A mapping $m : \mathcal{F} \to [0,1]$ is called an E-state, if the following conditions are satisfied:*

(i) $m((\mathbf{1},\mathbf{0}))=1$, $m((\mathbf{0},\mathbf{1}))=0$;
(ii) $\mathbf{A} \odot \mathbf{B} = (0,1) \Rightarrow m(\mathbf{A} \oplus_E \mathbf{B}) = m(\mathbf{A}) + m(\mathbf{B})$;
(iii) $\mathbf{A_n} \nearrow \mathbf{A} \Rightarrow m(\mathbf{A_n}) \nearrow m(\mathbf{A})$.

Example 1. Let $(\Omega, \mathcal{S}, p)$ be a probability space, then natural example of E-state is a function $m : \mathcal{F} \to [0,1]$ defined for fixed $a \in R, a > 1$ by the following

$$m((\mu_A, \nu_A)) = \int\limits_{\Omega} \frac{a^{\mu_A} - 1}{a-1} dp.$$

Proposition 1. *Let $\alpha \in [0,1]$ be real number, than any mapping $m_\alpha : \mathcal{F} \to [0,1]$ defined by the following*

$$m_\alpha((\mu_A, \nu_A)) = (1-\alpha)\int\limits_\Omega \frac{a^{\mu_A}-1}{a-1}dp + \alpha(\int\limits_\Omega \frac{a^{1-\nu_A}-1}{a-1}dr),$$

where $n \in N$ is fixed and p, r are probability measures, is E-state.

Proof. We will show, that $m_\alpha : \mathcal{F} \to [0,1]$ is an E-state:

(i) $m_\alpha((\mathbf{1},\mathbf{0})) = (1-\alpha)\int\limits_\Omega 1dp + \alpha(\int\limits_\Omega 1dr) = 1-\alpha+\alpha = 1;$
$m_\alpha((\mathbf{0},\mathbf{1})) = (1-\alpha)\int\limits_\Omega 0dp + \alpha(\int\limits_\Omega 0dr) = 0;$

(ii) If $\mathbf{A} \odot \mathbf{B} = (\mathbf{0},\mathbf{1})$, i.e. $\mu_A + \mu_B \leq 1, \nu_A + \nu_B \geq 1$ then

$$\begin{aligned} m_\alpha(\mathbf{A} \oplus_E \mathbf{B}) &= m_\alpha((\log_a(a^{\mu_A}+a^{\mu_B}-1)); 1-(\log_a(a^{1-\nu_A}+a^{1-\nu_B}-1))) = \\ &= (1-\alpha)\int\limits_\Omega \frac{a^{\mu_A}+a^{\mu_B}-2}{a-1}dp + \alpha\int\limits_\Omega \frac{a^{1-\nu_A}+a^{1-\nu_B}-2}{a-1}dr = \\ &= (1-\alpha)\int\limits_\Omega \frac{a^{\mu_A}-1}{a-1}dp + \alpha\int\limits_\Omega \frac{a^{1-\nu_A}-1}{a-1}dr + \\ &+ (1-\alpha)\int\limits_\Omega \frac{a^{\mu_B}-1}{a-1}dp + \alpha\int\limits_\Omega \frac{a^{1-\nu_B}-1}{a-1}dr = \\ &= m_\alpha(\mathbf{A}) + m_\alpha(\mathbf{B}), \end{aligned}$$

(iii)

$$\begin{aligned} m_\alpha(\mathbf{A_n}) &= (1-\alpha)\int\limits_\Omega \frac{a^{\mu_{A_n}}-1}{a-1}dp + \alpha\int\limits_\Omega \frac{a^{1-\nu_{A_n}}-1}{a-1}dr \nearrow \\ \nearrow (1-\alpha)\int\limits_\Omega \frac{a^{\mu_A}-1}{a-1}dp &+ \alpha\int\limits_\Omega \frac{a^{\nu_A}-1}{a-1}dr) = m_\alpha(\mu_A,\nu_A) = m_\alpha(\mathbf{A}). \end{aligned}$$

□

Let us suppose, that $\mathcal{P}$ maps $\mathcal{F}$ to $\mathcal{J}$. We will present this mapping with functions $\mathcal{P}^\flat, \mathcal{P}^\sharp : \mathcal{F} \to [0,1]$ in the following manner $\mathcal{P}(\mathbf{A}) = [\mathcal{P}^\flat(\mathbf{A}), \mathcal{P}^\sharp(\mathbf{A})], \mathbf{A} \in \mathcal{F}$. Shorter notation is used further on is $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp]$.

Theorem 1. *$\mathcal{P} : \mathcal{F} \to \mathcal{J}$, is an E-probability if and only if $\mathcal{P}^\flat, \mathcal{P}^\sharp : \mathcal{F} \to [0,1]$ are E-states.*

Proof. Let us suppose that $\mathcal{P}$ is an E- probability, then since

$$[1,1] = \mathcal{P}((\mathbf{1},\mathbf{0})) = [\mathcal{P}^\flat((\mathbf{1},\mathbf{0})), \mathcal{P}^\sharp((\mathbf{1},\mathbf{0}))],$$

we have $1 = \mathcal{P}^\flat((\mathbf{1},\mathbf{0}))$ and $1 = \mathcal{P}^\sharp((\mathbf{1},\mathbf{0}))$.

Further let $\mathbf{A} \odot \mathbf{B} = (\mathbf{0},\mathbf{1})$. Then

$$[\mathcal{P}^\flat(\mathbf{A})+\mathcal{P}^\flat(\mathbf{B}),\mathcal{P}^\sharp(\mathbf{A})+\mathcal{P}^\sharp(\mathbf{B})]=[\mathcal{P}^\flat(\mathbf{A}),\mathcal{P}^\sharp(\mathbf{A})]+[\mathcal{P}^\flat(\mathbf{B}),\mathcal{P}^\sharp(\mathbf{B})]=$$
$$\mathcal{P}(\mathbf{A})+\mathcal{P}(\mathbf{B})=\mathcal{P}(\mathbf{A}\oplus_E\mathbf{B})=[\mathcal{P}^\flat(\mathbf{A}\oplus_E\mathbf{B}),\mathcal{P}^\sharp(\mathbf{A}\oplus_E\mathbf{B})],$$

hence $\mathcal{P}^\flat(\mathbf{A})+\mathcal{P}^\flat(\mathbf{B})=\mathcal{P}^\flat(\mathbf{A}\oplus_E\mathbf{B})$ and $\mathcal{P}^\sharp(\mathbf{A})+\mathcal{P}^\sharp(\mathbf{B})=\mathcal{P}^\sharp(\mathbf{A}\oplus_E\mathbf{B})$. Finally

$$\mathbf{A_n}\nearrow\mathbf{A}\text{ implies }[\mathcal{P}^\flat(\mathbf{A_n}),\mathcal{P}^\sharp(\mathbf{A_n})]=\mathcal{P}(\mathbf{A_n})\nearrow\mathcal{P}(\mathbf{A}),$$

hence

$$\mathcal{P}^\flat(\mathbf{A_n})\nearrow\mathcal{P}^\flat(\mathbf{A})\text{ and }\mathcal{P}^\sharp(\mathbf{A_n})\nearrow\mathcal{P}^\sharp(\mathbf{A}).$$

The opposite implication can be proved similarly. □

3 E-Observables and P-Joint E-Observables

Definition 3. *A mapping $x:\mathcal{B}(R)\to\mathcal{F}$ is called an E-observable, if the following conditions are satisfied:*

(i) $x(R)=(1,0),x(\emptyset)=(0,1)$;
(ii) if $A\cap B=\emptyset$ *then* $x(A)\odot x(B)=(\mathbf{0},\mathbf{1})$, *and* $x(A\cup B)=x(A)\oplus_E x(B)$;
(iii) $A_n\nearrow A\Rightarrow x(A_n)\nearrow x(A)$.

Theorem 2. *Let $x:\mathcal{B}(R)\to\mathcal{F}$ be an E-observable, $\mathcal{P}=[\mathcal{P}^\flat,\mathcal{P}^\sharp]:\mathcal{F}\to\mathcal{J}$ be an E-probability. Then the functions $\mathcal{P}^\flat\circ x:\mathcal{B}(R)\to[0,1],\mathcal{P}^\sharp\circ x:\mathcal{B}(R)\to[0,1]$, are probability measures.*

Proof. The proof is straight forward. □

Theorem 3. *Let $x:\mathcal{B}(R)\to\mathcal{F}$ be an E-observable, $x(A)=(x^\flat(A),1-x^\sharp(A))$; $\omega\in\Omega$. Then the functions $p^\flat_\omega,p^\sharp_\omega:\mathcal{B}(R)\to[0,1]$ defined by*

$$p^\flat_\omega(A)=\frac{a^{x^\flat(A)(\omega)}-1}{a-1};p^\sharp_\omega(A)=\frac{a^{x^\sharp(A)(\omega)}-1}{a-1}$$

are probability measures.

Proof. Theorem follows by the Theorem 2.7 in [8]. □

Definition 4. *Let $x,y:\mathcal{B}(R)\to\mathcal{F}$ be E-observables. By the p-joint E-observable h of x,y we understand a mapping $h:\mathcal{B}(R^2)\to\mathcal{F}$ satisfying the following conditions*

(i) $h(R^2)=(\mathbf{1},\mathbf{0});h(\emptyset)=(\mathbf{0},\mathbf{1})$;
(ii) if $A\cap B=\emptyset$ *then* $h(A)\odot h(B)=(\mathbf{0},\mathbf{1})$ *and* $h(A\cup B)=h(A)\oplus_E h(B)$;
(iii) $A_n\nearrow A\Rightarrow h(A_n)\nearrow h(A)$;
(iv) $h(C\times D)=x(C).y(D)$ *for any* $C,D\in\mathcal{B}(R)$, *where* $(\mu_C,\nu_C).(\mu_D,\nu_D)=$
$(\log_a(\frac{(a^{\mu_C}-1)(a^{\mu_D}-1)}{a-1}+1)\wedge 1,1-\log_a(\frac{(a^{(1-\nu_C)}-1)(a^{(1-\nu_D)}-1)}{a-1}+1)\wedge 1)$

Remark 2. Analogously we can extend Definition 4 for finite collection of E-observables.

Theorem 4. *To any E-observables $x, y : \mathcal{B}(R) \to \mathcal{F}$ there exists their p-joint E-observable $h : \mathcal{B}(R^2) \to \mathcal{F}$.*

Proof. For fixed $\omega \in \Omega$ we define

$$p^\flat_\omega(A) = \tfrac{a^{x^\flat(A)(\omega)}-1}{a-1}, \qquad p^\sharp_\omega(A) = \tfrac{a^{x^\sharp(A)(\omega)}-1}{a-1},$$

$$q^\flat_\omega(A) = \tfrac{a^{y^\flat(A)(\omega)}-1}{a-1}, \qquad q^\sharp_\omega(A) = \tfrac{a^{y^\sharp(A)(\omega)}-1}{a-1},$$

We showed in Theorem 3 that $p^\flat_\omega, p^\sharp_\omega, q^\flat_\omega, q^\sharp_\omega$ are probability measures on $\mathcal{B}(R)$. Let us construct the following $p^\flat_\omega \times q^\flat_\omega : \mathcal{B}(R) \to [0,1], p^\sharp_\omega \times q^\sharp_\omega : \mathcal{B}(R) \to [0,1]$, and let us defined

$$h^\flat(A)(\omega) = \log_a((a-1)p^\flat_\omega \times q^\flat_\omega(A) + 1),$$
$$h^\sharp(A)(\omega) = \log_a((a-1)p^\sharp_\omega \times q^\sharp_\omega(A) + 1),\ h(A) = (h^\flat(A), 1 - h^\sharp(A)).$$

We shows, that $h : \mathcal{B}(R^2) \to \mathcal{F}$ holds properties $(i) - (iii)$:
(i) $h(R^2) = (h^\flat(R^2), 1 - h^\sharp(R^2))$, so for each $\omega \in \Omega$ there holds

$$h^\flat(R^2)(\omega) = \log_a((a-1)p^\flat_\omega \times q^\flat_\omega(R^2) + 1) = \log_a((a-1)+1) = 1,$$

and also there holds $h^\sharp(R^2)(\omega) = \mathbf{1}$ and therefore

$$h(R^2) = (h^\flat(R^2), 1 - h^\sharp(R^2)) = (\mathbf{1,0}).$$

Analogously $h(\emptyset) = (h^\flat(\emptyset), 1 - h^\sharp(\emptyset)) = (\mathbf{0,1})$.
(ii) Let $A \cap B = \emptyset$, then because the function $\varphi_E(u) = \frac{a^u-1}{a-1}$ is increasing, convex bijection for $a > 1$, the following holds for each $\omega \in \Omega$

$$\begin{aligned}
h^\flat(A)(\omega) + h^\flat(B)(\omega) &= \log_a\Big((a-1)(p^\flat_\omega \times q^\flat_\omega(A)) + 1\Big) + \\
&+ \log_a\Big((a-1)(p^\flat_\omega \times q^\flat_\omega(B)) + 1\Big) = \\
&= \log_a\left((a-1)(\int_0^1 q^\flat_\omega(A^x)dp^\flat_\omega(x)) + 1\right) + \\
&+ \log_a\left((a-1)(\int_0^1 q^\flat_\omega(B^x)dp^\flat_\omega(x)) + 1\right) \leq \\
&\leq \log_a\left((a-1)(\int_0^1 q^\flat_\omega(A^x)dp^\flat_\omega(x) + \int_0^1 q^\flat_\omega(B^x)dp^\flat_\omega(x)) + 1\right) = \\
&= \log_a\left((a-1)(\int_0^1 q^\flat_\omega((A \cup B)^x)dp^\flat_\omega(x)) + 1\right) = \\
&= h^\flat(A \cup B)(\omega) \leq 1,
\end{aligned}$$

let us denote by $A^x = \{(x,y) \in A\}$ the x cut of A.

Analogously there holds $h^\sharp(A)(\omega) + h^\sharp(B)(\omega) \leq 1$ and then

$$h(A) \odot h(B) = (h^\flat(A), 1 - h^\sharp(A)) \odot (h^\flat(B), 1 - h^\sharp(B)) =$$
$$((h^\flat(A) + h^\flat(B) - 1) \vee 0, (1 - h^\sharp(A) + 1 - h^\sharp(B)) \wedge 1) = \mathbf{(0,1)}.$$

Let us show that

$$\begin{aligned}
h(A) \oplus_E h(B) &= (h^\flat(A), 1 - h^\sharp(A)) \oplus_E (h^\flat(B), 1 - h^\sharp(B)) = \\
&= (\log_a \left((a-1) \left(\frac{a^{h^\flat(A)} - 1}{a-1} + \frac{a^{h^\flat(B)} - 1}{a-1} \right) + 1 \right), \\
&1 - \log_a \left((a-1) \left(\frac{a^{1-h^\sharp(A)} - 1}{a-1} + \frac{1 - a^{h^\sharp(B)} - 1}{a-1} \right) + 1 \right)) = \\
&= (\log_a \left((a-1)(p^\flat_\omega \times q^\flat_\omega(A) + p^\flat_\omega \times q^\flat_\omega(B)) + 1 \right), \\
&1 - \log_a \left((a-1)(p^\sharp_\omega \times q^\sharp_\omega(A) + p^\sharp_\omega \times q^\sharp_\omega(B)) + 1 \right)) = \\
&= (h^\flat(A \cup B), 1 - h^\sharp(A \cup B)) = h(A \cup B)
\end{aligned}$$

(iii) Let $A_n \nearrow A$ then $(A_n)^x \nearrow A^x$ then $q^\flat_\omega((A_n)^x) \nearrow q^\flat_\omega(A^x)$, and so

$$\begin{aligned}
h^\flat(A_n)(\omega) &= \log_a \left((a-1)(p^\flat_\omega \times q^\flat_\omega(A_n)) + 1 \right) = \\
&= \log_a \left((a-1) \int_0^1 q^\flat_\omega((A_n)^x) dp^\flat_\omega(x) + 1 \right) \nearrow \\
&\nearrow \log_a \left((a-1)(\int_0^1 q^\flat_\omega(A^x) dp^\flat_\omega(x)) + 1 \right) = \\
&= p^\flat_\omega \times q^\flat_\omega(A) = h^\flat(A)(\omega),
\end{aligned}$$

analogously

$$h^\flat(A_n)(\omega) \nearrow h^\flat(A)(\omega).$$

And so

$$h(A_n) = h^\flat(A_n), 1 - h^\sharp(A_n) \nearrow h^\flat(A), 1 - h^\sharp(A) = h(A).$$

Finally we prove (iv):

$$\begin{aligned}
x(C).y(D) &= (x^\flat(C), 1 - x^\sharp(c)).(y^\flat(D), 1 - y^\sharp(D)) = \\
&= (\log_a \left((a-1)(\frac{(x^\flat(C) - 1)}{a-1} . \frac{(y^\flat(D) - 1)}{a-1}) + 1 \right), \\
&1 - \log_a \left((a-1) \frac{(x^\sharp(C) - 1)}{a-1} . \frac{(y^\sharp(D) - 1)}{a-1} + 1 \right)) = \\
&= (\log_a \left((a-1)(p^\flat_\omega(C).q^\flat_\omega(D)) - 1 \right), 1 - \log_a \left((a-1)(p^\sharp_\omega(C).q^\sharp_\omega(D)) + 1 \right)) = \\
&= (\log_a \left((a-1)(p^\flat_\omega \times q^\flat_\omega(C \times D)) + 1 \right), 1 - \log_a \left((a-1)(p^\sharp_\omega \times q^\sharp_\omega(C \times D)) + 1 \right)) = \\
&= (h^\flat(C \times D), 1 - h^\sharp(C \times D)) = h(C \times D). \qquad \square
\end{aligned}$$

4 Application of E-Observables

Let us mention one version of Central limit theorem: let $(\xi_i)_{i=1}^{\infty}$ be a sequence of independent, equally distributed, square integrable random variables,

$$E(\xi_i) = a, \sigma^2(\xi_i) = \sigma^2 \text{ for all } i \in N.$$

Then for any $t \in R$ there holds

$$\lim_{n\to\infty} p\left(\left\{\omega; \frac{\overline{\zeta_n}(\omega) - a}{\sigma}\sqrt{n} < t\right\}\right) = \Phi(t).$$

Here $\overline{\zeta_n} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ and $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du$.
Now we are going to formulate an analogous assertion for E-observables.
First, we shall mention some useful definitions:

Definition 5. *For any probability $\mathcal{P} = [\mathcal{P}^\flat, \mathcal{P}^\sharp] : \mathcal{F} \to \mathcal{J}$ and any E-observable $x : \mathcal{B}(R) \to \mathcal{F}$ we define the expected values by*

$$E_\flat(x) = \int_R t d\mathcal{P}_x^\flat(t); \quad E_\sharp(x) = \int_R t d\mathcal{P}_x^\sharp(t)$$

and the variances by

$$\sigma_\flat^2(x) = \int_R (t - E_\flat(x))^2 d\mathcal{P}_x^\flat(t); \quad \sigma_\sharp^2(x) = \int_R (t - E_\sharp(x))^2 d\mathcal{P}_x^\flat(t),$$

where $\mathcal{P}_x^\flat = \mathcal{P}^\flat \circ x, \mathcal{P}_x^\sharp = \mathcal{P}^\sharp \circ x$, assuming that the integrals exist.

Assume $T = (\xi_1, ..., \xi_n) : \Omega^n \to R^n$ is a random vector and $g : R^n \to R$ is a Borel measurable function (e.g. $g(u_1, ..., u_n) = \frac{1}{n}\sum_{i=1}^{n} u_i$). Then $g(\xi_1, ..., \xi_n) = g \circ T : \Omega^n \to R^n$, is a transformation of T. Hence we get the following formula

$$(g \circ T)^{-1}(A) = T^{-1}(g^{-1}(A))$$

for any $A \in \mathcal{B}(R)$. The formula justifies the following definition.

Definition 6. *Let $g_n : R^n \to R$ be a Borel function, $x_1, ..., x_n : \mathcal{B}(R) \to \mathcal{F}$ be E-observables, $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ their joint observable. Then the g_n-transformation of h_n is an E-observable $y_n : \mathcal{B}(R) \to \mathcal{F}$ given by $y_n(A) = h_n(g_n^{-1}(A))$ for any $A \in \mathcal{B}(R)$.*

Definition 7. *Let $(x_n)_{n=1}^{\infty}$ be a sequence of E-observables,$(h_n)_{n=1}^{\infty}$ be a sequence of the joint E-observables $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ of $x_1, x_2, ..., x_n$ (for $n \in N$), $m : \mathcal{F} \to [0, 1]$ be a P-state. The sequence $(x_n)_{n=1}^{\infty}$ is independent (with respect to m), if for any $n \in N$ and any $C_1, C_2, ..., C_n \in \mathcal{B}(R)$ there holds*

$$m(h_n(C_1 \times C_2 \times ... \times C_n)) = m(x_1(C_1))...m(x_n(C_n)).$$

Definition 8. *A sequence* $(x_n)_{n=1}^{\infty}$ *of E-observables is equally distributed, if* $m(x_n(A)) = m(x_1(A))$ *for any* $n \in N$ *and* $A \in \mathcal{B}(R)$.

Theorem 5. (Central limit theorem)
Let $(x_n)_{n=1}^{\infty}$ be a sequence of independent, equally distributed, square integrable E-observables, where $E_\flat(x_n) = a^\flat, (E_\sharp(x_n) = a^\sharp)$ $\sigma_\flat^2(x_n) = \sigma_\flat^2$, $(\sigma_\sharp^2(x_n) = \sigma_\sharp^2)$ for each $n \in N$. Then for any $t \in R$ there hold

$$\lim_{n\to\infty} \mathcal{P}^\flat\left(\frac{x_1+\ldots+x_n-na^\flat}{\sigma_\flat\sqrt{n}}((-\infty,t))\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}}\,du,$$

$$(\lim_{n\to\infty} \mathcal{P}^\sharp\left(\frac{x_1+\ldots+x_n-na^\sharp}{\sigma_\sharp\sqrt{n}}((-\infty,t))\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}}\,du).$$

Proof. Theorem follows by the Theorem 4.1 in [8]. □

Acknowledgements

The paper was supported by Grant VEGA 1/0539/08.

References

1. Atanassov, K.: Intuitionistic Fuzzy Sets: Theory and Applications. Physica Verlag, New York (1999)
2. Atanassov, K.: Riečan, B.: On two new types of probability on IF-events. In: Proc.of the First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering, Warsawa (2007)
3. Čunderlíková - Lendelová, K., Riečan, B.: The probability theory on B-structures, pp. 355–383. EXIT, Warsawa (2008)
4. Gerstenkorn, T., Manko, J.: Probabilities of intuitionistic fuzzy events. In: Hrzniewicz, O., et al. (eds.) Issues in Inteligent Systems: Paradigms, pp. 63–58. EXIT, Warsawa
5. Grzegorzewski, P., Mrowka, E.: Probability of intuitionistic fuzzy events. In: Grzegorzewski, P., et al. (eds.) Soft Methods in Probablity, Statistics and Data Analysis, pp. 105–115. Physica Verlag, New York (2002)
6. Lendelová, K.: Strong law of large numbers for IF-events. In: Proc. Eleventh Int. Conf. IPMU, Paris, pp. 2363–2366 (2006)
7. Renčová, M., Riečan, B.: Probability on IF - sets: an elementary approach. In: First International Workshop on Intuitionistic Fuzzy Sets, Generalized Nets and Knowledge Engineering, pp. 8–17. University of Westminster, London (2006)
8. Renčová, M.: On the $\varphi-$probability and $\varphi-$observables. Submitted to Fuzzy Sets and Systems (2008)
9. Riečan, B.: General form of M-probabilities on IF-events. In: Proc. of IPMU, Spain, pp. 1675–1677 (2008)
10. Riečan, B.: On a problem of Radko Mesiar: general form of IF probabilities. Fuzzy Sets and Systems 152, 1485–1490 (2006)
11. Riečan, B.: Probability theory on intuitionistic fuzzy sets. In: Aguzzoli, S., Ciabattoni, A., Gerla, B., Manara, C., Marra, V. (eds.) ManyVal 2006. LNCS, vol. 4460. Springer, Heidelberg (2007)

Rough Ensemble Classifier: A Comparative Study

Suman Saha, Chivukula A. Murthy, and Sankar K. Pal

Center for Soft Computing Research, Indian Statistical Institute, Kolkata, India
{ssaha_r,murthy,sankar}@isical.ac.in

Abstract. Combining the results of a number of individually trained classification systems to obtain a more accurate classifier is a widely used technique in pattern recognition. In this article, we have introduced a rough set based meta classifier (RSM). Theoretical analysis of the proposed RSM is carried out in relation to Bayes classifier since Bayes classifier is the best classifier. It has been shown that the performance of the meta classifier is at least as good as the best constituent classifier, and if one of the base classifiers of RSM converges to Bayes then RSM converges to Bayes classifier. Experimental studies show that the meta classifier improves accuracy of classification and beats other ensemble approaches in accuracy by a decisive margin, thus demonstrating the theoretical results.

Keywords: Classification, Rough set, Ensemble classifier, Combination method, Redundant classifiers, Meta data, Base-level classifier.

1 Introduction

Ensemble of classification is a method for improving accuracy in supervised learning. Ensemble classifier combines the decisions of its constituent individual classifiers to results in a new classification rule. One of the most active areas of research in supervised learning has been to study methods for combining the decisions for improving classification accuracy [1].

The work presented here focuses on performance of a rough set based ensemble classifier which combines the predictions of base-level classifiers induced by applying different learning algorithms to a single data set. It adopts the stacking framework, where we have to learn how to combine the base-level classifiers.

While ensemble classifiers are accurate classifiers, there exists some problems that may limit their practical application. One problem is the need for a large number of base classifiers for achieving good performance. So an important line of research, therefore, is to find ways of converting the informations from base classifier into less redundant representations. Our approach tries to solve the problem of representing less redundant ensemble classifier by using a rough set based attribute reduction on the granular meta data generated by the decisions of base classifiers (section 2.3).

It may be noted that, no ensemble classification technique existing in the literature ensures that it will work better than its constituent individual classifiers.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 116–123, 2009.

In our article, it has been shown that the performance of the meta classifier is at least as good as the best constituent classifier (section 2.3).

Another difficulty with an ensemble classifier is that the classifier provides little insight into the correctness of the decision making process for the classification task. In this article, it has been shown that the rough set based meta classifier is an optimal combination technique with respect to a quality measure for evaluating combination techniques (section 2.3).

Comparison with respect to Bayes classifier is an effective way of evaluating the performance of a classifier. It has been shown that if one of the base classifiers of RSM converges to Bayes then the RSM converges to Bayes classifier (section 3).

In order to realize the specified objectives, section 2 describes the rough set theoretic analysis of the meta-level patterns and a new meta classifier termed as RSM. Sections 3 present the theoretical results of an RSM. Section 3 presents comparison of RSM with relation to Bayes classifier. Finally, the classification results of RSM on different data sets are reported in section 4.

2 Pattern Analysis of Meta-level Data in Rough Set Framework

2.1 Rough Set Theory

Rough set theory was developed by Pawlak in the early 1980's [2]. The rough set theory is a mathematical framework for analyzing granular data. It deals with the classificatory analysis of data tables. The main goal of the rough set analysis is to synthesize approximation of concepts from the acquired data. The theory sees the data in terms of equivalence classes i.e. partitions. A rough set is a set of objects that cannot be uniquely represented by these equivalence classes since the set only partly overlaps with at least one of them. It may be approximately described either by the equivalence classes completely contained in the set (the lower approximation) or the equivalence classes with at least one object in the set [3] (the upper approximation).

2.2 Ensemble Classification Problem in Rough Set Terminology

In the problem of classification we train a learning algorithm and validate the trained algorithm. This task is performed, using some test-train split on a given labelled dataset. In the notion of rough set, let U be the given categorized dataset and $P = \{C_1, C_2, ..., C_k\}$ where $C_i \neq \phi$ for $i = 1, 2, 3..., k$, $\cup_{i=1}^{k} C_i = U$ and $C_i \cap C_j = \phi$ for $i \neq j$ and $i, j = 1, 2, 3..., k$, be a partition on U which provides the given k categories of U. Output of a classifier determines a new partition on U. In rough set terminology each class of the given partition P is a given concept about dataset and output of a classifiers determines new concepts about the same dataset. The given concepts can be expressed approximately by upper and lower approximations constructed by generated concepts.

2.3 Rough Set Meta Classifier (RSM)

The rough set meta classifier (RSM) is designed to extract decision rules from trained classifier ensembles that perform classification tasks [4]. RSM utilizes trained ensembles to generate a number of instances consisting of prediction of individual classifiers as conditional attribute values and actual classes as decision attribute values. Then a decision table is constructed using all the instances with one instance in each row. Once the decision table is constructed, rough set attribute reduction is performed to determine core and minimal reducts. The classifiers corresponding to a minimal reduct are then taken to form classifier ensemble for RSM classification system. From the minimal reduct, the decision rules are computed by finding mapping between decision attribute and conditional attributes. These decision rules obtained by rough set technique are then used to perform classification tasks. Following theorems exists in this regard.

- **Theorem 1.** Rough set based combination is an optimal classifier combination technique [4].
- **Theorem 2.** The performance of the rough set based ensemble classifier is at least same as every one of its constituent single classifiers [4].

3 Comparison of RSM in Relation to Bayes Classifier

If prior probabilities and conditional density functions of the classes are known, Bayes decision rule is known to be the best decision rule, since it minimizes the probability of misclassification. However, in reality, the density functions are generally not known. Thus several classifiers exist for doing classification. However one needs to compare the performance of any proposed classifier with Bayes classifier to verify its validity. The comparison may be done in several ways. One way is to construct a few artificial data sets where the prior probabilities and class conditional probability density functions are known, and then compare the error rates of Bayes classifier with the proposed classifier. Another way is to show theoretically that as the number of observations tends to infinity, the error rate of the proposed classifier goes to the error probability of the Bayes classifier, whatever may be the prior probability and class conditional probability density functions are. In this article, we have chose the later one. In this section we shall establish a few theoretical results in this regard.

3.1 Notations, Definitions and Assumptions

(1) Let the number of instances be n, number of classes be k, number of base level classifiers be m and U_n be the set of n given observations. Let Ω be the space of all possible instances. Let $\pi_1, \pi_2, \ldots, \pi_k$ be the prior probabilities of the classes and $p_1, p_2, \ldots, p_k$ be the corresponding continuous class conditional density functions. Let $(\Omega_1, \Omega_2, \ldots, \Omega_k)$ be a partition of the Ω. Let for the said partition, the corresponding decision rule is: x is put in class i if $x \in \Omega_i$, $1 \leq i \leq k$. Then the probability of misclassifications of the partition $(\Omega_1, \Omega_2, \ldots, \Omega_k)$ denoted by $\varepsilon(\Omega_1, \Omega_2, \ldots, \Omega_k)$, is defined as: $\varepsilon(\Omega_1, \Omega_2, \ldots, \Omega_k) = \Sigma_{i=1}^{k} \pi_i \int_{\Omega_i^c} p_i(x)dx$.

(2) Let $p(x) = \Sigma_{i=1}^{k} \pi_i p_i(x)$, p is known as mixture density function. Let $x_1, x_2, \ldots, x_n, \ldots$ be a sequence of independent and identically distributed random vectors following p. Let $y_i = 1$ if x_i is misclassified using the decision rule for the partition $(\Omega_1, \Omega_2, \ldots, \Omega_k)$, $y_i = 0$ otherwise. Then it can be shown from Strong Law of Large Numbers that $\hat{\varepsilon_n}(\Omega_1, \Omega_2, \ldots, \Omega_k) = \frac{1}{n}\Sigma_{i=1}^{n} y_i \rightarrow_{a.e} \varepsilon(\Omega_1, \Omega_2, \ldots, \Omega_k)$. It is also true that $E(\frac{1}{n}\Sigma_{i=1}^{n} y_i) = \varepsilon(\Omega_1, \Omega_2, \ldots, \Omega_k)\ \forall\ \ n$ where $\frac{1}{n}\Sigma_{i=1}^{n} y_i$ is known as error rate for the partition $(\Omega_1, \Omega_2, \ldots, \Omega_k)$.

(3) Let Bayes classifier and RSM for n instances be represented by f_B^n and f_{RSM}^n respectively. Let partitions on the sample space and population corresponding to f_B^n be $(P_{1n}^B, \ P_{2n}^B, \ldots, P_{kn}^B)$, $(\Omega_1^B, \Omega_2^B, \ldots, \Omega_k^B)$ respectively. And partitions on the sample space and population corresponding to f_{RSM}^n be $(P_{1n}^{RSM}, P_{2n}^{RSM}, \ldots, P_{kn}^{RSM})$ and $(\Omega_{1n}^{RSM}, \Omega_{2n}^{RSM}, \ldots, \Omega_{kn}^{RSM})$ respectively. These notations have been followed throughout this section.

(4) Bayes decision rule is: put x in i^{th} class if $\pi_i p_i(x) \geq \pi_j p_j(x)\ \forall i \neq j$ (resolve ties arbitrarily). It has been shown that the Bayes decision rule provides minimum misclassification probability [5]. Error probability of any partition $(\Omega_1^B, \Omega_2^B, \ldots, \Omega_k^B)$ of Ω based on n observations corresponding to Bayes decision rule, is denoted by ε_0.

(5) Let X and Y be two compact subsets of $\mathcal{R}^a$, where a is the number of features. The Hausdorff distance $d_H(X, Y)$ is the minimal number r such that the closed r-neighborhood of X contains Y and the closed r-neighborhood of Y contains X. In other words, if $d(x, y)$ denotes the Euclidean distance in $\mathcal{R}^a$, then $d_{\mathrm{H}}(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\}$. The Hausdorff distance between two general bounded subsets can be defined as the Hausdorff distance between their closures.The Hausdorff distance between any two sets with the same closure is zero.

(6) A classifier f is said to be convergent to Bayes classifier if $d_H(\Omega_{in}^O \cap U_n, \Omega_i^B \cap U_n) \rightarrow 0 \quad as \quad n \rightarrow \infty$, where $(\Omega_{1n}^O, \Omega_{2n}^O, \ \ldots, \ \Omega_{kn}^O)$ is the partition of Ω corresponding to f based on n observations. Then it follows that $\varepsilon_n(\Omega_{1n}^O, \Omega_{2n}^O, \ldots, \Omega_{kn}^O) \rightarrow_{a.e} \varepsilon_0$.

(7) We assume that, for a partition $(\Omega_1, \Omega_2, \ldots, \Omega_k)$ of Ω, each Ω_i is compact. This assumption is possible because sets considered in real life are generally bounded and inclusion of boundary points in a set would generally not change their probability because the probability density functions are continuous.

3.2 Theorems

Theorem 3: If one of the base classifiers of RSM converges to Bayes then RSM converges to Bayes classifier.

Proof: Let $\Omega_n^O = \{\Omega_{1n}^O, \Omega_{2n}^O, \ldots, \Omega_{kn}^O\}$ be the partition corresponding to the classifier for the given set of n points which converges to Bayes. Therefore, $d_H(\Omega_{in}^O \cap U_n, \Omega_i^B \cap U_n) \rightarrow 0 \quad as \quad n \rightarrow \infty$, where d_H is the Hausdorff metric. Now, since SP_t belongs to a finer partition of Ω than Ω_n^o, for any $SP_t \in SP$

$$SP_t \cap (\Omega_{in}^O \cap U_n) = SP_t \quad \text{or} \quad SP_t \cap (\Omega_{in}^O \cap U_n) = \phi.$$

Since, $d_H(\Omega_{in}^O \cap U_n, \Omega_i^B \cap U_n) \to 0 \;\; as \;\; n \to \infty$.

Therefore, $\Omega_{in}^O = \{x : \pi_i p_i(x) \geq \pi_j p_j(x), \forall 1 \leq i, j \leq k \& j \neq i\}$ almost everywhere as $n \to \infty$.

For any $SP_t \in SP$, SP_t is assigned to class i

$$\Leftrightarrow |SP_t \cap P_i| \geq |SP_t \cap P_j| \;\; \forall \;\; j \neq i,$$

where $|.|$ denotes cardinality of a set. Now, $\frac{1}{n}\sum_{x \in SP_t} P(c_i|x)$ is an unbiased estimate of $\frac{|SP_t \cap P_i|}{n}$ and $\frac{1}{n}\sum_{x \in SP_t} P(c_j|x)$ is an unbiased estimate of $\frac{|SP_t \cap P_j|}{n}$. Therefore,

$$\sum_{x \in SP_t} P(c_i|x) \geq \sum_{x \in SP_t} P(c_j|x) \;\; \forall \;\; j \neq i.$$

Let $\xi_i = \{SP_t : |SP_t \cap P_i| \geq |SP_t \cap P_j| \;\; \forall j \neq i\}$ and $\alpha_i = \bigcup_{SP_t \in \xi_i} SP_t \;\; \forall i$.

Note that, $|SP_t \cap P_i| \geq |SP_t \cap P_j| \;\; \forall j \neq i$ and $SP_t \cap (\Omega_{j_0 n}^O \cap U_n) = SP_t$ for some $j_0 \neq i$is not possible, since $\forall x \in \Omega_{j_0 n}^O \cap U_n \Rightarrow P(c_{j_0}|x) \geq P(c_i|x) \;\; \forall i \neq j_0$ as $n \to \infty \Rightarrow |SP_t \cap P_j| \geq |SP_t \cap P_i| \;\; \forall i \neq j$ as $n \to \infty$.

Similarly, $|SP_t \cap P_i| \leq |SP_t \cap P_j| \;\; \forall j \neq i$ and $SP_t \cap (\Omega_{in}^O \cap U_n) = SP_t$ is not possible for sufficiently large n.

Now we have

$$\sum_{x \in \alpha_i} P(c_i|x) \geq \sum_{x \in \alpha_j} P(c_j|x) \;\; \forall \;\; j \neq i \quad \text{for sufficiently large} \;\; n.$$

Average misclassified points for class i or misclassification rate for class i for RSM classification,

$$= \frac{1}{n} \sum_{x \in \alpha_i} \sum_{t=1, t \neq i}^{k} P(c_t|x) \quad \text{for sufficiently large} \;\; n.$$

Average misclassified points or misclassification rate for RSM classification,

$$= \frac{1}{n} \sum_{i=1}^{k} \sum_{x \in \alpha_i} (1 - P(c_i|x)) \quad \text{for sufficiently large} \;\; n.$$

This is a consistent unbiased estimate of,

$$\frac{1}{n} \sum_{i=1}^{k} \sum_{x \in \Omega_i^B \cap U_n} (1 - \frac{\pi_i p_i(x)}{\sum_{t=1}^{k} \pi_t p_t(x)}) \to_{a.e.} \varepsilon_0.$$

Therefore,

$$\frac{1}{n} \sum_{j=1}^{k} \sum_{x \in \alpha_i} \frac{\sum_{i=1, i \neq j}^{k} \pi_i p_i(x)}{\sum_{t=1}^{k} \pi_t p_t(x)} \to_{a.e.} \varepsilon_0.$$

Hence the theorem is proved.

3.3 Remark

Generally k-nearest neighbor classifier estimates the Bayes error rate within a bound. Therefore considering k-nearest neighbor classifier as one of the base classifiers of RSM is an easy way of getting good classification results using RSM.

4 Experimental Results

The main goal of the experiments we performed is to evaluate the performance of RSM, especially in comparison to other methods for combining classifiers, such as bagging, boosting, voting and stacking with other machine learning approaches. We also investigate the use of different meta-level attributes in RSM. We performed experiments on a collection of data sets from the UCI Repository of Machine Learning Databases and text corpus data sets. These data sets have been widely used in other comparative studies. We listed all the base-level and meta-level learning algorithms used in this study and compared the performance of meta-level learning algorithms.

4.1 Base-Level Algorithms

Five learning algorithms have been used in the base-level experiments: tree-learning algorithm C4.5 [6], the rule-learning algorithm CN2 [7], the k-nearest neighbor (k-NN) algorithm [8], support vector machine (SVM) [9], and naive bayes method [10]. All algorithms have been used with their default parameters settings. The output of each base-level classifier for each example in the test set is the predicted class.

4.2 Meta-level Algorithms

At the meta-level, we evaluate the performances of six different algorithms for constructing ensembles of classifiers. Four of these make use of exactly the same set of five base-level classifiers induced by the five algorithms from the previous section. These four classifier combination techniques are P-vote [11], SCANN [12], MDT [13] and DECORATE [14]. P-vote algorithm performs stacking [11] with simple plurality vote. SCANN [12] performs stacking with nearest neighbor after analyzing dependencies among the base-level classifiers. MDT corresponds to meta decision tree learning algorithm [13]. DECORATE stance for Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples [14]. In addition, boosting [15] and bagging [16][17] of decision trees are considered, which create larger ensembles (100 trees). The competitive ensemble classifiers considered here are well known for their good performance. Accuracy of RSM with five heterogeneous base classifiers, and the accuracies of other methods are shown in table 1.

Table 1. Percentage of correct classification of RSM with other ensemble classifiers

Data set	Bagging	AdaBoost	P-vote	SCANN	MDT	DECORATE	S-RSM
australian	86.75	86.77	86.41	86.41	85.30	85.05	88.23
balance	83.46	75.47	85.95	89.71	92.91	92.92	94.04
breast-w	95.54	96.78	96.22	96.22	95.30	97.37	97.89
bridges-td	85.27	81.92	86.14	86.14	84.00	84.49	87.31
car	93.61	96.90	93.81	94.12	95.38	94.49	97.53
chess	99.44	99.68	99.05	99.05	99.48	99.44	99.62
diabetes	76.59	74.41	76.73	76.71	76.68	77.31	78.27
echo	69.29	65.29	68.63	67.60	67.29	67.46	70.50
german	75.54	74.39	74.42	74.63	73.21	75.45	76.41
glass	76.03	79.23	71.29	71.61	69.91	67.51	80.65
heart	81.04	79.89	83.48	84.03	80.02	84.20	85.28
hepatitis	82.98	83.50	82.64	84.12	81.64	83.37	85.13
hypo	99.24	98.94	98.99	99.18	99.22	99.28	99.51
image	97.65	98.64	97.63	98.05	97.38	96.72	99.07
ionosphere	93.02	93.99	91.37	92.42	90.19	89.88	94.22
iris	94.20	94.47	96.00	96.14	97.53	97.53	98.72
soya	93.23	93.38	93.57	93.80	93.87	94.32	95.03
tic-tac	94.88	99.15	90.06	96.28	97.20	98.51	99.18
vote	96.32	95.68	96.04	96.04	94.74	96.16	97.86
waveform	83.66	84.77	85.65	85.89	85.94	86.02	87.94
wine	96.17	96.91	98.48	98.59	97.96	97.96	99.24

4.3 Statistical Significance of the Results Corresponding to Table 1

Tests of significance were performed for the equality of means (of the accuracies) obtained using the RSM and the other ensemble classification techniques compared. Since both mean pairs and the variance pairs are unknown and different, a generalized version of t-test is appropriate in this context. The above problem is the classical Behrens-Fisher problem in hypothesis testing [18]. The level of significance considered, is 0.01. The accuracies obtained by our method are found to be significantly better than all other methods considered, on all datasets except for chess data, where our method is found to be not significantly different with AdaBoost classifier.

5 Conclusion

RSM views a classifier's output as a partition on the dataset. It tries to find the effectiveness of classifiers to build the combination classifier. Our method uses decision rules to make final prediction about the category of data instances. It has been shown mathematically that RSM is an optimal combination scheme, and is at least as good as the best base learner in terms of classification accuracy.

Experimental studies show that it provides more accurate classifications of the datasets than other ensemble classification techniques.

Acknowledgment

This paper was done when one of the authors, S. K. Pal, was a J.C. Bose Fellow of the Government of India.

References

1. Dietterich, T.G.: Machine-learning research: Four current directions. The AI Magazine 18(4), 97–136 (1998)
2. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)
3. Wang, G., Liu, Q., Yao, Y., Skowron, A. (eds.): RSFDGrC 2003. LNCS, vol. 2639, pp. 41–48. Springer, Heidelberg (2003)
4. Saha, S., Murthy, C.A., Pal, S.K.: Rough set based ensemble classifier for web page classification. Fundamentae Informetica 76(1-2), 171–187 (2007)
5. Fukunaga, K.: Introduction to statistical pattern recognition, 2nd edn. Academic Press Professional, Inc., San Diego (1990)
6. Ross Quinlan, J.: C4.5: Programs for machine learning, p. 302. Morgan Kaufmann Publishers Inc., San Francisco (1993)
7. Clark, P., Boswell, R.: Rule induction with CN2: Some recent improvements. In: Proc. Fifth European Working Session on Learning, pp. 151–163 (1991)
8. Wettschereck, D., Dietterich, T.G.: An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms. Machine Learning 19(1), 5–27 (1995)
9. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)
10. Gama, J.: Iterative bayes. Theoretical Computer Science 292, 417–430 (2003)
11. Wolpert, D.H.: Stacked generalization. Neural Networks 5, 241–259 (1992)
12. Merz, C.J.: Using Correspondence Analysis to Combine Classiers. Machine Learning 36(1/2), 33–58 (1999)
13. Zenko, B., Todorovski, L., Dzeroski, S.: A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods. In: ICDM 2001: Proceedings of the 2001 IEEE International Conference on Data Mining, pp. 669–670. IEEE Computer Society, Washington (2001)
14. Melville, P., Mooney, R.: Constructing diverse classifier ensembles using artificial training examples. In: Proc. of 18th Intl. Joint Conf. on Artificial Intelligence, Acapulco, Mexico, pp. 505–510 (2003)
15. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: International Conference on Machine Learning, pp. 148–156 (1996)
16. Breiman, L.: Bagging predictors. Machine Learning 24, 123–140 (1996)
17. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles. Machine Learning 51, 181–207 (2003)
18. Best, D.J., Rayner, J.C.W.: Welch's approximate solution for the Behrens-Fisher problem. Technometrics 29(2), 205–210 (1987)

A Fuzzy One Class Classifier for Multi Layer Model

Giosuè Lo Bosco and Luca Pinello

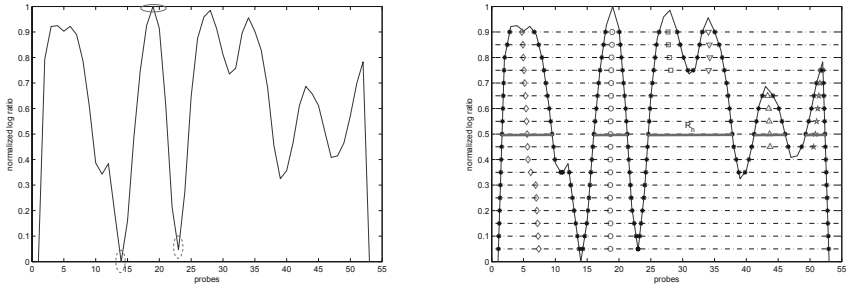
DMA - Università di Palermo, Italy
{lobosco,pinello}@unipa.it

Abstract. The paper describes an application of a *fuzzy one-class* classifier (FOC) for the identification of different signal patterns embedded in a noise structured background. The classification phase is applied after a preprocessing phase based on a Multi Layer Model (MLM) that provides a preliminary signal segmentation in an interval feature space. The FOC has been tested on synthetic and real microarray data in the specific problem of DNA nucleosome and linker regions identification. Results have shown, in both cases, a good recognition rate.

1 Introduction

Classification algorithms are usually based on a training set, where both positive and negative examples are considered. However, in many cases either only positive examples are available or the two classes are very much unbalanced. Within these hypothesis, we talk about Novelty or Outlier Detection, and more generally, One Class Classification [1]. One-class classifiers have been introduced in order to discriminate a target class from the rest of the feature space using only the target training examples. Generally, the classical approach is based on finding the smallest volume hypersphere (in the feature space) that encloses most of the training data. Many real cases needs the use of such approach to classification: documents analysis [2], masquerader intrusion detection [3],[4],[5] facial expression [6], online signature verification [7], signal identification [8], showing also, in [4,7] that the combination of several one-class classifiers may improve the overall performances. In this paper, a *fuzzy one class* classifier (FOC) [4], is applied in order to identify signal patterns slightly differentiated and embedded in a noise structured background. Examples of such kind of signal are those provided by microarray data of the *Saccharomyces cerevisiae* [9], where the goal, in this case, is the identification of nucleosomes and linker regions across DNA. This problem is very challenging because nucleosomes are the fundamental repeating units of eukaryotic chromatin and their organization can result in a variety of diseases, including cancer [10,11,12]. To measure nucleosome positions on a genomic scale a DNA microarray method has been developed [9], allowing, to represent microarray data as a signal of green/red ratio values that shows nucleosomes as peaks of about 150 base pairs long, surrounded by lower ratio values corresponding to linker regions. Note that the identification of nucleosomes as patterns around peaks, is not simple

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 124–131, 2009.



(a) (b)

Fig. 1. *(a) Input Signal:* The input signal is the logarithmic ratio of the *green channel* to *red channel* value for each spot of the microarray. Nucleosomes correspond to peaks (marked by black circle), surrounded by lower ratio values corresponding to linker regions (marked by dashed circles) that are nucleosome free. *(b) Pattern identification and extraction:* In this example 6 patterns are retrieved, identified by rhombus, circle, square, triangle down, triangle up, star. Each pattern identifier is replicated for each of its feature values and pointed in each one of its middle point.

because of noise in the data signal and large-scale trends in mean hybridization values. Therefore, a Hidden-Markov Model (HMM) approach has been used to discriminate two classes of patterns, the *well positioned nucleosomes* (WPN) and the *linkers* (LN). However, this approach suffers from the constrain imposed by its static topology, as a consequence lots of potential good input data are discarded from the analysis. Moreover, this method does not take into account the shape information of the green/red ratio values. For an example of an input signal which also shows the two pattern classes see Figure 1a.

The fuzzy one-class classifier is applied after a preprocessing phase, based on a Multi Layer Model (MLM) [13,14] that provides a preliminary signal segmentation in an interval feature space. In this work the FOC has been tested on real microarray data of the *Saccharomyces cerevisiae*, showing a good accordance with the HMM for nucleosome and linker regions discrimination. The paper is organized as follow: Section 2 describes the FOC classifier and its assessment; the MLM preprocessing phase is outlined in Section 3; in Section 4 the adaptation of the MLM on a particular biological problem is outlined and the results of the FOC on real biological data are shown; final remarks and discussion are given in Section 5.

2 Fuzzy One Class Classifier

A *classifier* for an M classes problem can be defined as a *mapping* from the set of observation $\mathbf{X}$ to a set of labels $Y = \{y_0, \cdots, y_{M-1}\}$. In the case of *supervised classification*, the classifier is based on a training (learning) set $T^{(m)}$ for each class $0 \leq m \leq M-1$, moreover, if $y_m \in \{0, 1\}$ we talk about *crisp classifiers*, otherwise, if $y_m \in [0, 1]$, about *fuzzy classifiers*.

In the specific case of binary classifiers ($M = 2$), the *fuzzy classifier* can be defined by a membership function $\mu_m : \mathbf{X} \rightarrow [0,1]$ for each $m = 0,1$. In particular, *one class classifiers* are binary classifiers able to learn by using the training set $T_U = \{t_1, t_2, .., t_R\}$ of the target class only. Generally, a classifier can be defined by using a proper *dissimilarity function*, δ between elements in X. By using δ, we can define the set $\Delta = \{\delta(x,x')|x,x' \in T_U, x \neq x'\}$ of all the difference between distinct elements of T_U, and estimate the p.d.f. of Δ, f_Δ by kernel density estimation. Finally, the proposed new *fuzzy one class* classifier, called FOC, is based on a fuzzy membership function μ_0 of an unknown element, x, defined as it follows:

$$\mu_0(x,a,b) = \begin{cases} 0, & \overline{\delta}(x,\cdot) \leq a \\ 2 * (b - \frac{\overline{\delta}(x,\cdot)}{b-a})^2, & a \leq \overline{\delta}(x,\cdot) \leq \frac{a+b}{2} \\ 1 - 2 * (\frac{\overline{\delta}(x,\cdot)-a}{b-a})^2, & \frac{a+b}{2} \leq \overline{\delta}(x,\cdot) \leq b \\ 1, & \overline{\delta}(x,\cdot) \geq b \end{cases}$$

Where $\overline{\delta}(x,\cdot) = \Sigma_{y \in T_U, x \neq y} \delta(x,y)/|T_U|$ is the average dissimilarity of x in T_U. Ultimately, the performance of the FOC depends on the values of a and b, such value can be set by properly using f_Δ, depending on the classification problem to solve. Finally, the defuzzification phasecan be done by a simple threshold at a fixed value α. More formally:

$$\chi_0(x,a,b,\alpha) = \begin{cases} 1, & \mu_0(x,a,b) > \alpha \\ 0, & otherwise \end{cases} \quad (1)$$

Where χ_0 defines the crisp membership to the target class, while $\chi_1 = 1 - \chi_0$ the membership of the class of negatives. In section 4 the details about the choice of a,b and α on the specific biological problem of nucleosome positioning is described.

3 Multi Layer Model

In this section an outline of the MLM for the analysis of mono-dimensional signal is provided. It can be considered a preliminary step that provides to the one-class classifier the proper feature space and input pattern to be classified. The MLM procedure is carried out as follows:

Preprocessing. A preprocessing is necessary in order to reduce the effect of the signal noise. Starting from the input signal, $\mathbf{S}$, each fragment S_t, $1 \leq t \leq T$, is convolved by a generic kernel window. After this process, $\mathbf{X}$ represents the convolved signal.

Training set and model construction. A model set $\mathbf{M} = \{X_t | X_t \subseteq \mathbf{X}\}$ is built by extracting subsignals X_t of the convolved signal $\mathbf{X}$ that satisfy a particular set of conditions. Such conditions are defined with respect to the shape of the pattern we want to identify. For example, in the case of the signals resulting

from tiling microarray data, we select particular bell shaped subsignals centered on the local maxima of the input signal.

Interval identification.The core of the method is the interval identification by considering H threshold levels t_h $(h = 1, ..., H)$ of the convolved signal $\mathbf{X}$. For each t_h a set of intervals $R_h = \left\{I_h^1, I_h^2, \cdots, I_h^{n_h}\right\}$ is obtained; where, $I_h^i = [b_h^i, e_h^i]$, where b_h^i, e_h^i are the lower and upper limits of the interval and $\mathbf{X}(b_h^i) = \mathbf{X}(e_h^i) = t_h$.

Interval merging, pattern definition and selection. This step is performed by taking into account again the shape of the pattern to classify. In particular, a set of rules $\mathcal{R}$ on the intervals extracted in the previous step is defined respecting several conditions inspired from the knowledge of the pattern to search (shape, persistence, etc.). The application of such rules on the set of intervals $\mathbf{R} = \{R_h | 1 \leq h \leq H\}$ groups such intervals, defining the patterns:

$$P_i = \{I_j^{i_j}, I_{j+1}^{i_{j+1}}, \cdots, I_{j+l}^{i_{j+l}} \mid \mathcal{R}(I_j^{i_j}, I_{j+1}^{i_{j+1}}, \cdots, I_{j+l}^{i_{j+l}}) \;\; is\; satisfied\}$$

Feature extraction. Each pattern P_i is identified by $I_j^{i_j}, \cdots, I_{j+l}^{i_{j+l}}$. Straightforwardly, the feature vector of P_i is a $2 \times l$ matrix where each column represents the lower and upper limits of each interval from the lower threshold j to the upper threshold $j + l$. The representation in this multi-dimensional feature space is used to characterize different types of patterns.

Dissimilarity function. A dissimilarity function between patterns is defined in order to measure their dissimilarity. This is fundamental in the case of Classification.

4 Experiments and Results

The *FOC* has been tested on mono-dimensional signals provided by microarray data of *Saccharomyces cerevisiae* DNA [9]. Here, the goal is the identification of nucleosomes and linker regions. In particular, each spot of the microarray represents a *probe* i of resolution $r = 50$ base pairs overlapping $o = 20$ base pairs with probe $i + 1$. The microarray follows the tiling approach, where the chromosome is spanned by moving a window (probe) i of width r base pairs from left to right, measuring both the percentage of mononucleosomal DNA G_i (*green channel*) and whole genomic DNA R_i (*red channel*) within such window, respecting also that two consecutive windows (probes) have an overlap of o base pairs. The resulting signal $V(i)$ for each probe i is the logarithmic ratio of the *green channel* G_i to *red channel* R_i. Intuitively, nucleosomes presence is related to peaks of V which correspond to higher logarithmic ratio values, while lower ratio values shows nucleosome free regions called *linker regions*.

4.1 *MLM* Rules on Biological Data

The key steps of the *MLM* are the *Training set and model construction*, the *Interval merging, pattern definition and selection*, which have to be defined in a personalized way depending on the property of the pattern to retrieve. In

the following, these steps are defined in the case of the mono-dimensional signal showing nucleosome positioning information. Note that, in such case, the *Preprocessing* phase consists in a convolution by an averaging kernel window $w = [\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$.

Training set construction. Since we know that WPN are shown as peaks of a bell shaped curve, in order to locate the position of a nucleosome, all local maxima of the input signal are automatically extracted from the convolved signal $\mathbf{X}$ of $\mathbf{S}$. Each convolved fragment X_t is processed in order to find $L(X_t)$ local maxima $M_t^{(l)}$ for $l = 1, \cdots, L(X_t)$. The extraction of each sub-fragment for each $M_t^{(l)}$ is performed by assigning all values in a window of radius os centered in $M_t^{(l)}$ to a vector, F_t^l of size $2 \times os + 1$: $F_t^l(j) = X_t(M_t^{(l)} - os + j - 1)$, for $j = 1, 2, ..., 2 \times os + 1$. The selection of the *significant* sub-fragments - to be used in the model definition - is performed by satisfying the following rule:

$$\begin{cases} F_t^l(j+1) - F_t^l(j) > 0 & j = 1, \cdots, os \\ F_t^l(j+1) - F_t^l(j) < 0 & j = os+1, \cdots, 2 \times os \end{cases} \tag{2}$$

After the selection process $G(X_t)$ sub-fragments remain for each X_t. The training set of the *interesting pattern* is $T_U = \{F_t^l | 1 \leq t \leq T \ , \ 1 \leq l \leq G(X_t)\}$.

Interval merging and pattern definition. This step is performed by taking in account that bell shaped pattern must be extracted for the classification phase. Such kind of patterns are characterized by sequences of intervals $\left\{I_j^1, \cdots, I_{j+l}^n\right\}$ such that $I_j^i \supseteq I_{j+1}^{i+1}$; more formally a pattern P_i is defined as:

$$P_i = \{I_j^{i_j}, I_{j+1}^{i_{j+1}}, \cdots, I_{j+l}^{i_{j+l}} \mid \forall I_h^{i_h} \ \exists! I \in R_{h-1} : I = I_{h-1}^{i_{h-1}} \supseteq I_h^{i_h}\}$$

where, j defines the threshold, t_j, of the widest interval of the pattern. From the previous definition it follows that P_i is build by adding an interval $I_{h+1}^{i_{h+1}}$ only if it is the unique in R_{h+1} that includes $I_h^{i_h}$. Note that, this criterion is inspired by the consideration that a nucleosome is identified by bell shaped fragment of the signal, and the intersection of such fragment with horizontal threshold lines results on a sequence of nested intervals.

Pattern selection. In this step the *interesting patterns* $\mathbf{P^{(m)}}$ are selected following the criterium:

$$\mathbf{P^{(m)}} = \{P_i \ : \ |P_i| > m\}$$

i.e. patterns containing intervals that persists at least for m increasing thresholds. This further selection criterion is related to the height of the shaped bell fragment, in fact a small value of m could represents noise rather than nucleosomes. The value m is said the *minimum number of permanence*.

Dissimilarity function. A dissimilarity function between patterns is defined in order to characterize their shape:

$$\delta(P_r, P_s) = \sum_{h=1}^{H} (a_h^{r_h} - a_h^{s_h}) \tag{3}$$

where $a_h^{r_h} = e_h^{r_h} - b_h^{r_h}$, $a_h^{s_h} = e_h^{s_h} - b_h^{s_h}$.

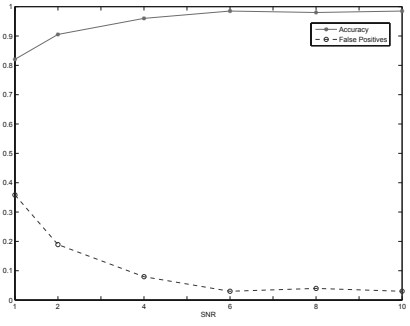


Fig. 2. Best Accuracy and FPR values versus SNR

In particular, this dissimilarity is a distance that takes into account the *shape* of a pattern. An example of signal and the relative interesting patterns is given in Figure 1b.

4.2 Results

The following experiments have been carried out by measuring the correspondence between Nucleosome and linker regions. In the case of the synthetic signal, the output of the classifier has been compared with a binary mask M revealing the real nucleosomal (linker) regions RNR (RLR) as contiguous sequence of 1's (0's)., in the case of the real data set the mask M is built by the output of the Hidden Markov model (HMM) used in the paper of Yuan et Al [9] optimally converted into a binary string, and considered as truth.

Moreover, by biological consideration, the radius os has been set to $os = 4$. The performances have been evaluated in terms of *Recognition Accuracy*, RA: a nulceosomal (linker) region CNR (CLR) is classified correctly if there is a match of at least $l = 0.7 \times L$ contiguous 1's (0's) between CNR (CLR) and the corresponding RNR (RLR) in M where L is the length RNR (RLR). The value 0.7 has been chosen because it represents a 70% of regions overlap very unlikely to be due to chance. Moreover, the parameters b,a,α of the FOC has been set to $b = E(\Delta)$,a such that $f_{\Delta}(a) = 10^{-}5$ and $\alpha = 0,5$. The choice of a and b is motivated by the fact that the distribution f_{Delta} has resulted very close to a normal distribution, thus b represent the mean value, and a is the $10^{-}5$ critical value. The value of $\alpha = 0,5$ has been decided because the condition $\mu_0(x, a, b) > \alpha$ corresponds to the defuzzification by the highest membership since $\chi_1 = 1 - \chi_0$ (see eq. 1).

4.3 Results on Synthetic Data

Such experiments allows to test the robustness of the FOC to signal noise. A procedure to generate synthetic signal has been recently developed allowing us to assess the feasibility of our method on controlled data. For details about the procedure and the setting parameters see [8,14]. All parameters used in

the generation of synthetic data have been inspired by biological considerations and are $nn = 200$, $nl = 250$, $\lambda = 200$, $r = 50$, $o = 20$, $nr = 100$, $dp = 0$, $dr = 0$, $pur = 0.8$, $nsv = 0.01$, $SNR = \{1, 2, 4, 6, 8, 10\}$ and $ra = 4$, resulting in 6 synthetic signals at different SNR. The training set T_U is represented by all WPN's that fit better the conditions in Eq. 2 with $os = 4$, because, by biological consideration, we know that a nucleosome is around 150 base pairs which corresponds to 8 probes. Thus, the training set T_U and consequently its size TL, are automatically selected by the MLM depending on the generated input signal, resulting that, for the specific experiments reported here, $TL = \{63, 98, 127, 142, 145, 147\}$ for $SNR = \{1, 2, 4, 6, 8, 10\}$ respectively. The optimal parameters for the MLM are derived by a calibration phase described in [13] and have resulted $H = 40$ and $m = 6$. The performances have been evaluated measuring the correspondence between the classified WPN or LN regions and the ones imposed in the generated signal. Fig. 2 reports the best *Accuracy* and FPR values versus SNR. From this study, it results that the average accuracy and FPR over the 6 experiments is 94% and 12% respectively.

4.4 Results on Real Data

In this experiment, we have compared the accordance of the FOC with the Hidden Markov model (HMM) used by Yuan et Al [9] on the *Saccharomyces cerevisiae* real data. The training set T_U has been decided as described above, moreover, we have chosen $H = 40$, $m = 6$ by the calibration phase ($m = 0.15 \times 40$). and, by biological consideration, the radius os has been set to $os = 4$. The confusion matrices which show the RA of HMM considering FOC as the truth classification and RA of FOC considering HMM as the truth classification are reported in Table 1. The results can be summarized in an overall RA of (0.82) for the HMM (FOC true) and 0.69 for FOC (HMM true). From this studies we can conclude that FOC does not fully agree with HMM on the nucleosome patterns. In particular, the FOC recognizes most of the nucleosomes classified by HMM, and other patterns that could be nucleosomes too. Such considerations indicate that the integration of the two methods could improve the overall classification.

Table 1. Agreement between the HMM and FOC (and viceversa) on the Saccharomyces cerevisiae data set for Nucleosomes (N) and Linker (L) regions. The table on the left shows the RA results of HMM when considering FOC as the truth classification, while the opposite is shown on the right table.

	F	O	C
H		L	N
M	L	0.77	0.23
M	N	0.13	0.87

	H	M	M
F		L	N
O	L	0.54	0.46
C	N	0.15	0.85

5 Final Remarks

In this paper we have shown that a fuzzy One class classifier, by using the preprocessing of the Multi Layer method, is able to distinguish between nucleosome

and linker patterns in the particular problem of the nucleosome positioning. We have performed our experiment on a real and synthetic data-set, comparing, in the real case the result of the new classifier with an HMM classifier recently used for the same purpose. In the future we aim to properly integrate the two methods of classification (HMM and MLM) in the same biological problem in order to increase the overall classification performances.

References

1. Tax, D.M.J.: One-class classification, Ph.D. thesis (June 2001)
2. Manevitz, L., Yousef, M.: One-Class SVMs for Document Classification. Journal of Machine Learning Research 2, 139–154 (2001)
3. Wang, K., Stolfo, S.J.: One-Class Training for Masquerade Detection. In: Workshop on Data Mining for Computer Security, Melbourne, Florida, November 19-22, 2003, pp. 10–19 (2003)
4. Di Gesù, V., Lo Bosco, G.: Combining one class fuzzy kNN's. In: Masulli, F., Mitra, S., Pasi, G. (eds.) WILF 2007. LNCS, vol. 4578, pp. 152–160. Springer, Heidelberg (2007)
5. Di Gesú, V., Lo Bosco, G., Friedman, J.H.: Intruder Pattern Identification. In: Proc. of ICPR 2008 19th International Conference on Pattern Recognition, Tampa, Florida, USA, December 8-11, 2008, pp. 1–4. IEEE Computer Society Press, Los Alamitos (2008)
6. Zeng, Z., Fu, Y., Roisman, G.I., Wen, Z., Hu, Y., Huang, T.S.: One-Class Classification for Spontaneous Facial Expression Analysis. In: IEEE Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, pp. 281–286 (2006)
7. Nanni, L.: Experimental comparison of one-class classifiers for online signature verification. Neurocomputing 69(7-9), 869–873 (2006)
8. Di Gesù, V., Lo Bosco, G., Pinello, L.: A one class classifier for Signal identification: a biological case study. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part III. LNCS, vol. 5179, pp. 747–754. Springer, Heidelberg (2008)
9. Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., Rando, O.J.: Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. Science 309, 626–630 (2005)
10. Corona, D.F.V., Tamkun, J.W.: Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochim Biophys Acta 1677(1-3), 113–119 (2004)
11. Saha, A., Wittmeyer, J., Cairns, B.R.: Chromatin remodelling: The industrial revolution of DNA around histones. Nature Reviews Molecular Cell Biology 7, 437–447 (2006)
12. Jacobson, S., Pillus, L.: Modifying chromatin and concepts of cancer. Current Opinion in Genetics and Development 9, 175–184 (1999)
13. Corona, D.F.V., Di Gesú, V., Lo Bosco, G., Pinello, L., Yuan, G.C.: A new Multi-Layers Method to Analyze Gene Expression. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part III. LNCS, vol. 4694, pp. 862–869. Springer, Heidelberg (2007)
14. Di Gesù, V., Lo Bosco, G., Pinello, L., Yuan, G.C., Corona, D.F.V.: A Multi-Layer Method to Study Genome-Scale Positions of Nucleosomes. Genomics 93(2), 140–145 (2009)

An Experimental Validation of Some Indexes of Fuzzy Clustering Similarity

Stefano Rovetta[1,2] and Francesco Masulli[1,2,3]

[1] Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova,
[2] CNISM Genova Research Unit, Via Dodecaneso 35, I-16146 Genova, Italy
[3] Center for Biotechnology, Temple University, Philadelphia, USA
{rovetta,masulli}@disi.unige.it

Abstract. Measuring the similarity between clusterings is a classic problem with several proposed solutions. In this work we focus on measures based on co-association of data pairs and perform some experiments to investigate whether specificities can be highlighted in their behaviour. A unified formalism is used, which allows easy generalization of several indexes to a fuzzy setting. A selection of indexes is presented, and experiments investigate simplified cases and a paradigmatic real-world case, as an illustration of application.

1 Introduction

Fuzzy clustering [1] is a well established procedure and a useful data analysis tool, often providing more flexibility and expressive power than crisp techniques. In general, clustering quality assessment is a problem without a satisfactory solution [2] due to the unsupervised nature of the task. Several cluster validity indexes [3,4] take into account cluster size and composition. However, if we have additionally class labels available, comparisons may be performed according to this external information, although the fact that these reflect the natural grouping of the data is only to be regarded as a working hypothesis. Another form of quality assessment for clusters is stability analysis [5], where comparison is necessary to evaluate cluster variability.

Cluster similarity (or diversity) can also be used to achieve better performance by ensemble clustering [6]. Another use of cluster comparison is biclustering [7] i.e., clustering rows and columns at the same time. Fuzzy biclustering [8] is available. In both crisp and fuzzy cases, many methods only produce one bicluster for each run, so the question arises of whether two biclusters are similar enough to be considered the same.

In all these examples, the problem may be reduced to measuring the similarity between two fuzzy partitions. In this paper, we analyze some methods to compare fuzzy partitions. The contributions of this work consist in:

- A unified formalism useful for the implementation of several clustering comparison indexes, which are all represented by means of co-association matrices;
- The use of this framework to generalize several indexes to a fuzzy setting; some indexes are presented, but the generalization can be extended to many other cases;
- Some experimental insight on the behaviour of these indexes in both simplified cases, where the relationship between the partitions is clear, and a paradigmatic real-world case, as an illustration of application.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 132–139, 2009.

2 Measures Based on Data Pairs

Clustering induces a partitions of the data, so measuring the agreement or overlap between two clusterings amounts to measuring the similarity between two partitions.

There are several partition similarities available in the literature. A first distinction can be made between *pairwise* similarity indexes, which apply to pairs of partitions, and *non-pairwise* ones, which can work on an arbitrary number of partitions. The difference is of a practical nature only, since in principle a non-pairwise index can always be obtained by computing a pairwise index for all possible pairs, and then averaging.

A more fundamental distinction refers to the way partitions are compared. The two main approaches include comparing matching subsets, and comparing co-association information. The first approach is not reliable when the partitions are not very similar, and in any case require some criterion for matching subsets. In principle, we can expect these methods to be of linear complexity w.r.t. data cardinality.

In this study, we concentrate on the second approach. Co-association information is obtained by analyzing whether pairs of points in the data set are co-attributed to the same cluster by both partitions. These methods require building a co-association matrix [9], by scanning all possible data pairs, so they run quadratically with cardinality.

Given a data set X, suppose we have two fuzzy or soft partitions A and B of X. Soft partitions means that $\forall x \in X$ there is a membership $\mu(x, a_i)$ for each subset $a_i \in A$ (similarly for B – and this comment applies throughout). We assume normal memberships. For a proper partition we ask that $\sum_i \mu(x, a_i) = 1$. (Note that possibilistic subsets –not proper partitions– are also possible by removing this constraint.)

Each data point is thus represented by a coordinate vector, whose dimension is the number of subsets (clusters) in the partition, and whose components are the membership values, which we assume normalized in $[0, 1]$. Each data pair is described by the degree of similarity between the two objects x and y under the partition A.

Similarity of strings of memberships can be measured by Hamming distance for crisp bits, which is equivalent to summing the bits of the bitwise-AND between the two words. The fuzzy generalization of this operation is defined once we appropriately define the conjunction connective AND [10]. We adopt the product t-norm [11], which is appealing because it is related to the scalar product operation between vectors, which in turn can be used to define popular distance measures, and also to the concept of a joint probability. This allows some generality in the indexes studied, although the product logic arising from this particular choice does not have some of the properties found in other cases [12] (e.g., Gödel or Łukasiewicz fuzzy logics). In particular, some of the derivations have been obtained by assuming a specific relationship between the AND and OR connectives which is not necessarily satisfied by all possible definitions.

Given two fuzzy memberships/truth values μ and ν, the conjunction logical connective is defined as $\mu \,\mathrm{AND}\, \nu = \mu\nu$. The co-association of a given pair of data points to a given cluster a_i is the conjunction of the respective point memberships to a_i, and the degree of similarity of two points is the average of these values. The *co-association matrix* s^A under partition A is:

$$s^A_{ij} = s^A(x_i, x_j) = \sum_{l=1}^{|A|} \mu(x_i, a_l)\mu(x_j, a_l) \tag{1}$$

Note that in the crisp case this definition collapses to the proposition "partition A puts x and y in the same cluster", but in the fuzzy case it is necessary to take all clusters into considerations because, in general, none of them will be exactly zero or one.

3 Indexes

Once the co-association matrices s^A and s^B are built, we can treat their entries as two paired samples and compare them with appropriate measures.

To simplify notations, we will serialize the matrices s^A and s^B so that they may be indexed as vectors, so that: $s^A_{ij} = \sigma^A_h$. To avoid redundancy, the index h scans only the upper triangular matrix, excluding the diagonal (which is trivial), so $i \in [1, |X| - 1]$, $j \in [i + 1, |X|]$, and $h = |X|(i - 1) - i(i + 1)/2 + j$. Moreover, we define $H = |X|(|X| - 1)/2$ so that $h \in [1, H]$.

Indexes of partition similarity based on the co-association matrix can be computed by several approaches. Some of them are reviewed in [13] and some are experimentally compared in [6]. These may include Ward's linkage criterion [14], Student's t formula [15], information-theoretic criteria as in [6] and [13], and well-known partition overlap measures like Jaccard's [16] or Rand's [17] or Fowlkes and Mallows' [18] indexes, and subsequent work [19].

In general, measures for paired samples can be based on different criteria. Methods like Ward or Student are focused on comparing average and dispersion of the two samples. Another approach is to compute some statistic on the differences between the two data in each pair (that is, $\sigma^A_h - \sigma^B_h \ \forall h$). Yet another criterion is to exploit the $[0, 1]$ range of the values and analyze other types of combination between data, as in the Pearson correlation and the Jaccard index. We will focus on some indexes which are representative of each approach.

Average linkage – This measure of distance is based on comparing the centroids of two sets, i.e., on computing the average of the two sets and measuring their distance.

This criterion is simple-minded, since it is prone to false positives: two sets with the same centroid are considered coincident even if one has a larger variance than the other. However, for normalized data, it may be reasonable.

Correlation – The standard Pearson correlation coefficient, a measure of similarity:

$$\mathrm{corr} = \frac{\frac{1}{H}\sum_h(\sigma^A\sigma^B) - \frac{1}{H}\sum_h(\sigma^A)\frac{1}{H}\sum_h(\sigma^B)}{\sqrt{\frac{1}{H}\sum_h((\sigma^A)^2) - \left(\frac{1}{H}\sum_h(\sigma^A)\right)^2}\ \sqrt{\frac{1}{H}\sum_h((\sigma^B)^2) - \left(\frac{1}{H}\sum_h(\sigma^B)\right)^2}} \quad (2)$$

Since this is a similarity measure between -1 and 1, the correlation distance is

$$C(A, B) = (1 - corr)/2. \quad (3)$$

Jaccard – The Jaccard coefficient [16] is a classic measure of set similarity, and one of the most general. It is the ratio of the intersection of two sets to their union: $J(A, B) = |A \cap B|/|A \cup B|$. In the crisp case, this pairwise index can be practically computed by counting the number $N_{11} = |A \cap B|$ of points put in the same cluster by both partitions,

the number N_{10} of points assigned to the same cluster only by partition A, and the number N_{01} similarly defined, so that $N_{10} + N_{01} + N_{11} = |A \cup B|$ and

$$J(A, B) = \frac{N_{11}}{N_{10} + N_{01} + N_{11}}. \tag{4}$$

In the fuzzy case, the concept of coincidence must be redefined as a degree of coincidence. By taking advantage of the reasonable assumption that a generalization of De Morgan's law holds, for the product t-norm we can define the associated disjunction operator as the *probabilistic sum* t-conorm, so that $\mu \,\mathrm{OR}\, \nu = \mu + \nu - \mu\nu$. Therefore, in terms of σ^A and σ^B, the actual computation in this case is:

$$A \cap B = \sum_h \sigma_h^A \sigma_h^B \qquad \text{and} \qquad A \cup B = \sum_h \left(\sigma_h^A + \sigma_h^B - \sigma_h^A \sigma_h^B\right) \tag{5}$$

The Jaccard distance between A and B is $1 - J(A, B)$.

Student distance – This index exploits the well-known Student's t statistic to obtain a distance measurement which takes into account not only the estimated overlap, but also its significance in terms of variance.

$$S(A, B) = \frac{\left(\sum_h |\sigma_h^A - \sigma_h^B|\right)}{0.5 + \sum_h |\sigma_h^A - \sigma_h^B|^2 / H - (\sum_h |\sigma^A - \sigma^B|/H)^2} \tag{6}$$

With respect to the Student's t formula, this is compensated to avoid a vanishing denominator.

The Rand index – This index was explicitly proposed for comparing clusterings [17], but it is computed similarly to the Jaccard index which was introduced in a more generic setting. It is defined as

$$R(A, B) = \frac{\sum_h (\sigma_h^A + \sigma_h^B) + 1}{H} \tag{7}$$

4 Experiments

We present some experiments aimed at highlighting the behaviors of these indexes. We use three kinds of data: two families of simple datasets, represented through the membership (directly as partitions, regardless of how they were obtained); then we use the Iris data set as an illustration of real-world application.

4.1 Experiments with Toy Problems

Experiment set 1 – The aim of this experiment set is to compare several synthetic partitions. The family of toy datasets used is composed of 20 data objects in 2 partitions, one of 2 clusters, the other of 3 clusters. The presented methods are insensitive to the number of clusters in partitions, although this sensitivity may be easily introduced. The datasets used are tabulated in Table 1. Table 2 presents the results.

Experiment set 2 – The second toy dataset family is composed of 20 data objects; 2 partitions of 2 clusters each are used. The second partition is fixed as follows:

Table 1. The family of datasets used in the first experiment set

toy0																		
A	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
B	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
toy1																		
A	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
B	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0
toy2																		
A	0.8	0.8	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
B	0.8	0.8	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
toy3																		
A	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
B	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.8	0.8	0.8	0.8	0.8	0.2	0.2	0.2	0.2	0.2	0.2
	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2	0.2	0.8	0.8	0.8	0.8	0.8	0.8
toy4																		
A	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
B	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0
toy5																		
A	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0	1.0	0.9	0.0	0.0	0.0	0.0	0.0
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	0.0	0.1	1.0	1.0	1.0	1.0	1.0
B	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0

object	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
cluster1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
cluster2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

In the first experiment of this set we investigate the variation of indexes in a crisp case. The data start equally clustered in both cases (two equal partitions). Then, in 10 further steps, each data object is moved from one cluster to the other in the first partition; the second is left as is. In the second experiment, the variation of indexes in a simple fuzzy case is analyzed. The partitions start identical, with data equidistributed in the two clusters. One data object in the first partition is gradually moved from the first cluster to the second by changing its memberships in steps of 0.1, as follows:

step	0	1	2	3	4	5	6	7	8	9	10
cluster1	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
cluster2	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

The third experiment involves changing the memberships of all data in a partitions from crisp to totally fuzzy, i.e., in the first partition all memberships of the first 10

Table 2. Results on the first toy problem set

	Average	Rand	Student	Correl.	Jaccard
toy0	0.0000	0.0000	0.0000	0.0000	0.0000
toy1	0.8497	0.4248	0.5707	0.8211	0.6566
toy2	0.0000	0.4352	0.0000	0.0000	0.5976
toy3	0.4455	0.2227	0.3845	0.1610	0.4207
toy4	1.0980	0.5490	0.7344	1.0811	0.8000
toy5	0.9707	0.4854	0.8205	0.9489	0.7595

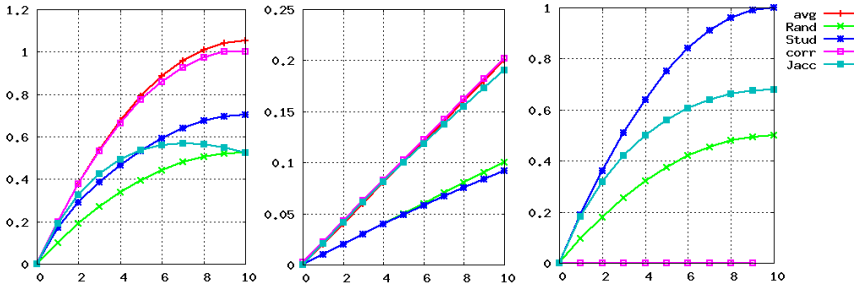


Fig. 1. Results on the second experiment set

points in the first cluster are gradually moved from 1 to 0.5 and all memberships of points 11-20 are moved from 0 to 0.5; memberships in the second cluster are obviously complementary. The second partition is held fixed.

The results of these experiments are illustrated by the graphs in Figure 1. We can observe that in general all the indexes considered feature similar behavior, having value 0 for equal partitions and monotonically increasing behavior (they are not normalized on a single scale, however). The graph also show that, while this similarity holds for most types of variation, there are cases where some index does not agree with others (see Jaccard on the first experiment).

4.2 Real Data: Iris

Anderson's Iris data [20] is an almost mandatory testbed, since it is so well-known. Here we use it to show how the properties of the indexes may be exploited. For this dataset the following two features are used: sepal width $\times$ sepal length; petal width $\times$ petal length. These allow a low error with linear separation. In this experiment, two partitions are compared: the ground truth provided by the true classes and the central clustering obtained by taking the averages of each class as centroids. Therefore there are 3 clusters in both the target and class-induced clustering.

Experiments have been made with both crisp and fuzzy partitions. In the crisp case, the memberships are obtained with the simple nearest centroid rule. In the fuzzy case, a membership model similar to the "Maximum Entropy" approach [21] has been used, where membership of data point x_i is computed as $\mu_j(x_i) = \frac{e^{-d_{ij}/\beta}}{\sum_l e^{-d_{Ila}/\beta}}$, where d_{ij} is

Table 3. Results on the Iris dataset

Crisp case	**Average**	**Rand**	**Student**	**Correl.**	**Jaccard**
Complete	0.5589	0.2795	0.3985	0.5962	0.5552
Setosa vs Versicolor	0.3790	0.1895	0.2899	0.4492	0.5043
Setosa vs Virginica	1.1055	0.5527	0.7397	1.1562	0.8761
Virginica vs Versicolor	0.1523	0.0762	0.1335	0.2175	0.2902
Fuzzy case	**Average**	**Rand**	**Student**	**Correl.**	**Jaccard**
Complete	0.8010	0.4005	0.5639	0.5506	0.6311
Setosa vs Versicolor	0.2916	0.1458	0.2828	0.0763	0.4260
Setosa vs Virginica	0.9465	0.4733	0.8101	0.8406	0.7576
Virginica vs Versicolor	0.1624	0.0812	0.1608	0.0127	0.2939

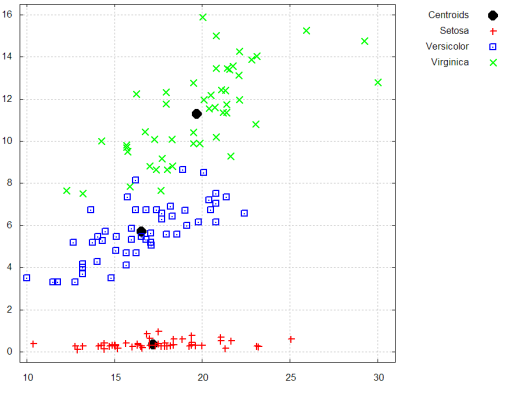


Fig. 2. The Iris data, with class centroids

the (Euclidean) distance between data point x_i and the j-th centroid y_j and β is a scale parameter imposing the degree of fuzziness, here set to 1.

Table 3 presents the results. In both the crisp and the fuzzy case, the first row refers to the whole dataset (three classes), while the remaining rows represent all possible pairing of the three classes. We notice that representation with three centroids is not very good, even if these are chosen as the mean of the respective classes. This can be easily inferred from Fig. 2, showing centroids and data. The results indicate that, although with different numerical values, all the indexes correctly reflect the well-known fact that the separation between the *Versicolor* and *Virginica* varieties is hard.

5 Conclusion

We introduced a possible fuzzy framework for applying traditional and novel partition similarity measures to fuzzy clustering. Some indexes have already been generalized to the fuzzy setting in the literature [22], but we provide a more systematic procedure while also introducing some new indexes. Several other formulations are of course possible in addition to the three proposed here; we only provide these as examples.

The experimental result show that, while all indexes retain the same general behaviour, there are indeed some occasional differences that may deserve to be studied and exploited. Further research will focus on characterizing these fuzzy indexes and on their application in tasks involving clustering comparison, e.g., ensemble clustering.

References

1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell (1981)
2. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)
3. Bezdek, J.C.: Cluster validity with fuzzy sets. Cybernetics and Systems 3(3), 58–73 (1973)
4. Wang, W., Zhang, Y.: On fuzzy cluster validity indices. Fuzzy Sets Syst. 158(19), 2095–2117 (2007)
5. Filippone, M., Masulli, F., Rovetta, S., Zini, L.: Comparing fuzzy approaches to biclustering. In: Tagliaferri, R., Masulli, F. (eds.) CIBB 2008 proceedings (in press) (2008)
6. Kuncheva, L.I., Vetrov, D.P.: Evaluation of stability of k-means cluster ensembles with respect to random initialization. IEEE Trans. Pattern Anal. Mach. Intell. 28(11), 1798–1808 (2006)
7. Cheng, Y., Church, G.M.: Biclustering of expression data. In: Proc. Int. Conf. Intell. Syst. Mol. Biol., vol. 8, pp. 93–103 (2000)
8. Filippone, M., Masulli, F., Rovetta, S., Mitra, S., Banka, H.: Possibilistic approach to biclustering: An application to oligonucleotide microarray data analysis. In: Priami, C. (ed.) CMSB 2006. LNCS (LNBI), vol. 4210, pp. 312–322. Springer, Heidelberg (2006)
9. Fred, A.L.N., Jain, A.K.: Data clustering using evidence accumulation. In: International Conference on Pattern Recognition, vol. 4 (2002)
10. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)
11. Menger, K.: Statistical metrics. Proceedings of the National Academy of Sciences of the United States of America 28(12), 535–537 (1942)
12. Klement, E.: A survey on different triangular norm-based fuzzy logics. Fuzzy Sets and Systems 101(2), 241–251 (1999)
13. Meilă, M.: Comparing clusterings–an information based distance. Journal of Multivariate Analysis 98(5), 873–895 (2007)
14. Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58, 236–244 (1963)
15. Student: The probable error of a mean. Biometrika 6(1), 1–25 (March 1908)
16. Jaccard, P.: Étude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 547–579 (1901)
17. Rand, W.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66, 846–850 (1971)
18. Fowlkes, E.B., Mallows, C.L.: A method to compare two hierarchical clusterings. Journal of the American Statistical Association 78, 553–569 (1983)
19. Hubert, L., Arabie, P.: Comparing partitions. Journal of Classification, 193–218 (1985)
20. Anderson, E.: The irises of the gaspe peninsula. Bulletin of the American Iris Society 59, 25 (1935)
21. Rose, K., Gurewitz, E., Fox, G.: A deterministic annealing approach to clustering. Pattern Recognition Letters 11, 589–594 (1990)
22. Campello, R.J.G.B.: A fuzzy extension of the rand index and other related indexes for clustering and classification assessment. Pattern Recognition Letters 28(7), 833–841 (2007)

Combining Fuzzy C-Mean and Normalized Convolution for Cloud Detection in IR Images

Anna Anzalone[1], Francesco Isgrò[2], and Domenico Tegolo[3]

[1] INAF - Istituto di Astrofisica e Fisica Cosmica, Palermo, Italy
[2] Dipartimento di Scienze Fisiche, Università degli Studi di Napoli Federico II, Napoli, Italy
[3] Dipartimento di Matematica ed Applicazioni, Università di Palermo, Palermo, Italy
anna.anzalone@ifc.inaf.it, francesco.isgro@unina.it, domenico.tegolo@unipa.it

Abstract. An important task for the cloud monitoring in several frameworks is providing maps of the cloud coverage. In this paper we present a method to detect cloudy pixels for images taken from ground by an infra-red camera. The method is a three-steps algorithm mainly based on a Fuzzy C-Mean clustering, that works on a feature space derived from the original image and the output of the reconstructed image obtained via normalized convolution. Experiments, run on several infra-red images acquired under different conditions, show that the cloud maps returned are satisfactory.

Keywords: Cloudiness mask, fuzzy set, infra-red images.

1 Introduction

Automated cloud detection is a challenging issue, crucial for cloud monitoring. Clouds play an important role in several fields such as the global climate change, weather forecast, climate modeling. For example the global change of the Sea Surface Temperature (SST) due to the greenhouse effect, is a study where it is necessary filtering out cloudy pixels from the satellite data, in order to avoid cloud contamination in the measurements of the temperature. For this purpose auxiliary cloud masks are used that indicate wether the single pixel is affected or not by clouds. Cloud detection process in the images from space, should take into account that cloud appearance is similar to other entities and also changes depending on the region, hence the discrimination is not a straightforward task. It is difficult to discriminate ice or snow from clouds in the polar regions, because of their similar reflectiveness in the visible and small contrast in the infra-red wavelengths. Cloud and volcanic ashes in volcanic areas, cloud and fire, cloud and dust over the deserts etc. [1]. Moreover in both space and ground observations, it must be taken into account that edges in cloud images are generally smoothed and sharp outlines are hard to be detected. The recent literature proposes methods for clear sky determination based on the idea that brightness temperature of cloudy pixels show different relationships and properties from those expected for

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 140–147, 2009.

clear conditions. A set of combined tests are applied on images from space and generally infra-red and visible bands are used. In [2] a pixel is labeled as cloudy if all tests indicate the presence of cloud. Different approaches exploit features non related to the physics of the clouds but to the spatial/temporal relationships between pixels of the same or consecutive frames [3] or determined by machine learning processes using Bayesian classification [4], decision trees [1], or support vector machines [5]. It is important to take into account that validation of a cloud mask is a very difficult issue. A way to proceed for quality estimation is assessing the agreement with the human analysis, combined with comparisons against masks detected with different algorithms.

In this paper we present an algorithm for detecting clouds from infra-red images. The algorithm is based on the Fuzzy C-Mean clustering method, and it is structured in three steps: Fuzzy C-mean clustering on the whole image, Gaussian classification on the detected clusters, Fuzzy C-mean clustering only on one of the two final clusters. The features used for clustering are the original grey-level and the value of the same image reconstructed using the normalized convolution.

The paper is structured as follows. Next section describes the method, together with a short description of the techniques used. Section 3 discusses the assessment of the method on a data set of infra-red images including different kinds of clouds. Section 4 is left to the final remarks.

2 Description of the Method

The proposed method for cloud mask detection is mainly based on three steps: Fuzzy C-mean clustering on the whole image, Gaussian classification on the detected clusters, Fuzzy C-mean clustering only on one of the two final clusters. The first step is a coarse segmentation of the image in three clusters representing sky, cloud or other scene objects. These three clusters have been used to include the most representative entities of a general cloudy scene. But in our experiments the sky and the other non-cloud elements (sky in the following) were joined because this paper focusses the attention only on cloud/non-cloud segmentation. The second step analyzes the outliers of each clusters and reassigns them by means of the normal distribution. Finally the last step is applied on the sky cluster to identify the misclassified cloud pixels, typically the border ones.

The representation of information has a relevant role in the image analysis, some different features have been considered to identify the map of cloud coverage (Grey Level, Mean and STD, Gradient, Local Histogram, Entropy,...), and an exhaustive analysis has been done to identify the best set of features. In the end the grey level feature resulted good for a first rough segmentation of the IR image, but the final mask needed to be improved by means of the support of other different features. To the scope a transformation of the grey level image based on normalized convolution, has been adopted.

Normalized Convolution: The Normalized Convolution is a method for signal analysis that takes into account uncertainties in signal values and at the same

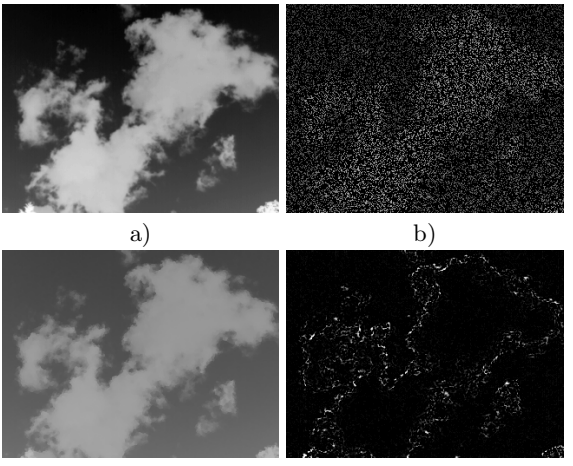


Fig. 1. Results of the normalized convolution algorithm on an IR cloud image. a) Original image, b) 10% Sampled image, c) Reconstructed image d) Difference between the original image and the reconstructed one.

time allows spatial localization of the possible unlimited functions by means their analysis.

The equation $U(\epsilon) = \sum_X a(x)B(x) \odot c(\epsilon - x)T(\epsilon - x)$ expresses the general formulation of convolution, where $\odot$, in the standard convolution, is the scalar multiplication. While the normalized convolution of aB and cT can be formulated as $U_N = \{aB\widehat{\odot}cT\}_N = N^{-1}D$ where D=$\{aB\widehat{\odot}cT\}$ and N=$\{aB \odot B * \widehat{\cdot}c\}$. Summation is an example to produce D and N over the corresponding indexes, and to give an example for a we can use the following function:

$$a = \begin{cases} r^{-\alpha} \cos^{\beta}(\frac{\pi r}{2r_{max}}) & r < r_{max} \\ 0 & \text{otherwise} \end{cases}$$

where r is the distance between the center and the nearest pixel, α and β are two positive integers. [6,7]. In figure 1 an example of the normalized convolution applied on an infra-red cloud image is shown.

The normalized convolution step extracts a sample of the pixels from the IR image (figure 1.b) and reconstructs the original image (figure 1.c). Dissimilarities among the input image and the reconstructed image result especially in the border pixels (figure 1.d). From the thermographic point of view, in these points there is a big variety of temperatures with large mutual exchanges and the uncertainty to assign these points to the sky or to the cloud sets, grows. The normalized convolution is a good support for the next step that tries to reassign those points. Therefore the original and the reconstructed grey levels are the only two characteristics considered by the next step to cluster the clouds. Note that the normalized convolution grey level is used as feature just to improve the border pixel classification.

A Fuzzy C-mean algorithm for three clusters has been chosen for the first step of the segmentation.

Fuzzy C-Mean: Standard segmentation is based on the use of attributes or features to distinguish different objects inside an image. Such characteristics are extracted during the low level vision phase to characterize afterwards objects in the high level vision step. In image data acquisition, all pixels of each frame are acquired in synchronous way. Let is k the number of observations of different elements, they can be grouped in N-dimensional vectors $z_k = [z_{1k}, z_{2k}, ...z_{nk}]^T, z_k R^n$ We can define a set of N observations as follows:

$$Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{21} & z_{22} & \cdots & z_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nN} \end{pmatrix} \tag{1}$$

In the case of dynamic systems the matrix Z can contain sample of signals or different scale of signal. In our case two columns have been considered at different level of signal. Aim of the clustering is to segment the data in different classes. K-means clustering, is one of the simplest unsupervised classification algorithm, it can be arranged to segment objects that appear in the images [8]. Defined the number of clusters, the procedure follows a simple way to classify the data-set. The algorithm partitions a set of N vectors $X = \{x_j, j = 1..N\}$ into C classes $c_i, i = 1, .., C$. It finds a starting cluster centre for each class, named cluster centroid, then an objective function of dissimilarity has to be minimized [9]. If an Euclidean function is considered, the function $P = \sum_{i=1}^{c} \left(\sum_{k, x_k \in c_i} \parallel x_k - v_i \parallel^2 \right)$ must be minimized where v_i is the centroid of the cluster v_i. A binary matrix $U = (u_{ij})$ defines a membership matrix as

$$u_{ij} = \begin{cases} 1 & \text{if } \parallel x_j - v_i \parallel^2 \leq \parallel x_j - v_k \parallel^2, \forall k \neq i \\ 0 & \text{otherwise} \end{cases}$$

where $v_i = \frac{\sum_{x_j \in v_i} \sum_{j=1}^{N} X_j}{\|c_i\|}$. Outcome of this step is a rough segmentation of the image according to the selected features and the defined clusters. The next two steps refine the retrieved clustering improving the classification of the outliers and resuming the likely misclassified cloudy pixels. In fact the Gaussian classification step identifies some few outliers present in each cluster and switch them to the correct cluster. Then the Fuzzy C-mean algorithm, assuming that the cloud cluster is correct, extracts from the sky cluster some points that could be included in the other one.

3 Experiments

The method described in the previous section can be summarized as follows:

Norm_Conv: Computes the Normalized Convolution for a 10% pixels of the original image;

Fuzzy_Im: Runs the Fuzzy C-Mean algorithm on the whole image to segment sky and cloud areas;

Gauss_Im: Applies a Gaussian function to include outlier pixels to the correct cluster;

Fuzzy_Sky: Runs Fuzzy C-Mean algorithm only on the sky cluster, to move wrongly classified pixels into the cloud cluster.

The step **Norm_Conv** requires some parameters where windows, on which to apply the Gaussian function for the convolution, and radius, are the most important ones. Their values have been set to 2 and 4 based on experimental tests. Moreover the method works on 10% uniform random selection of pixels of the input image. The step **Fuzzy_Im** segments the image in sky and cloud pixels, according to a set of parameters: maximum number of iterations, minimum amount of improvement, and number of clusters present in the data set. We have defined 14 as maximum number of iterations, e^{-5} as minimum amount of improvement and 3 as cluster number. Cloudy and non cloudy pixels characterize the step **Gauss_Im**. Mean and standard deviation are calculated for each cluster, then two gaussian probability values are evaluated for each pixel. Their maximum value allows to reassign the pixel to another cluster leaving the outliers moving from a cluster to another one. In the step **Fuzzy_Sky** has been applied a Fuzzy C-Mean on the non cloudy cluster with the cluster parameter equal to 2 and the other options defined equal to the ones in the step Fuzzy_Im.

3.1 Data

In order to test the algorithm efficiency, we selected some different infra-red cloud images and figure 2 shows some of them. Cloud images acquired in the infra-red

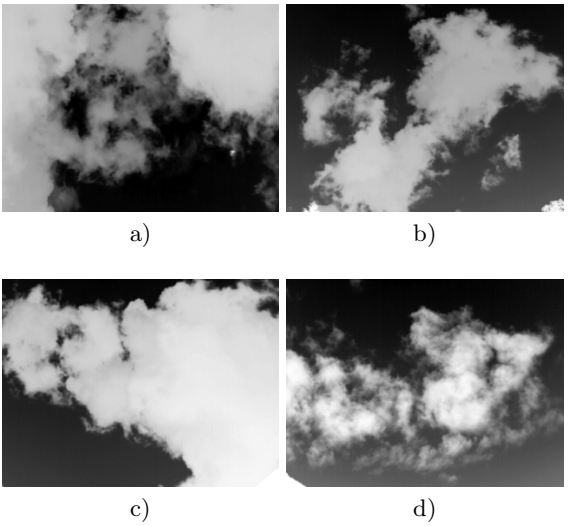


Fig. 2. Some examples of infrared cloud images

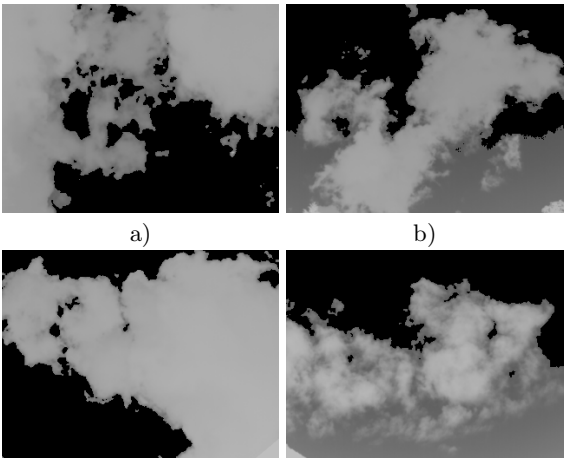


Fig. 3. Results after the first step

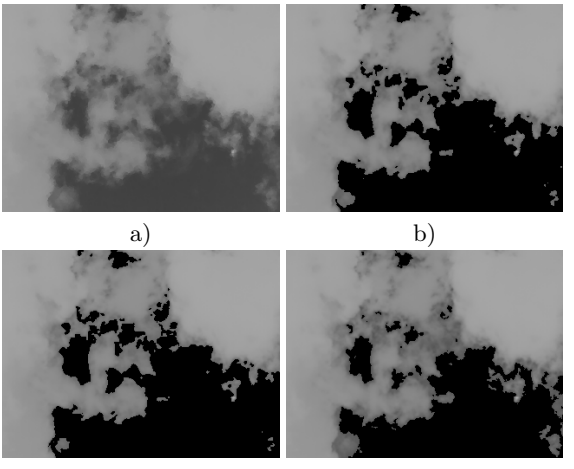


Fig. 4. Intermediate and final results for the image in figure 2 a). a) Normalized convolution step, b) First Fuzzy C-mean on whole image, c) Gaussian step; d) Second Fuzzy C-mean step on the sky cluster.

spectrum were selected from the archives of the Department of Mathematics and Application (DMA), University of Palermo, they include different kind of clouds such as cirrus, strato-cumulus, cumulonimbus,etc. The image sequences were acquired by the FLIR S-65 thermal-camera with a 18mm lens, spectral range 7.5-13μm and 320×240 pixels. The sequences have been processed by a FLIR software to extract single frames. At present our database includes a

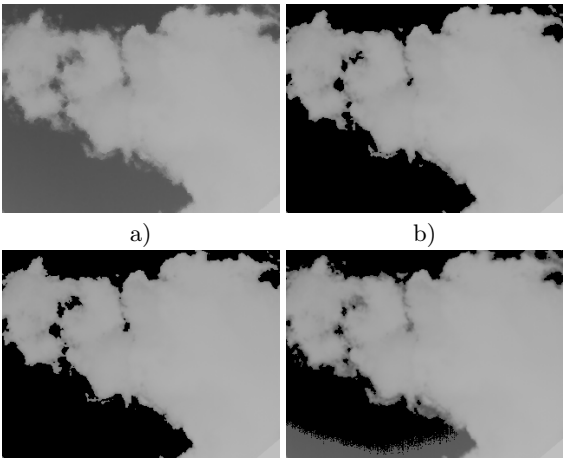


Fig. 5. Intermediate and final results for the image in figure 2 c). a) Normalized convolution step, b) First Fuzzy C-mean step on whole image, c) Gaussian step; d) Second Fuzzy C-mean step on the sky cluster.

large amount of sequences and many frames have been analyzed to evaluate the goodness of the method.

3.2 Results

Figure 3 depicts the images proposed in figure 2 after the first step of the method. Although the result of this step gives a reasonable classification of the pixels, we can observe that some pixels of cloud are lost and some other pixels are wrongly assigned as cloud, around the edges of the image. The background or sky pixels will be eliminated from the cloud cluster by the next step.

Figures 4 and 5 show the output of each step of the method for two portions of sky present in figure 2 a) and c) : sub-image a) shows the first step (Fuzzy C-mean algorithm), sub-images b) and c) illustrate the middle steps and finally the sub-image d) displays the cloudiness mask.

4 Conclusion

In this paper we tackled the problem of cloud segmentation from infra-red images, proposing a method that combines the normalized convolution with the Fuzzy C-mean clustering algorithm. Cloudiness masks were computed on infrared images acquired from ground including mainly two entities: sky and clouds. Furthermore the method was assessed considering several kind of clouds with different configurations. The quality of the results is satisfactory, although we could also give a qualitative validation. In the near future we plan to validate the

algorithm on different kind of images from different sensors and also on images from space where different entities should be taken into account such as land, snow, sea etc. Finally it needs to be tested against other methods.

References

1. Shiffman, S., Nemani, R.: Evaluation of decision trees for cloud detection from AVHRR data. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 5610–5613. IEEE Press, Los Alamitos (2005)
2. Ackerman, S., et al.: Discriminating clear sky from clouds with MODIS. J. Geophysical Research. 103, 141–157 (1998)
3. Seiz, G., Baltisavias, E.P., Gruen, A.: Cloud Mapping from the Ground:Use of Photogrammetric Methods. J. Photogrammetric Engineering and Remote Sensing, 941–951 (2002)
4. Murtagh, F., Barreto, D., Marcello, J.: Decision Boundaries Using Bayes Factors:The Case of Cloud Masks. J. IEEE Trans. on Geoscience and Remote Sensing 14, 2052–2958 (2003)
5. Lee, Y., Wahba, G., Ackerman, S.A.: Cloud classification of satellite radiance data by multicategory support vector machines. J. of Atmospheric and Oceanic Technology 21(2), 159–169 (2004)
6. Knutsson, H., Westin, C.F.: Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data. In: Proceedings of Computer Vision and Pattern Recognition 1993, pp. 515–523 (1993)
7. Knutsson, H., Westin, C.F., Westelius, C.J.: Filtering of uncertain irregularly sampled multidimensional data. In: Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers 1993, vol. 2, pp. 1301–1309 (1993)
8. Lee, J., Liu, R.: A Fuzzy Clustering Algorithm Based on Fuzzy Distance Norms for Asynchronously Sampled Data. In: 11th IEEE International Conference on Computational Science and Engineering, CSC 2008, pp. 361–368 (2008)
9. El-Melegy, M., Zanaty, E.A., Abd-Elhafiez, W.M., Farag, A.: On Cluster Validity Indexes in Fuzzy and Hard Clustering Algorithms for Image Segmentation. In: IEEE International Conference on Image Processing, vol. 6, pp. 5–8 (2007)

Fuzzy C-Means Inspired Free Form Deformation Technique for Registration

Edoardo Ardizzone, Roberto Gallea, Orazio Gambino, and Roberto Pirrone

Università degli Studi di Palermo
DINFO - Dipartimento di Ingegneria Informatica
Viale delle Scienze - Ed.6 - 3° piano - 90128 Palermo (ITALY)
{robertogallea,ardizzon,pirrone,gambino}@unipa.it

Abstract. This paper presents a novel method aimed to free form deformation function approximation for purpose of image registration. The method is currently feature-based. The algorithm is inspired to concepts derived from Fuzzy C-means clustering technique such as membership degree and cluster centroids. After algorithm explanation, tests and relative results obtained are presented and discussed. Finally, considerations on future improvements are elucidated.

Keywords: free form deformation, image registration, fuzzy clustering.

1 Introduction

Image registration is the process of overlaying two or more datasets (generally 2d or 3d), by finding a transform function which allows to match every correspondent pixel or voxel in the datasets. Such task is useful, among others, in many fields of Imaging and Computer Vision and is of great importance for Medical Imaging purposes. The registration procedure is an indispensable step for several purposes, such merging data captured from different modality (for example MR and PET), evaluate the effects of a therapy observable by the changes in two images acquired in different times, or to compare datasets available in the atlases containing known structures.

Feature based registration techniques compute the datasets-matching transform using a sparse set of data, i.e. the features. Such features are salient correspondences of the images, such as points or lines. Knowing the displacement vectors of these features is possible to interpolate the complete displacement function.

Several interpolation schemes have been proposed in literature, one of most popular is the Thin-Plate Spline surface fitting [1], with a number of variants, such as [2], [3], [4] and [5].

Thin-Plate Splines can model a surface constrained to contain several node points using an elegant algebraic solution. This is done by decomposing the whole transformation function in a linear (affine) part and the superposition of principal warps, each one independent from the others. Although the need for a search procedure is removed due to the existence of an affordable analytical solution,

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 148–154, 2009.

the computing cost of the TPS is still high. In this paper a novel registration method is proposed. This method requires no minimization at all, strongly improving the overall performance of the registration task. The algorithm is based on concepts derived from Fuzzy C-means clustering algorithm. Even though C-means relies on the determination of the clustering centroids, and this requires a minimization iterative procedure, for our method purposes the disposition of such centroids is given as a known starting condition. As a result, the optimization step is removed and the displacement function determination is the results of quick and simple sum and product operations.

The paper is arranged as follows: in sect. 2, after some remarks on the Fuzzy C-means algorithm, the registration algorithm is presented, in sect. 3 the tests conducted to validate the algorithm are described and the results, compared with TPS approach, are summarized. Section 4 contains considerations on the algorithm performance and the evaluation of the results obtained. Section 5 discusses how method performance can be improved and states the roadmap for further development of this Fuzzy-Inspired Registration project.

2 Methods

The proposed method works like other free form deformation feature-basaed techniques: provided a list of landmarks correspondence between a pair of images, namely the input and the target image, the algorithm finds a transformation function $f(x, y)$ which realizes a mapping between the pixels of the two images. The approximation of the function is constrained by the value of the displacements between the landmark points. These features can be manually chosen or automatically detected. The result is the interpolation of the displacements of every points pair in the images.

Provided that the landmarks points represent the evidence of the unknown displacement field required to align the two datasets, an inference should be made to estimate the whole displacement function.

In our framework the actual computation of this function is not obtained probabilistically but simplificated using Fuzzy C-means clustering derived technique. [6]. Anyway, even tough the centroids and the fuzzy membership concepts are useful for our deformation purpose, no clustering process is required, so, as a result, no minimization is performed actually.

Each landmark point could be considered as a cluster centroid in the feature space defined by spatial coordinates x and y. Then, for each pixel of the image a membership value relative to the spatial region centered in each of the landmarks could be assigned. In this way the displacement of each pixel can be chosen as the sum of every landmark displacement weighted by the relative membership value. This allows every landmark to have a stronger influence on closer pixels rather than on farther, in a proportional way. The amount of decay related to distance variation, and consequently the membership value, is controlled by means of the s exponent, the unique adjustable parameter in this registration scheme.

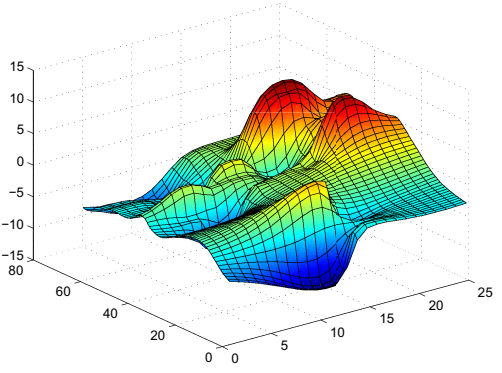


Fig. 1. Example of plot of a transform function for x values displacements

Given the above conditions, the displacement vector v_ij of the pixel at position i, j is given by:

$$v_{ij} = \sum_{l=1}^{m} u_{ij,l} v_l \tag{1}$$

where u_ij, l is the membership degree of the pixel i, j to the l-th centroid while v_l is the displacement of the l-th landmark, which is known.

This results in a transformation function with no discontinuities that exhibits nice properties such as interpolation with exact approximation of the landmark points and governable smoothness without no needs for a regularization parameter. An example of such functions is shown in Fig. 1.

Another remarkable advantage is that no time-consuming minimization procedure is required to obtain the transformation function since each of the required values is known in advance, what needs to be computed is just the membership matrix U using (2). The metric used is the Euclidean distance because is sufficient to take into account just spatial closeness between pixels and landmarks.

Such method can be used both for 2-D images and 3-D models, unique difference is the number of transformation functions to compute, one for each dimension.

$$u_{ij} = \frac{1}{\sum_{l=1}^{m} \left(\frac{d(x_i, c_j)}{d(x_i, c_l)} \right)^{\frac{2}{s-1}}} \tag{2}$$

3 Results

Once the method was designed, several kinds of tests were conducted on sets of photographic and synthetic medical images provided by Brainweb [7,8,9,10].

A preliminary test was operated on a dummy picture consisting of a white circle on black background, Fig. 2a. The image was strongly deformed as in Fig. 2b Then the registration procedure was applied using a different number

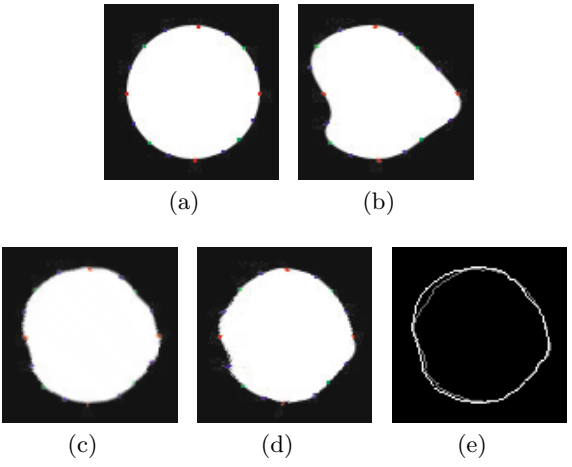


Fig. 2. From left to right: original test pattern (a), deformed version of test pattern (b), registered versions with 16 landmarks: proposed method (c), TPS approximation (d), difference between two results, edges shown (e)

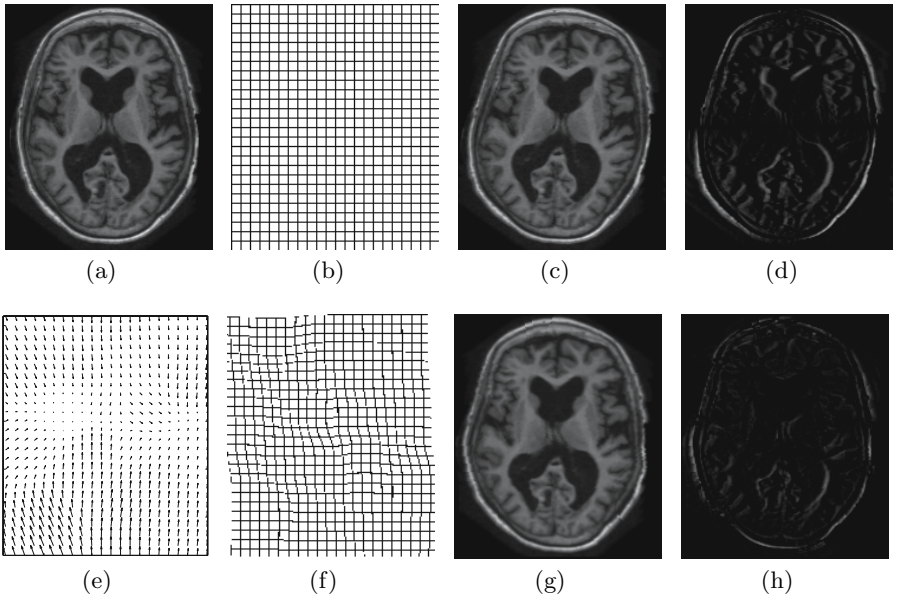
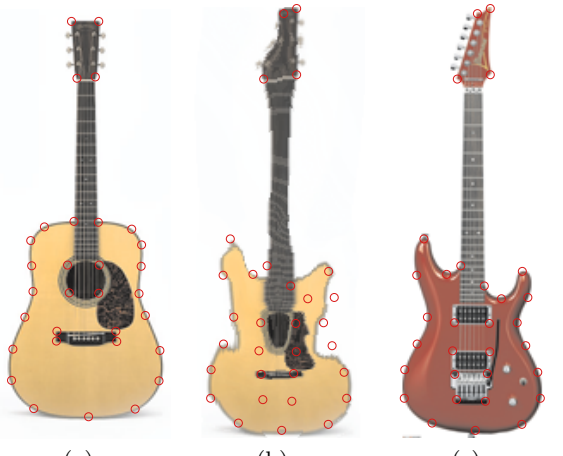


Fig. 3. From left to right: ground truth (a), initial registration grid (b), deformed versione of the image (c), initial error image (d), deformation field (e), deformed registration grid (f), registered version with 18 landmark points (g), final error image (h)

of landmarks both with our and TPS algorithms, the results are shown in Fig. 2c and Fig. 2d. Differences between the two methods are elucidated in Fig. 2e,

Table 1. Results summary, best measures underlined

	Fuzzy Approximation				TPS Approx			
s	MSE	SSD	MI	SSIM	MSE	SSD	MI	SSIM
1.2	30.0057	1098447	1.2673	0.7842				
1.4	29.6038	1083737	1.2690	<u>0.7903</u>				
1.6	<u>29.3464</u>	<u>1074314</u>	<u>1.2693</u>	0.7894				
1.8	30.6666	1122642	1.2650	0.7723				
2.0	32.2718	1181406	1.2584	0.7424	31.0487	113631	1.2669	0.7889
2.2	34.6014	1266688	1.2509	0.7133				
2.4	36.9765	1353635	1.2435	0.6808				
2.6	38.7948	1420200	1.2378	0.6544				
2.8	40.1140	1468493	1.2331	0.6337				
3.0	41.4389	1516997	1.2292	0.6153				



(a) (b) (c)

Fig. 4. Example of morphing: acoustic guitar (a) morphed (b) into an electric guitar (c) using 31 landmark points

where the brighter contour represents the TPS approximation scheme, while the darker one refers to the proposed method.

The results exhibit that the two methods act in a similar way; small differences are present just in correspondence of large deformations (like in the middle part of the pattern).

The second class of tests consists on the registration of a manually deformed image onto its original version. Method performance is then evaluated using several similarity metrics: sum of squared difference (SSD), mean squared error (MSE) and mutual information (MI) as objective measures, and Structural Similarity ($SSIM$) [11] as the subjective one. The algorithm was ran using different fuzziness values s, visual results for the proposed method are depicted in Fig. 2 and measures are summarized in Table 1. Comparisons with thin-plate spline approach are also presented.

Other qualitative tests, merely visual, were performed on various photographic images in order to evaluate the goodness of the algorithm for other purposes,

such as morphing. In Fig. 4 are shown the results obtained executing the morph of two sample images.

4 Discussion

The tests conducted showed that the framework exhibits good performance for landmarks based registration of 2-D images.

The comparisons with TPS technique showed that the two methods have similar performances. However, the presented algorithm results preferable since it is by far faster: for example on a P4 processor machine equipped with Matlab R2008a, for the registration of a 208x176 pixels image using 16 landmarks, our technique takes around 17% of the time taken by TPS approximation (approximately 6.4 seconds versus 37.2 seconds). This difference is due to the fact that using TPS approximation implies, in order to obtain the transform function, to compute the solution of a variational problem, i.e. executing an optimization procedure, which is not required for our technique that relies on simple distance measures and weighted sums.

Even tough there exists the problem of the tuning of the smoothing parameter s, experiments shown that the optimal value generally lies in a range between 1.4 and 2, and a few monodimensional search attempts (on average 3-4) using bisectional strategies such as golden ratio, are enough to find the right solution, keeping the method convenient.

Comparing the size of data structures it can be seen that in our approach M values need to be stored for the landmarks displacements and M values for the membership degrees of each point. However, once every single pixel has been transformed, its membership degree can be dropped, so the total data structure is $2M$ large. TPS approximation has a little more compact structure, in fact it needs just to maintain the $M + 3$ surface coefficients (M for the non-linear part and 3 for the linear one). However, the storing complexity is $O(M)$ for both methods, i.e. linear in the number of landmarks used, and thus equivalent.

5 Conclusions and Future Works

A novel method aimed to free form deformation and registration of 2d images has been presented. This technique was inspired by the Fuzzy C-means clustering algorithm, although just the concept of centroids and fuzzy membership degree are used and no clustering operation needs to be performed at all. The proposed methods performance was evaluated with several tests using both objective and subjective measures and the results were compared to free form deformation technique based on Thin-Plate Spline surface approximation. From the study it resulted that both techniques provide similar results. However, our method is about 6 times faster due to its straightforward computation requiring no optimization process.

The next step of the study will be the use of the same technique for registering 3-D medical volume datasets usingn an automatic landmarks detection. The

feature points are intended to be chosen by means of contour curvature information and similarity considerations on the inner structures. Subsequently, the landmark points-based approach will be left toward a fully automatic pixel-based elastic registration framework.

References

1. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
2. Arad, N., Dyn, N., Reisfeld, D., Yeshurun, Y.: Image warping by radial basis functions: Application to facial expressions. Computer Vision, Graphics, and Image Processing. Graphical Models and Image Processing 56(2), 161–172 (1994)
3. Sprengel, R., Rohr, K., Stiehl, H.S.: Keywords— thin-plate splines, regularization, image registration i. thin-plate spline interpolation
4. Bartoli, A., Perriollat, M., Chambon, S.: Generalized thin-plate spline warps. In: IEEE International Conference on Computer Vision and Pattern Recognition, cvpr (2007)
5. Johnson, H.J., Christensen, G.E., Sc, D.: Consistent landmark and intensity-based image registration. IEEE Transactions on Medical Imaging 21, 450–461 (2002)
6. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms (Advanced Applications in Pattern Recognition). Springer, Heidelberg (1981)
7. Cocosco, C.A., Kollokian, V., Kwan, R.K.-s., Pike, G.B., Evans, A.C.: Brainweb: Online interface to a 3d mri simulated brain database. In: NeuroImage, p. 425 (1997)
8. Kwan, R.K.S., Evans, A.C., Pike, G.B.: Mri simulation-based evaluation of image-processing and classification methods. IEEE Transactions on Medical Imaging 18(11), 1085–1097 (1999)
9. Kwan, R.K.-S., Evans, A.C., Pike, G.B.: An extensible mri simulator for post-processing evaluation. In: Höhne, K.H., Kikinis, R. (eds.) VBC 1996. LNCS, vol. 1131, pp. 135–140. Springer, Heidelberg (1996)
10. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imaging 17(3), 463–468 (1998)
11. Wang, Z., Bovik, A.C., Sheikh, H.R., Member, S., Simoncelli, E.P., Member, S.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612 (2004)

Interpretability Assessment of Fuzzy Rule-Based Classifiers

Corrado Mencar, Ciro Castiello, and Anna Maria Fanelli

Department of Informatics, University of Bari,
via E. Orabona, 4 -70125- Bari, Italy
{mencar,castiello,fanelli}@di.uniba.it

Abstract. Interpretability is one of the most important driving forces for the adoption of fuzzy rule-based classifiers. However, it is not given for granted, especially when fuzzy models are acquired from data. Therefore, evaluation of interpretability should be regarded as a major research topic. In this paper, we describe a technique for automatic interpretability assessment, based on the co-intension of the semantics of the knowledge base with the intrinsic semantics designated by linguistic labels. The core of the evaluation technique relies on the propositional view of rules and on logical operations. An illustrative example shows how the proposed approach can be useful in detecting lacks of interpretability for a simple knowledge base.

1 Introduction

Interpretability is one of the most important driving forces for the adoption of fuzzy rule-based systems, since they are intended to perform intelligent tasks while allowing for a representation of knowledge that can be easily read and understood by their users. Interpretability, however, is *not* given for granted when fuzzy models are used, especially when they are acquired from data. The main problem is that data-driven design has a great number of degrees of freedom (number of fuzzy sets, their shape, position, etc.) and may end up with fuzzy models that are very accurate but very hard to comprehend. For this reason, interpretability constraints have been defined so as to bind data-driven design in order to derive interpretable fuzzy models [1]. This usually comes to a price, that is a lower accuracy with respect to unconstrained design. Furthermore, often interpretability is accounted without taking care of accuracy. This approach has been criticized, since interpretable but inaccurate models are as useless as very accurate but not interpretable ones [2]. Therefore, a proper evaluation of interpretability should be regarded as a major research topic.

However, interpretability assessment is an ill-posed problem because the definition of interpretability eludes any formal characterization. In [3], Michalski gives a referential definition of interpretability, the so-called "comprehensibility postulate". In short, a rule base is interpretable if it is defined by symbolic structures, semantically and structurally similar to human knowledge, so as to be interpretable in natural language (for a more extensive discussion of the postulate,

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 155–162, 2009.

see [4]). This postulate justifies the use of linguistic values in rule-based fuzzy systems, but that is not enough to guarantee interpretability. In [5] Zadeh introduces the notion of *co-intension*, a semantic relation between concepts. Roughly speaking, two concepts are co-intensive if they refer to almost the same objects. In fuzzy rule-based systems, rules are defined by composition of linguistic terms, which are related to the two different semantics defined by the fuzzy model and designated by the linguistic label. By merging the notion of co-intension with the comprehensibility postulate, we derive a formulation of interpretability that can be more helpful for designing assessing techniques: a rule base is interpretable if the two semantics related to each linguistic label are co-intensive.

On the basis of this definition, we propose an automatic technique for evaluating interpretability. Our approach evaluates interpretability by assessing the co-intension of the semantics of the rule base of a fuzzy model with the intrinsic semantics designated by linguistic labels. The core of the evaluation technique relies on the propositional view of rules and on logical operations. We expect that, for interpretable knowledge bases, logical operations on rules do not change their semantics and, hence, do not affect accuracy. If this is assumption is violated, we deduce a lack of interpretability.

We focus our research on fuzzy rule-based classifiers, described in Section 2. The proposed approach is then described, by first focusing on its rationale and then on the methodology (Sections 3.1 and 3.2, respectively). An illustrative example is reported in Section 4, to show how the proposed approach can be useful in detecting lacks of interpretability for a simple knowledge base. The paper is concluded with some final remarks.

2 Fuzzy Rule-Based Classifiers

We consider a classifier as a system computing a function of type:

$$f : \mathbf{X} \longrightarrow \Lambda, \tag{1}$$

where $\mathbf{X} \subseteq \mathbf{R}^n$ is an n-dimensional input space, and $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_c\}$ is a set of class labels. If a dataset D of pre-classified data is given, i.e.

$$D = \{(\mathbf{x}_i, l_i) | \mathbf{x}_i \in \mathbf{X}, l_i \in \Lambda, i = 1, 2, \ldots, N\}, \tag{2}$$

then the classification error can be computed as:

$$E(f, D) = \frac{1}{N} \sum_{i=1}^{N} (1 - \chi(l_i, f(\mathbf{x}_i))), \tag{3}$$

being $\chi(a, b) = 1$ iff $a = b$ and 0 otherwise.

A fuzzy rule-based classifier (FRBC) is a system that carries out classification (1) through inference on a knowledge base. The knowledge base includes the definition of a linguistic variable for each input. Thus, for each $j = 1, 2, \ldots, n$, linguistic variables are defined as:

$$V_j = (v_j, X_j, Q_j, S_j, I_j), \tag{4}$$

being:

- v_j the name of the variable;
- X_j the domain of the variable (it is assumed that $\mathbf{X} = X_1 \times X_2 \times \cdots \times X_n$);
- $Q_j = \{q_{j1}, q_{j2}, \ldots, q_{jm_j}, \text{ANY}\}$ is a set of labels denoting linguistic values for the variable (e.g. SMALL, MEDIUM, LARGE);
- $S_j = \{s_{j1}, s_{j2}, \ldots, s_{jm_j+1}\}$ is a set of fuzzy sets on X_j, $s_{jk} : X_j \to [0,1]$;
- I_j associates each linguistic value q_{jk} to a fuzzy set s_{jk}. We will assume that $I_j(q_{jk}) = s_{jk}$.

We assume that each linguistic variable contains the linguistic value "ANY" associated to a special fuzzy set $s \in S_j$ such that $s(x) = 1, \forall x \in X_j$.

The knowledge base of a FRBC is defined by a set of R rules. Each rule can be represented by the schema:

$$\text{IF} \quad v_1 \text{ IS } [\text{NOT}]\ q_1^{(r)} \text{ AND } \cdots \text{ AND } v_n \text{ IS } [\text{NOT}]\ q_n^{(r)} \text{ THEN } \lambda^{(r)}, \quad (5)$$

being $q_j^{(r)} \in Q_j$ and $\lambda^{(r)} \in \Lambda$. Symbol NOT is optional for each linguistic value. If for some j, $q_j^{(r)} = \text{ANY}$, then the corresponding atom "v_j IS ANY" can be removed from the representation of the rule.[1]

Inference is carried out as follows. When an input $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is available, the strength of each rule is calculated as:

$$\mu_r(\mathbf{x}) = s_1^{(r)}(x_1) \otimes s_2^{(r)}(x_2) \otimes \cdots \otimes s_n^{(r)}(x_n), \quad (6)$$

being $s_j^{(r)} = \nu_j^{(r)}(I_j(q_j^{(r)}))$, $j = 1, 2, \ldots, n$, $r = 1, 2, \ldots, R$. Function $\nu_j^{(r)}(t)$ is $1 - t$ if NOT occurs before $q_j^{(r)}$, otherwise it is defined as t. The operator $\otimes : [0,1]^2 \to [0,1]$ is usually a t-norm, such as minimum or product.

The degree of membership of input $\mathbf{x}$ to class λ_i is computed by considering all the rules of the FRBC as:

$$\mu_{\lambda_i}(\mathbf{x}) = \frac{\sum_{r=1}^{R} \mu_r(\mathbf{x}) \chi(\lambda_i, \lambda^{(r)})}{\sum_{r=1}^{R} \mu_r(\mathbf{x})}. \quad (7)$$

Finally, since just one class label has to be assigned for the input $\mathbf{x}$, the FRBC assigns the class label with highest membership (ties are solved randomly):

$$f_{FRBC}(\mathbf{x}) = \lambda \Leftrightarrow \mu_\lambda(\mathbf{x}) = \max_{i=1,2,\ldots,c} \mu_{\lambda_i}(\mathbf{x}). \quad (8)$$

3 Interpretability Assessment

We assume the availability of an interpretable FRBC, verifying a number of interpretability constraints so that the rule base is described in terms of linguistic values.

[1] The sequence NOT ANY is not allowed.

3.1 Rationale

The proposed approach for interpretability assessment relies on the formal structure of the FRBC. The rationale behind this approach comes from the observation that the rule base is the linguistic interface of the FRBC to the user. For an interpretable knowledge base, the user should be able to understand the classification rules by simply observing their linguistic representation. All the semantic information (fuzzy sets attached to linguistic values, t-norm used for conjunction, etc.) should be hidden to the user because – this is the key point of interpretability – the semantics of FRBC knowledge should be co-intensive with the user's knowledge, recalled by the linguistic terms.

To assess interpretability, we exploit the cognitive structures that are shared by users and FRBC. In particular, we observe a strict affinity of a FRBC rule base to logical propositions. Actually, rules are formed so as to resemble propositions, so that they can be understood by users. In consequence of this, FRCB and users share the propositional view of rules. Being like propositions, rules could be transformed by truth-preserving operators without any change of the semantics. This is not completely true since the application of such operators may distort the semantics of rule (which is fuzzy); however, we should expect small distortions due to the shared propositional view between FRBC and users.

The core of our approach is the following: given a rule base of a FRBC, we represent it as a collection of propositions, then we transform it through a truth-preserving operator, thus obtaining a new set of propositions, that represents a new rule base. We then compare the two rule bases on the basis of their classification ability: if they differ too much, then we conclude that the logical view of rules is wrong. Also, since rules are defined so as to resemble logical propositions, we derive that the semantics of rules (which is responsible of classification) is such that logical view is not possible. This means that the semantics of rules is not co-intensive with user knowledge, since users expect that truth-preserving transformations of propositions do not change (too much) their semantics.

An issue arises in the choice of the truth-preserving transformation. Actually, several transformations can be considered, but we choose to apply a transformation that minimizes the number of linguistic terms used. This choice has a twofold advantage. First, by eliminating as many terms as possible, we test if logical view of rules is preserved with the minimum required information. Second, if assessment leads to positive results, we could retain the simplified rule base because it is easier to read than the original.

3.2 Methodology

The proposed approach for interpretability assessment is based on a four-stage strategy.

Definition of Truth Tables. Each rule of the FRBC is seen as a proposition, i.e. a combination of propositional variables that is considered true for a class.

For each class label $\lambda_i \in \Lambda$ and for each rule r, a truth function π_i is defined on the propositional variables defined for the FRBC as:

$$\pi_i(\chi_{11}^{(r)},\ldots,\chi_{1m_1}^{(r)},\chi_{21}^{(r)},\ldots,\chi_{2m_2}^{(r)},\ldots,\chi_{n1}^{(r)},\ldots,\chi_{nm_n}^{(r)}) = \chi(\lambda_i,\lambda^{(r)}), \quad (9)$$

being $\chi_{jk}^{(r)} = \chi(q_j^{(r)}, q_{jk})$. Inputs $\chi_{jk}^{(r)}$ assume value X ("don't care") if the corresponding linguistic value is ANY or, in other words, the linguistic variable V_j does not occur in the r-th rule. For any other combination of inputs, the output of π_i is undefined, i.e. any truth value is possible (again, this condition is usually referred as "don't care").

Each truth function π_i can be represented as a truth table, which enumerates any combination of assignments to the propositional variables of the FRCB and associates the value of π_i to each combination. Combinations associated to undefined values of π_i are not included in the table. The number of rows of each truth table matches the number of rules of the FRCB. This prevents the combinatorial explosion of rows that would be expected in the general case of truth function representation.

Minimization. Once each truth table has been built, it can be processed so as to be minimized. The minimization procedure produces a new truth table without modifying the truth function (where it is defined). The new truth table has a number of rows not greater than the original truth table. It also has a number of X values in its inputs not smaller than in the original truth table. Furthermore, minimization guarantees that any further simplification (in terms of rows and inputs) provides for a truth function different from the original.

The Quine-McCluskey (QMC) algorithm represents an effective mechanism for minimization of truth tables [7]. It is mainly based on the distributive property, which simplifies propositions according to the law: $ABC + A\bar{B}C \equiv AC$.

Although computationally expensive, the QMC algorithm can be implemented by an efficient procedure that exploits the peculiar structure of truth tables derived from FRBC rules to perform minimization quickly.

Reconstruction. After minimization, a new FRBC is built from the rows of the minimized truth table. For each class label $\lambda_i \in \Lambda$ we consider the minimized table associated to the truth function π_i. A rule is built for each row with output equal to 1. It is easy to show that for each j there is at most one k such that $\chi_{jk}^{(r)} \neq X$. Therefore, the antecedent of the rule can be defined by atoms v_j IS [NOT] $q_j^{(r)}$ where:

- $q_j^{(r)} = q_{jk}$ if $\chi_{jk}^{(r)} \neq X$;
- NOT occurs if $\chi_{jk}^{(r)} = 0$ and it does not occur if $\chi_{jk}^{(r)} = 1$;
- $q_j^{(r)} = $ ANY if $\forall k : \chi_{jk}^{(r)} = X$;

Atoms with ANY are removed to improve the readibility of the rule. The consequent of the rule is λ_i.

Comparison. The two FRBCs, the first with the original rule base and the second with the minimized version, are compared in terms of classification error on the same data.

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is High THEN Class 3
If Magnesium is Medium AND Flavonoids is High AND Proline is Medium THEN Class 3

Fig. 1. The rule base of the FRBC considered in the example

If the two errors differ too much, we conclude that the original FRBC lacks of interpretability. Its accuracy is mainly due to the semantic definition of linguistic values, which do not correspond to the propositional view of rules. The FRBC can be used for classification as a "grey box", but its labelling is arbitrary and not co-intensive with user knowledge. Attaching natural language terms to such FRCB is useless and potentially misleading.

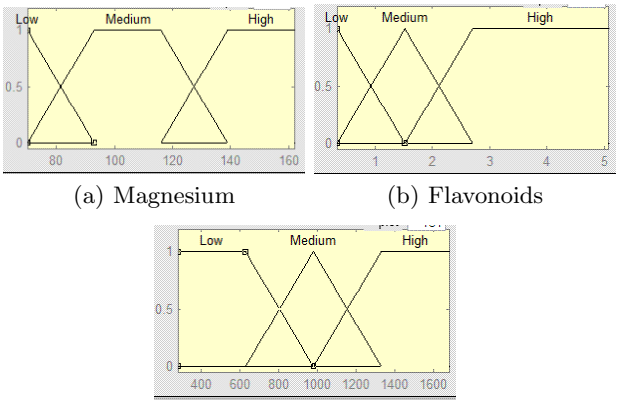
If the two errors are very similar, we conclude that the original FRBC is interpretable. The semantic definition of linguistic values is coherent with the logic operators used in minimization. In this sense, the semantic of linguistic values is co-intensive with user knowledge. We could retain the simplified FRBC because its interpretability is greater than the original (due to its higher simplicity) while its accuracy is almost the same of the original.

There is no threshold to decide if a FRBC is interpretable or not, but rather a continuous spectrum of possibilities. Interpretability – as expected – is a matter of degree, and the degree of interpretability is, in our approach, inversely proportional to the difference of accuracies. Even with small variations of accuracy, some considerations can be drawn on the interpretability of the FRBC, as shown in the next Section.

4 Illustrative Example

We consider a FRBC obtained from the application of HILK [6], a tool for building interpretable fuzzy rule-based systems. HILK allows for the definition of fuzzy rule-based systems from empirical learning, expert knowledge, or both. The resulting systems are considered highly interpretable because fuzzy sets defined for each input satisfy a number of interpretability constraints. Furthermore, the number of fuzzy sets per input, the number of inputs and the number of rules are kept as small as possible since, in principle, the simplest is the knowledge representation, the easier is to read and understand.

In our experimentation, the FRBC obtained from HILK was acquired from data, in order to classify Wine data, a well-known benchmark dataset, freely available from UCI repository [8]. The knowledge base of the FRBC is reported in fig. 1, while the linguistic variables are shown in fig. 2. The FRBC provided 10.67% of classification error on the entire dataset. We transformed the rule base of the FRBC into three truth tables – one for each class – minimizing them with QMC algorithm; then we rebuilt the FRBC obtaining the simplified rule base reported in fig. 3.



(a) Magnesium (b) Flavonoids

(c) Proline

Fig. 2. The linguistic variables used in the example

If Flavonoids is Low THEN Class 1
If Flavonoids is Medium THEN Class 2
If Flavonoids is High AND Proline is Low THEN Class 2
If Flavonoids is High AND Proline is NOT Low THEN Class 3

Fig. 3. The rule base of the FRBC after minimization

We observe that the number of rules has been reduced to four, and the linguistic variable "Magnesium" has been removed. From the logical viewpoint, the two rule bases are equivalent since they compute the same truth functions. However, after applying the minimized rule base to the dataset, we obtained 11.24% of classification error, i.e. an increase of +0.57%, corresponding to one more misclassified pattern over 178, which is defined as (Magnesium=162, Flavonoids=2.27, Proline=937). For the misclassified pattern we observe that, according to the definitions of the linguistic variables, Magnesium is high, Proline is medium, and Flavonoids is both medium and high to a significant degree. However, in the original rule base, there is not any classification rule activated when Magnesium is high, Proline is medium and Flavonoids is high. From the logical viewpoint, the truth functions are undefined for this condition. As a consequence, the simplified rule base subsumes this condition in the fourth rule, assigning the pattern to class 3, while the original FRBC arbitrarily assigns the pattern to class 2. The correct classification of the pattern performed by the original FRBC is due to the semantic specification of linguistic values, which does not emerge from the propositional view of rules. In this sense, the original FRBC lacks of interpretability for a specific case. However, due to the reduced increase of classification error, we conclude that the original FRBC is highly interpretable and we can retain the simplified version, which offers a further simplified knowledge base.

5 Conclusion

Assessment of interpretability is not an easy task. Difficulty is mainly due to the blurry definition of interpretability, which requires co-intension with human knowledge. Evaluating interpretability is a subjective task, which could be tedious and even ill-posed, therefore automatic techniques for interpretability assessment are useful, but they should embody some information on semantic co-intension. In this paper, we have proposed an approach for automatically evaluating interpretability of rule-based fuzzy classifiers, exploiting the propositional view of rules as a mean to define co-intension. On the basis of this hypothesis, a procedure has been devised, so as to evaluate how much the semantics of fuzzy rules is coherent with their logical view.

The illustrative example shows the effects of the proposed approach for a simple knowledge base. Actually, we were able to detect a flaw in interpretability for a FRBC that verifies several interpretability constraints.

Research on this topic is in progress, especially in the direction of enhancing the efficiency of the minimization algorithm. That will allow an extensive experimentation over a wider class of knowledge bases, in order to promote deeper insights of the proposed technique. Further enhancements could spring from the use of additional information to refine the definition of co-intension in computational terms.

References

1. Mencar, C.: Theory of Fuzzy Information Granulation: Contributions to Interpretability Issues. PhD thesis, University of Bari, Italy (2005)
2. Riid, A., Rustern, E.: Interpretability of Fuzzy Systems and Its Application to Process Control. In: Proc. of Fuzzy Systems IEEE International Conference, 2007, FUZZ-IEEE 2007, pp. 1–6 (2007)
3. Michalski, R.S.: A theory and methodology of inductive learning. Technical Report UIUCDCS-R-83-1122, Department of Computer Science, University of Illinois, Urbana (January 1983)
4. Mencar, C., Castellano, G., Fanelli, A.M.: On the role of interpretability in fuzzy data mining. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (IJUFKS) 15(5), 521–537 (2007)
5. Zadeh, L.A.: Is there a need for fuzzy logic? Information Sciences 178, 2751–2779 (2008)
6. Alonso, J.M., Magdalena, L., Guillaume, S.: HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. International Journal of Intelligent Systems 23, 761–794 (2008)
7. McCluskey, E.J.: Minimization of Boolean functions. Bell Labs Technical Journal 35(5), 1417–1444 (1956)
8. UCI Machine Learning Repository Content Summary, http://www.ics.uci.edu/~mlearn/MLSummary.html

Metaclustering and Consensus Algorithms for Interactive Data Analysis and Validation

Ida Bifulco, Carmine Fedullo, Francesco Napolitano, Giancarlo Raiconi, and Roberto Tagliaferri

NeuRoNe Lab, DMI, University of Salerno,
via Ponte don Melillo, 84084 Fisciano (SA) Italy
{ibifulco,cfedullo,fnapolitano,gianni,rtagliaferri}@unisa.it
http://www.neuronelab.dmi.unisa.it

Abstract. Clustering of real-world datasets is a complex problem. Optimization models seeking to maximize a fitness function assume that the solution corresponding to the global optimum is the best clustering solution. Unfortunately, this is not always the case, mainly because of noise or intrinsic ambiguity in the data. In this work we present a set of tools implementing classical and novel techniques to approach clustering in a systematic way, with an application example to a complex biological dataset. The tools deal with the problem of generating multiple clustering solutions, performing cluster analysis on such clusterings (i.e. Meta Clustering) and reducing the final number of clusterings by the appropriate application of different Consensus techniques. A subsequent crossing of prior knowledge to the obtained clusters helps the user in better understanding its meaning and validates the solutions.

Keywords: meta clustering, data visualization, consensus clustering.

Introduction

Two main problems arise when dealing with clustering of multidimensional data: the former regards multiple different solutions, the latter concerns the visualization of the results, in terms of different ways of grouping high-D data.

The most used clustering algorithms start from a random or arbitrary initial configuration and then evolve to a local minimum of the objective function. In complex problems (in many real cases) there are several minima and more than one can explain in a convincing manner the data distribution. We described the process of generating many solutions corresponding to local minima in [7].

At this point, the problem of analyzing many different solutions arises. One possible approach to this problem is to merge more solutions to obtain a new clustering: that is called Consensus Clustering [2,6,14,19,11]. Another approach consists in the process of clustering such solutions, namely Meta Clustering [12]. The problem is that both the approaches alone are not adequate to give the required information to the domain expert, because consensus algorithms give a unique solution which is "the most similar" to all the clusterings, while the meta

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 163–170, 2009.

clustering gives only subsets of solutions. We put the two approaches together to extract few solutions from each meta cluster.

As a final step, we cross the obtained multiple solutions with different kinds of prior knowledge. We analyze the matching between clusterings and prior knowledge labels by searching for the most enriched clusters, that is clusters that are best characterized for some labels. This can help the users on two sides: validating the obtained clusters with respect to the chosen labels and stimulate hypotheses about unlabeled data falling into strongly characterized clusters.

Our software implementation of the system is called Modular Interactive Dendrogram Analyzer (MIDA [23]) and is developed for the Matlab environment. We will show an application of the ideas presented in this paper through the use of MIDA with an elaboration of the HeLa-cell dataset [22,17].

1 Our Approach to Clustering

When clustering complex datasets, the existence of a unique optimal solution is usually questionable. In such cases, the problem of extrapolating a small number of good different solutions becomes crucial. Our approach to clustering includes the generation of a large number of solutions, the clustering of such solutions, their reduction through consensus clustering applied to the clusters of solutions and the analysis of the obtained final solutions supported by prior knowledge. In the following, some details are given for each of the steps.

1.1 Generation of Multiple Clustering Solutions

The generation of multiple clustering solutions is performed exploiting a Global Optimization approach. However, in many applications the function of interest is multi-modal and possibly not differentiable and it is with this view in mind that the Controlled Random Search (CRS) algorithm [18] was initially developed. Some CRS algorithms also include genetic techniques [9,10]. We applied such technique to cluster analysis. We used a Price based Global Optimization algorithm to explore local minima of the K-means or Expectation Maximization (EM) objective functions. Instead of sticking to the best found solution, we collect all the solutions corresponding to local minima.

1.2 Clustering the Clustering Solutions

The first step towards the analysis of the solutions obtained as previously shown, consists in grouping together the similar ones, that is in turn a clustering problem. This process is called Meta Clustering [12]. The subset of clusterings can be evaluated by the domain expert or by quality indices. Meta-Clustering is completely based on a similarity measure between partitioning. Many such similarities are known in literature, such as Minkowski Index, Jaccard Coefficient, Correlation and Matching Coefficients (all found in [3]). Once the similarity between each couple of solutions is obtained (similarity matrix), a hierarchical clustering can be directly

applied to it. The result is a meta-clustering dendrogram in which leaves represent the clustering solutions obtained in the previous step.

1.3 Reducing the Number of Solutions

Consensus clustering, also known in literature as clustering ensembles or clustering aggregation, is the process of extrapolating a single clustering solution from a collection, in such a way to maximize a measure of agreement. This problem is NP complete [2]. We implemented software modules for seven different Consensus techniques, including an intersection-based method developed by us [6]. The other modules resulted from an integration of the software released by the respective authors of the implemented methods into our interactive framework. They include: Furthest Consensus Algorithm (FC) [14] and BallClust [14], based on pairwise similarity; Best [14], based on objective function; Cluster-based Similarity Partitioniong Algorithm (CSPA) [19]; Iterative Voting Consensus [11] and Iterative Probabilistic Voting Consensus (IPVC) [11] based on iterative approach.

1.4 Validating the Final Solutions

The analysis of the final (small) set of clustering solutions is the real objective of the entire process. For this step we propose an exhaustive enrichment analysis based on prior knowledge. This information can be exploited in at least two ways: to validate a clustering result and to infer new knowledge. The validation of a clustering can be obtained comparing the prior knowledge and the knowledge obtained from the clustering itself. The validation of the clustering methodology can work only if there exist a clear correspondence between the variables used to produce the clustering and prior knowledge used to validate it: this is not always the case.

The prior knowledge can also be used to produce new knowledge inferred by the presence or absence of objects of a given class in a cluster. This is the approach embraced here, where we seek for clusters (in the final solutions set) that are the most enriched with respect to some prior knowledge about the data. When a cluster is highly characterized by the presence of a given type of points, hypotheses can be made about the points that belong to different classes or that are not known to belong to any class. We perform an exhaustive search for such enriched clusters: consensus clustering is applied to all the nodes of the meta-clustering tree and each obtained solution is searched for enriched clusters. Of course, this phase is done in a non-interactive way, while the results are shown visually as explained in section 2.

2 Visual and Interactive Tools

The human eyes and brain together make a formidable pattern detection tool, but, for them to work, the data must be represented in a low-dimensional space, usually of two or three-dimensions. Our tools enable the user to go through the entire clustering process with the support of ad hoc visualizations and interactive tools. The software we developed runs in the MATLAB environment and is mainly integrated in the MIDA (Modular Interactive Dendrogram Analyzer, [23], Fig. 1) project.

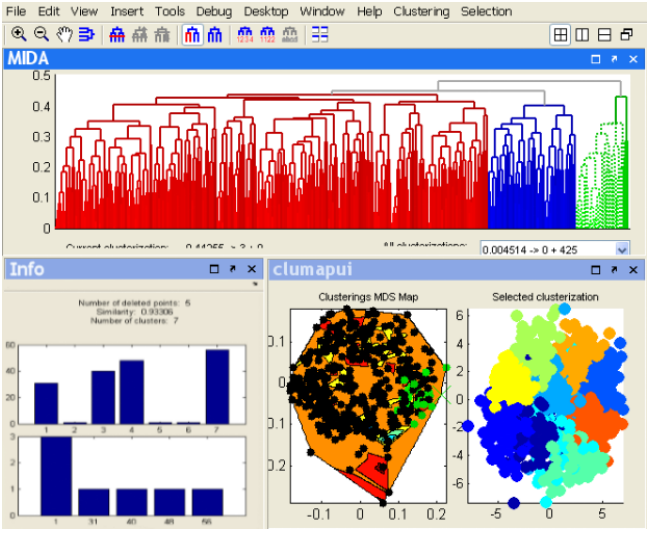


Fig. 1. The HeLa-Cell dataset as analyzed in MIDA. Top line: MIDA core, displaying the interactive dendrogram. Bottom line: Info Module, showing statistics about current Consensus (left) and Clustering Map module, showing the Clustering Map on the left part and currently selected clustering on the right part (right)

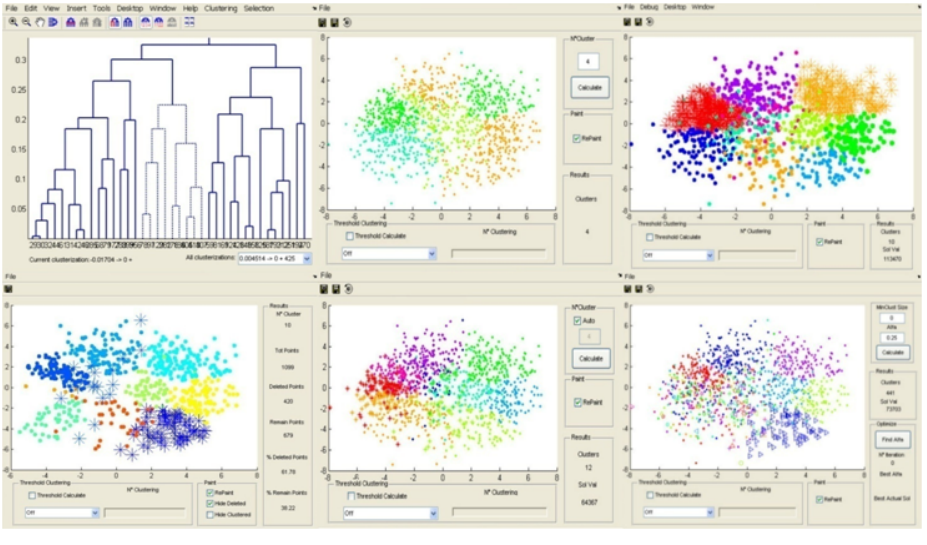


Fig. 2. MIDA running consensus modules. When the user selects a subtree on the meta-clustering dendrogram, each consensus module apply a different consensus technique and shows the resulting clustering. From top to bottom, left to right: FC, Intersection, IPVC, CSPA and BallClust modules are shown.

2.1 MIDA

MIDA (Modular Interactive Dendrogram Analyzer, [23], Fig. 1) is the software framework we use to integrate most of our visualizations and interactive tools. The core of MIDA is an interactive dendrogram in which, when used for meta clustering, each leaf is associated to a clustering. A dashed line, representing the current threshold, can be dragged to obtain the corresponding clustering in real time. Each subtree of the dendrogram corresponds to a meta cluster. With a mouse click, consensus modules start applying consensus directly on the solutions belonging to such subtree and all currently open visualizations are updated accordingly.

In a dedicated module, indicators of the consensus solution are summarized, such as distortion value, number of clusters, number of points per cluster, number of points eliminated by the intersection method, the mean distance between the solution and the leaves of the subtree, and other information about the current solution.

The space of clustering solutions is visualized through a dedicated module. The MDS technique can be exploited to represent each clustering as a point in the plane, with the distance between points representing the dissimilarity between the corresponding clusterings. Finally, adding the fitness value as the third coordinate for each point, we obtain a *clustering map* as that in Fig. 1.

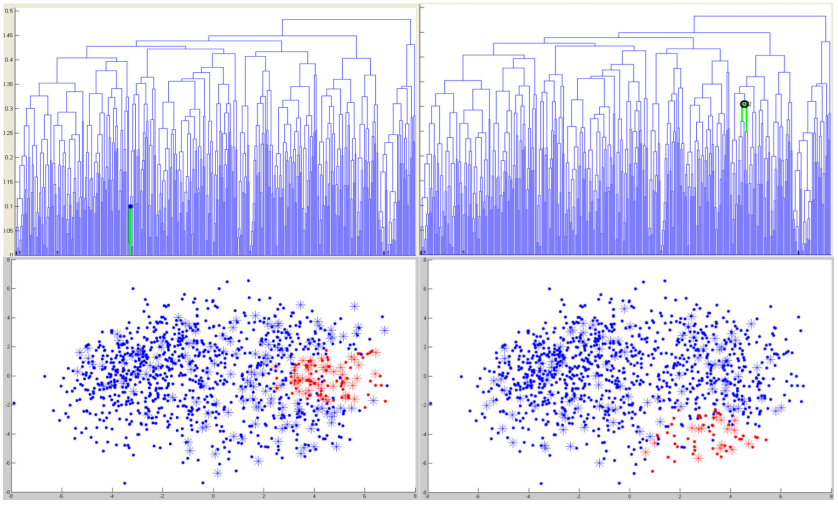


Fig. 3. Same clustering population, different labellings. Top row: meta-clustering dendrograms obtained optimizing the distortion function. The circles show the roots of the subtrees whose consensus produces the most enriched clusters for the Cell Cycle GO annotation (left) and Transport GO annotation (right). Bottom row: MDS with red dots showing points belonging to the most enriched cluster for the Cell Cycle label (left) and the Transport label (right). Stars show positively annotated genes for the corresponding label. The corresponding consensus methods are respectively CSPA and FC. The enrichment for the two clusters is respectively 2.46 and 2.30.

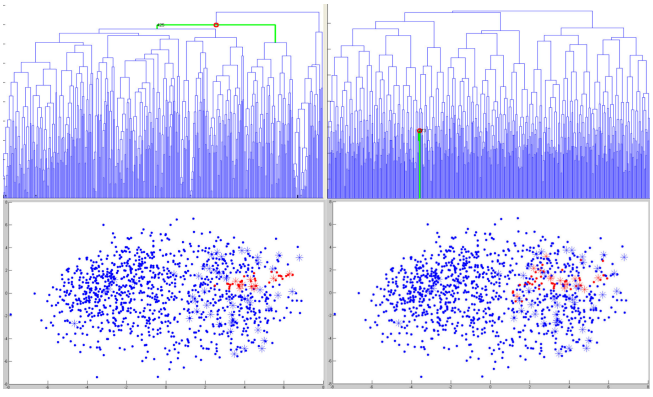


Fig. 4. Different clustering populations, same labelling. Top row: meta-clustering dendrograms obtained optimizing the distortion function (left) or the EM objective function (right) for the Cell Proliferation GO annotation. The circles show the roots of the subtrees whose consensus produces the most enriched clusters for such labelling. Bottom row: MDS with red dots showing points belonging to the most enriched cluster for the distortion function clustering population (left) and the EM objective function clustering population (right). Stars show positively annotated genes for the corresponding label. The consensus method is IPVC for both of the populations. The enrichment for the two clusters is respectively 4.58 and 4.80.

Clustering maps allow to intuitively represent the space of clustering solutions for a given dataset, independently from the algorithm producing the clusterings.

2.2 Visualization of Validation

In section 1.4 we described the validation step, that is a non-visual process. Here we describe how the results of such process can be visualized. We are interested in analyzing mainly two aspects of the validation phase: where is located the consensus solution containing the most enriched clusters on the meta-clustering dendrogram and where the points belonging to such clusters are localized in the data space. Our tools visualize such information in a simple way. For the former we highlight the node of the meta-clustering tree that is the root of the best subtree. This gives important insight on the usefulness of the whole process: if the best solution was found on the root of the tree, than meta clustering would not be necessary (the best solution is obtained applying consensus to all the solutions); if the best solution was found on a leaf of the tree, than both meta clustering and consensus would not be necessary (the best solution is just one of the starting solutions). Information about localization of the most enriched clusters inside the data space is shown through a simple 2D MDS visualization: we plot all the points using a different color for those belonging to the cluster and a different shape for the points belonging to the class whose enrichment is high for that cluster.

3 Application Example

As an example, we applied our framework to the well known HeLa-cell dataset [22,17]. In this context, we are not interested in actual biological findings. We chose the HeLa dataset just for its complexity, due to its size and ambiguity from the standpoint of clustering. The dataset we consider is an elaboration of the HeLa cell dataset, as explained in [17,13]. The dataset is composed by 1099 genes with 10 values for each gene (as resulting from a non-linear PCA reduction of 48 expression values). As prior knowledge, we used two labellings obtained from the Gene Ontology (GO) database, namely Cell Cycle, Cell Proliferation and Transport. We performed the exhaustive application of 8 consensus methodologies to all the nodes of the meta clustering dendrogram. Some results are shown in Fig. 3 and 4.

4 Conclusions

In this paper we showed how Meta Clustering, Consensus Clustering and prior knowledge can be exploited to extract well characterized groups of data. We showed how a systematic approach to clustering can help in extensively searching the space of clustering solutions towards a better understanding of data structure. Superimposition of prior knowledge can help in both validating clusters and formulate hypotheses about missing information. We developed Matlab visual and interactive tools that simplify and speed up the entire process. Most of the tools are integrated in a modular environment called MIDA [23].

References

1. Amato, R., Ciaramella, A., Deniskina, N., et al.: A Multi-Step Approach to Time Series Analysis and Gene Expression Clustering. Bioinformatics 22(5), 589–596 (1995)
2. Barthélemy, J.P., Leclerc, B.: The median procedure for partitions. In: Cox, I.J., Hansen, P., Julesz, B. (eds.) Partitioning Data Sets, American Mathematical Society, Providence, RI, pp. 3–34 (1995)
3. Ben-Hur, A., Elisseeff, A., Guyon, I.: A stability based method for discovering structure in clustered data. In: Pacific Symposium on Biocomputing, vol. 7, pp. 6–17 (2002)
4. Bertolacci, M., Wirth, A.: Are approximation algorithms for consensus clustering worthwhile? In: 7th SIAM International Conference on Data Mining, pp. 437–442 (2007)
5. Bertoni, A., Valentini, G.: Random projections for assessing gene expression cluster stability. In: Proceedings IEEE International Joint Conference on Neural Networks, vol. 1, pp. 149–154 (2005)
6. Bifulco, I., Fedullo, C., Napolitano, F., Raiconi, G., Tagliaferri, R.: Robust Clustering by Aggregation and Intersection Methods. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part III. LNCS, vol. 5179, pp. 732–739. Springer, Heidelberg (2008)

7. Bifulco, I., Murino, L., Napolitano, F., Raiconi, G., Tagliaferri, R.: Using Global Optimization to Explore Multiple Solutions of Clustering Problems. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part III. LNCS, vol. 5179, pp. 724–731. Springer, Heidelberg (2008)
8. Bishehsari, F., Mahdavinia, M., Malekzadeh, R., Mariani-Costantini, R., Miele, G., Napolitano, F., Raiconi, G., Tagliaferri, R., Verginelli, F.: PCA based feature selection applied to the analysis of the international variation in diet. In: Masulli, F., Mitra, S., Pasi, G. (eds.) WILF 2007. LNCS, vol. 4578, pp. 551–556. Springer, Heidelberg (2007)
9. Brachetti, P., De Felice Ciccoli, M., Di Pillo, G., Lucidi, S.: A new version of the Price's algorithm for global optimization. Journal of Global Optimization 10, 165–184 (1997)
10. Bresco, M., Raiconi, G., Barone, F., De Rosa, R., Milano, L.: Genetic approach helps to speed classical Price algorithm for global optimization. Soft Computing Journal 9, 525–535 (2005)
11. Nguyen, N., Caruana, R.: Consensus Clustering. In: Perner, P. (ed.) ICDM 2007. LNCS, vol. 4597, pp. 607–612. Springer, Heidelberg (2007)
12. Caruana, R., Elhawary, M., Nguyen, N., Smith, C.: Meta Clustering. In: Perner, P. (ed.) ICDM 2006. LNCS, vol. 4065, pp. 107–118. Springer, Heidelberg (2006)
13. Ciaramella, A., Cocozza, S., Iorio, F., Miele, G., Napolitano, F., Pinelli, M., Raiconi, G., Tagliaferri, R.: Interactive data analysis and clustering of genomic data. Neural Networks 21, 368–378 (2008)
14. Gionis, A., Mannila, H., Tsaparas, P.: Clustering aggregation. ACM Trans. Knowl. Discov. Data 1 (1 article 4) (2007)
15. Kerr, M.K., Churchill, G.A.: Bootstrapping cluster analysis: Assessing the reliability of conclusions from microarray experiments. PNAS 98, 8961–8965 (2001)
16. Kuncheva, L.I., Vetrov, D.P.: Evaluation of Stability of k-Means Cluster Ensembles with Respect to Random Initialization. PAMI 28(11), 1798–1808 (2006)
17. Napolitano, F., Raiconi, G., Tagliaferri, R., Ciaramella, A., Staiano, A., Miele, A.: Clustering and visualization approaches for human cell cycle gene expression data analysis. International Journal Of Approximate Reasoning 47(1), 70–84 (2008)
18. Price, W.L.: Global optimization by controlled random search. Journal of Optimization Theory and Applications 55, 333–348 (1983)
19. Strehl, A., Ghosh, J.: Cluster ensembles - a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research 3, 583–617 (2002)
20. Valentini, G., Ruffino, F.: Characterization Of Lung Tumor Subtypes Through Gene Expression Cluster Validity Assessment. RAIRO-Inf. Theor. Appl. 40, 163–176 (2006)
21. Xui, R., Wunsch, D.: Survey of clustering algorithms. IEEE Transactions on Neural Networks 16(3), 645–678 (2005)
22. Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., Botstein, D.: Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. Molecular Biology of the Cell 13, 1977–2000 (2002)
23. MIDA software, NeuRoNe lab, DMI, University of Salerno, http://www.neuronelab.dmi.unisa.it

Neuro-Fuzzy Approach for Reconstructing Fissures in Concrete's Reinforcing Bars

Matteo Cacciola, Giuseppe Megali, Diego Pellicanò, Michele Buonsanti, Salvatore Calcagno, Mario Versaci, and Francesco C. Morabito

University Mediterranea of Reggio Calabria
Via Graziella Feo di Vito, I-89100 Reggio Calabria, Italy
{matteo.cacciola,giuseppe.megali,diego.pellicano,michele.buonsanti,
calcagno,mario.versaci,morabito}@unirc.it
http://www.ing.unirc.it

Abstract. In concrete, metallic bars are used to reinforce the mechanic resistance of the structures. When the structural element is subject to strong traction stresses, the main efforts load just on the bars. Thus, they are mainly subject to problems of ruptures within the concrete. Therefore, a very useful application of Non Destructive Testing could be the implementation of decisional tool for characterizing the status of the bars, and the eventually existing breaks and cracks. This relevant inverse problem is solved by means of a system which extracts information on the specimen under test from the measurements and implements a priori constraints to facilitate the detection of defect, if any. A Neuro-Fuzzy approach is proposed in this paper to locate defects on reinforcing bars in concrete specimens applying eddy current-based measurements. The method exploits the concepts of fuzzy inference to localize and estimate the defect. A comparison with Neural Network estimators is presented.

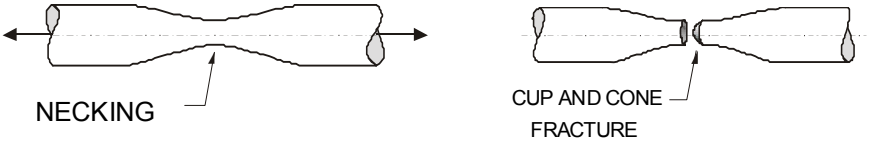
Keywords: Slim Structures, Eddy-current Non Destructive Tests, Neuro-Fuzzy Systems, Non Linear Systems.

1 Introduction

Non Destructive Testing and Evaluation (NDT/E) are more and more utilized to quickly and cheaply recognize and identify flaws within the inspected materials with special regard to those sectors where the integrity of the material is strictly required. This is the case of civil engineering, where the safety of buildings is related to the specific purposes (e.g. private houses, hospitals, stations and soon). Anyway, the general requirement is a good state of the structural integrity, in order to have no sudden risk of collapses. In concrete structures the rule of the reinforcing bars stands out when the structural element is subject to strong traction stresses, since concrete cannot withstand these strains but only compression stresses. Several theoretical model have been developed by researchers on thin reinforced concrete structures [1,2] with the aim to reproduce with adequate reliability the rising and spreading of fissuring, and the

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 171–178, 2009.

loads leading to collapses. In our work, in-lab experimentations have been carried out on a defectless reinforced concrete specimen, which underwent to rising traction up to the fissuring. After each traction, the specimen has been analyzed by the well-known eddy current (EC) NDT/E. The obtained maps of magnetic field distribution constitute an electromagnetic representation of the tensional condition of the specimen at the different values of the applied traction [3]. In this way, it has been possible to collect a suitable experimental database, useful to extract inferences for assessing the growth of defects in reinforced concrete elements similar to the exploited specimen. An heuristic approach has been exploited in order to characterize such kind of flaw in bars, thus regularizing the inverse problem [4].



(a) Typical necking occurring in a steel beam after the elastic behavior and before the cracking

(b) Graphical example of a cracked steel beam

Fig. 1. A sketch of the behavior of necking and cracking in steel beams

2 Mechanical Behavior of Reinforced Concrete under Fatigue Stress

Considering the Cervenska's model [5], possible slots can appear and develop in a normal direction than the main traction's stress. The behavior of the reinforcement steel is considered elastic-perfectly plastic; after the yielding the perfect plasticity is imposed. In purely membrane condition, the stress into the concrete is zero, while it is maximum inside the rods. Therefore, limitations in medium main traction stress should be introduced. These are due to loading capacity limits subsequently the plasticization of the steel. If the critical limit is exceeded due to further load increases, the final strength relies on the integrity of the tight rods. The adherence between steel and concrete is expressed according to the following well-known relationship, in case of planar and lateral traction, $\mathbf{F} = \sigma_s \cdot A_s = \tau \cdot p \cdot l$. Here $\mathbf{F}$ is the applied mechanical force, A_s is the steel area, σ_s is the tension along the external section, l is the length of the bar, τ is the adherence tension, p is the adherence perimeter. If the yielding of the rod occurs, the Hooke's law cannot be applied anymore [4]. In particular, when traction takes place inside the structural element, fracture phenomena arise inducing, next to the cracks, a widespread tensional state, at the same time with zero traction inside the concrete. Consequently, reinforcing bars are subjected to stretching, with a corresponding increase of their tensional state. In order to ensure structural safety, the bars must withstand the traction due to the fracture

without overcoming neither their yielding limit nor exhibiting high plasticity. In modern structures these aspects are rarely involved, whereas the problem has a major impact when the integrity of old decaying structural elements is taken into consideration. In order to understand how much the full integrity of the bars in reinforced concrete is concerned to guarantee in-service structural safety, just consider oxidation-reduction phenomena, imperfect soldering of the bars, very reduced adherence due to smooth bars. Both during the construction and inspection after lengthy in-service periods the need comes out to get info on the structural integrity of the concerned element without damaging it. Up-to-date technologies present remarkable procedural methodologies to properly evaluate the state of health of structural elements. But most of them are highly invasive, often leading to the destruction of the element itself.

3 Evaluation of Fatigue Cracked Reinforced Concrete with Fuzzy Based Systems

In order to improve manufacturing quality and ensure public safety, components and structures are commonly inspected for early detection of defects. In experimental NDT/E, the available measurement data are explored in order to some clues may emerge, with the advantage of leaving the specimen undamaged after the inspection [5,6]. In our in-study case, the concrete surrounding the bars doesn't introduce any disturbing noise, being totally transparent to the employed electromagnetic fields. If the microscopic structure of the defect has no relevance, the aim of the reconstruction problem being to characterize the defect. An heuristic approach could be helpful in solving this kind of ill-posed inverse problems (since the nonlinearity of the direct electromagnetic problem) [7], in a sort of learning by sample paradigm. In particular, we exploited the concept of Neuro-Fuzzy Systems (NFSs) [8]. NFS models are basically feed-forward networks that use a Fuzzy Inference System (FIS) [9] as a first guess model of the underlying dynamic process and then tune the initial choice of the model's parameters according with the available input-output pairs. FISs allows us to treat and exploit uncertainty [9]. Thus the inputs of the procedure are interpreted as fuzzy variables, characterized by Fuzzy Membership Functions (FMFs). FMFs are overlapped to allow an input to be distributed across a number of rules, giving rise to an interpolation effect. Typical FMFs are continuous, monotonic and piecewise-differentiable functions, such as the commonly used trapezoidal or triangular-shaped functions. However, the FIS can be useful as a first guess model, whose performances can be improved both using an algorithm of automatic extraction of FIS [7] from numerical data and introducing learning [10]. According to the previously proposed sketch of the FIS and NFIS theory, the identification problem can be formulated as the search of a suitable mapping between the set of available measurements and the selected set of parameters. Our goal was to associate the position and the entity of the defect in the bar to a particular pattern of measurements.

3.1 The Experimental Database

Non destructive damage analysis of slim structures in thin walls is an aspect of the structural mechanics, which offers interesting perspectives of practical relevance. NDT/E has been carried out by exciting currents over the specimen, thus inducing eddy currents inside it and finally picking-up by a FLUXSET© sensor [6] a signal which is a measure of the variations of the overall magnetic field, including the contribution due to the presence of a defect. The FLUXSET© sensor consists of three coils: exciting coil (to induce eddy currents in the sample), driving coil (to saturate the FLUXSET© core material), pick-up coil (the sensor strictly speaking). A hard device is able and used to convert the magnetic variations to electric variations [11], so providing in output a voltage signal, called pick-up voltage signal (or more easily pick-up voltage). The sample was a concrete block, with reinforcement steel bar located along the longitudinal axis (Fig. 2). An artificial cut, roughly two third of the cross section, was done in the rod [5]. According to the geometry of the experimental set-up, the distance between the sensor and the rod is approximately 10 mm. Phase demodulation of the pick-up voltage gives a measurement proportional to the magnetic field parallel to the plane of the sample. The specimen has been investigated with a 10 Vpp sinusoidal signal at a frequency of 1.022 kHz, suitable to explore the whole cross section of the bar. The sensor was moved over the block by means of a 0.5 mm step-by-step automatic scanning, along x and y axes, in order to map the area neighboring the cut. Figs. 3(a) and 3(b) clearly show the presence of the cut, which exhibits an inclination with respect to the axis of the rod itself.

The collected database is composed by: 10 ECT maps characterized by presence of defect and 10 ECT maps showing absence of defect, with different tensional states and analyzed with different current values. Each flawed specimen has an unflawed correspondent specimen, and thus it has been possible to normalize the collected data. Subsequently, the dataset has been split in training and test subsets. In order to set the amount of training signals, we made a trade-off between the requirements of an as large as possible training subset and a significant availability of testing signals. Thus, in our experimentations, training set has been composed by randomly selected areas taken from 7 ECT maps with defect's presence and 7 ECT maps of unflawed specimens. Remaining maps compose the test subset. Inputs system are: normalized impedance $\tilde{Z} = \frac{\hat{Z}_i}{Z_{0,i}}$ where

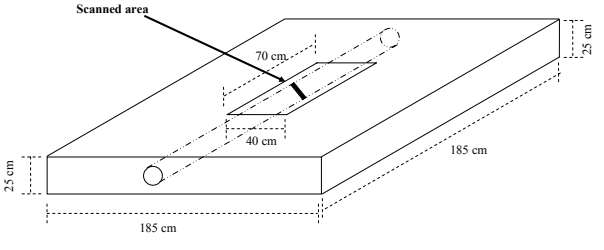
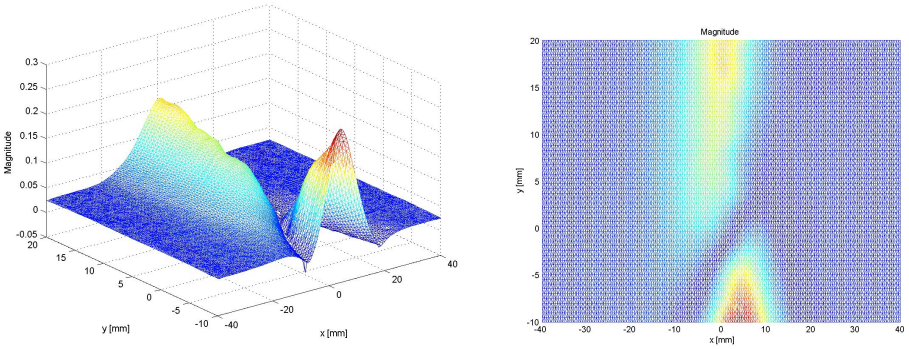


Fig. 2. Sketch of the investigated specimen



(a) Cross section view (b) Front view

Fig. 3. Magnitude of the pick-up voltage

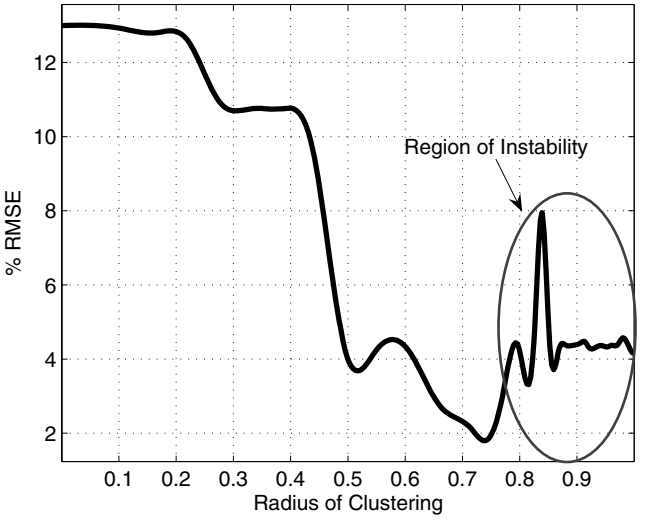


Fig. 4. Graphic depiction of system performances according to the variation of the %RMSE values

$\hat{Z}_i$ represents the impedance of the i-th point on computed map of specimen with defect and $Z_{0,i}$ the corresponding value for the specimen without crack; $\mu_{ratio} = \frac{\hat{\mu}_i^{10x10}}{\mu_i^{10x10}}$, $\sigma_{ratio} = \frac{\hat{\sigma}_i^{10x10}}{\sigma_i^{10x10}}$, $S_{ratio} = \frac{\hat{S}_i^{10x10}}{S_i^{10x10}}$, $K_{ratio} = \frac{\hat{K}_i^{10x10}}{K_i^{10x10}}$, where $\hat{\mu}_i^{10x10}$, $\hat{\sigma}_i^{10x10}$, $\hat{S}_i^{10x10}$, $\hat{K}_i^{10x10}$ are the average, standard deviation, skewness and kurtosis of a 10x10 mm squared centered on the i-th pixel of flawed specimen, and μ_i^{10x10}, σ_i^{10x10}, S_i^{10x10}, K_i^{10x10} are the corresponding values on the undamaged specimen. Voltage signal ($|V|$) represents the FIS' output.

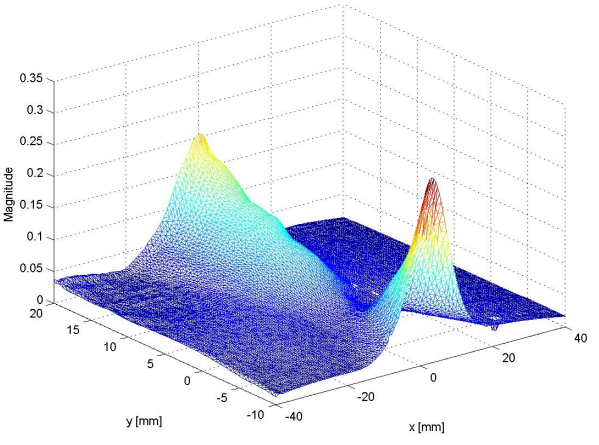


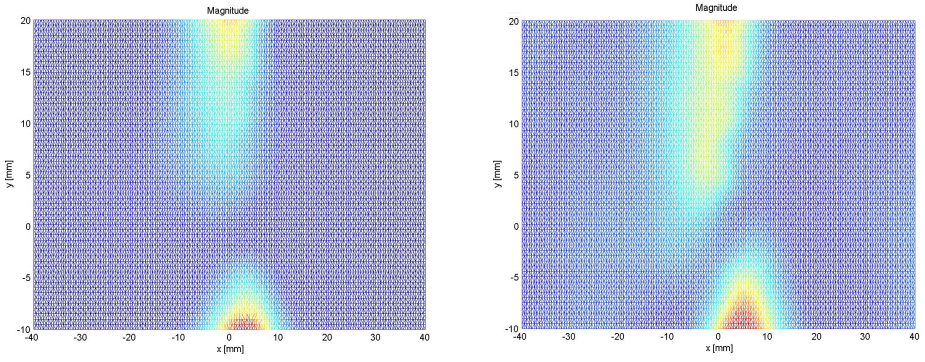
Fig. 5. Three-dimensional view of the magnitude of the pick-up voltage reconstructed by a NFS with a three rules FIS

3.2 Comparison between Experimental and Numerical Results

In order to increase the flexibility of the model, Gaussian FMFs were run in our work. The procedure to design a FIS is usually the following: fuzzification of the input-output variables; fuzzy inference through the bank of fuzzy rules; defuzzification of the fuzzy output variables. Artisan-type fuzzy behaviour rules are usually obtained from a simple visual analysis of the single bi-dimensional plots, relevant to input and input-output pairs. MATLAB® GENFIS system [10,12] generates fuzzy rules with multiple antecedents in number equal to the number of the inputs. FMF labeling has been automatically done in growing numbers. Performances have been evaluated by calculating Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(x_{ij} - \hat{x}_{ij}\right)^2}{N \cdot M}} \tag{1}$$

where x_{ij} and $\hat{x}_{ij}$ are the observed and estimated values for the point at coordinates (i, j) on the inspected specimen, respectively. Generally, obtained results are very encouraging, with an average performances of about 98.2%. The lowest %RMSE value is equal to 1.8%, corresponding to a clustering radius equal to 0.74 (three rules). Thus, several FISs have been carried out to tune the system, by proposing them as the starting point of NFS approach. In this way, the MATLAB® ANFIS tool, starting from the generated FIS, allowed us to evaluate the performances of such a kind of regressors. Maps in Figs. 5 and 6(a) correspond to maps shown in Figs. 3(a) and 3(b), respectively. They were obtained applying the Neuro-Fuzzy Inference to the experimental database with radius of subtractive clustering [13] equal to 0.74. What can be seen is that the presence of the cut clearly stands up, but the info relevant to its inclination is lost. By using the same FIS, but adding three more rules (which means a computational



(a) Front view (three rules FIS) (b) Front view (six rules FIS)

Fig. 6. Magnitude of the pick-up voltage reconstructed by a NFS

increase of third order) an outcome NFS comparable to Fig. 3(b) was achieved (see Fig. 6(b)).

4 Conclusions

The proposed approach offers the possibility to build the model (i.e. to determine the rules) both by a-priori knowledge of the system under investigation and learning evidence. The resulting system has a "readable" structure given by the FIS rules, and the advantage of the neural learning adopted by the additional Neural knowledge, in contrast to the "black box" structure of simple Neural Networks. Performing a careful analysis of the experimental database it has been possible to significantly reduce the number of the inputs and the cardinality of the fuzzy data bank. This feature turns out to be decisive in practical real time applications, where the reduction of the inspection time is a major demand, and in the design of novel large systems, like nuclear fusion reactors, where an effort must be made to reduce diagnostic concerns. The NFS technique offers in addition a novel opportunity given to designers and technicians to solve the inverse problem by inspecting the rules which are automatically generated. In authors' opinion, the most remarkable conclusion of the work is that it is possible to design inference models with reduced computational complexity by NFSs. Since the presence of Neural paradigm, they are able to estimate defects in terms of their position and size, and the attained accuracy has been improved by adding qualitative knowledge on the problems in terms of fuzzy expert linguistic rules.

References

1. American Concrete Institute: Fracture Mechanics of Concrete: Concepts, Models and Determination of Material Properties. Technical report, American Concrete Institute (1991)
2. Bazant, Z.P., Planas, J.: Fracture and Size Effect. CRC Press, New York (1998)

3. Cacciola, M., La Foresta, F., Morabito, F.C., Versaci, M.: Advanced Use of Soft Computing and Eddy Current Test to Evaluate Mechanical Integrity of Metallic Plates. NDT&E International 40(5), 357–362 (2007)
4. Morabito, F.C., Versaci, M.: A fuzzy neural approach to localizing holes in conducting plates. IEEE Transaction on Magnetics 37(5), 3534–3537 (2001)
5. Chady, T., Enokizono, M., Takeuchi, K., Kinoshita, T., Sikora, R.: Eddy current testing of concrete structures. In: Takagi, T., Uesaka, M. (eds.) Applied Electromagnetics and Mechanics. JSAEM Studies, pp. 509–510. IOS Press, Amsterdam (2001)
6. Gasparics, A., Daroczi, C.S., Vertesy, G., Pavo, J.: Improvement of ECT probes based on Fluxset type magnetic field sensor. In: Albanese, R., et al. (eds.) Electromagnetic Nondestructive Evaluation, vol. II, pp. 146–151. IOS Press, Amsterdam (1998)
7. Morabito, F.C., Coccorese, E.: A fuzzy modeling approach for the solution of an inverse electrostatic problem. IEEE Transaction on Magnetics 32(3), 1330–1333 (1996)
8. Abraham, A.: Adaptation of Fuzzy Inference System Using Neural Learning, Fuzzy System Engineering: Theory and Practice. In: Nedjah, N., et al. (eds.) Studies in Fuzziness and Soft Computing, ch. 3, pp. 53–83. Springer, Heidelberg (2005)
9. Zadeh, L.: Fuzzy logic and its application to approximate reasoning. In: Proceedings of the Information Processing (IFIP74), pp. 591–594 (1974)
10. Jang, R.: ANFIS: Adaptive-Network-based Fuzzy Inference Systems. IEEE Transaction on Systems, Man, and Cybernetics 23(3), 665–685 (1993)
11. Buonsanti, M., Calcagno, S., Morabito, F.C., Versaci, M.: Eddy Current and Fuzzy Inference to Control Defects Growth in Reinforced Concrete. Key engineering materials, 345–346, 1291–1294 (2007)
12. Ross, T.J.: Fuzzy Logic with engeneering applications, II edn. Wiley and Sons, New York (2001)
13. Chiu, S.: Fuzzy Model Identification Based on Cluster Estimation. Journal of Intelligent & Fuzzy Systems 2(3), 267–278 (1994)

Fuzzy Relational Calculus and Its Application to Image Processing

Etienne E. Kerre and Mike Nachtegael

Ghent University
Department of Applied Mathematics & Computer Science
Fuzziness & Uncertainty Modelling,
Krijgslaan 281 - S9, B-9000 Gent, Belgium
{etienne.kerre,mike.nachtegael}@ugent.be
http://www.fuzzy.ugent.be

Abstract. The calculus of relations has been very important during the past 40 years from theoretical as well as from practical point of view. The development of fuzzy set theory, particularly in the framework of relational calculus has substantially increased the interest in this domain of science. In this paper we will give a brief overview of the recent developments in crisp as well as in fuzzy relational calculus and illustrate its applicability in image processing.

1 New Concepts in Classical Relational Calculus

The concept of a relation is fundamental since only in a few steps one can introduce this concept in the framework of set theory. Indeed as soon as the meaning of the so-called classifier $\{z|P\}$, i.e., the class of all objects z that satisfy a given property P, has been introduced one may define consecutively a singleton, a doubleton, an ordered pair, the cartesian product of two sets and finally a relation from X to Y as a subset of the cartesian product $X \times Y$. This concept may be extended to a relation between n universes $X_1, X_2, \ldots, X_n$ as a subset of $X_1 \times X_2 \times \ldots \times X_n$. Some auxiliary notions with respect to a relation R from X to Y are:

- The domain $\mathrm{dom}(R)$ consisting of all elements of X that are linked by R to at least one element of Y.
- The range $\mathrm{rng}(R)$ consisting of all elements of Y that are linked to at least one element of X.
- The inverse relation R^{-1} consisting of all ordered pairs (y, x) such that (x, y) belongs to R.
- The R-afterset of x (denoted xR) consisting of all elements of Y that are linked to x.
- The R-foreset of y (denoted Ry) consisting of all elements of X that are linked to y.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 179–188, 2009.

Due to the last two notions introduced by Bandler and Kohout [1] in the 80's, a lot of new images and compositions could be introduced. These concepts substantially enlarge the toolkit of relational calculus.

Since relations are sets all set-theoretic operations such as union, intersection, complementation, difference and symmetric difference can be applied to relations. For example suppose that R_1 and R_2 are relations from X to Y, then the union $R_1 \cup R_2$ consists of all ordered pairs (x, y) that belong to R_1 or to R_2. It is interesting to note that all these operations can be directly applied to aftersets and foresets, i.e., the $(R_1 \cup R_2)$-afterset of $x \in X$ equals the union of xR_1 and xR_2, i.e., the family of aftersets $(xR)_{x\in X}$ contains all relevant information concerning the relation R and similarly for the family of foresets $(Ry)_{y\in Y}$.

Important notions in mathematics like continuity and measurability are based on the concept of direct and inverse image of a set under a (functional) relation.

Let R be a relation from X to Y, A a subset of X and B a subset of Y, then

- the direct image of A under R is given by:

$$\begin{aligned} R(A) &= \{y | (\exists x \in A)((x, y) \in R)\} \\ &= \{y | A \cap Ry \neq \emptyset\} \\ &= \bigcup_{x\in A} xR \end{aligned}$$

- the inverse image of B under R is given by:

$$\begin{aligned} R^{-1}(B) &= \{x | (\exists y \in B)((x, y) \in R)\} \\ &= \{x | B \cap xR \neq \emptyset\} \\ &= \bigcup_{y\in B} Ry \end{aligned}$$

Inspired by the work of Bandler-Kohout [1] on the new compositions, De Baets-Kerre [2]–[5] have introduced some new images that could be defined using after- and foresets:

- the subdirect image of A under R:

$$\begin{aligned} R^{\triangleleft}(A) &= \{y | A \cap Ry \neq \emptyset \quad and \quad A \subseteq Ry\} \\ &= \{y | A \neq \emptyset,\ Ry \neq \emptyset \quad and \quad A \subseteq Ry\} \end{aligned}$$

- the superdirect image of A under R:

$$\begin{aligned} R^{\triangleright}(A) &= \{y | A \cap Ry \neq \emptyset \quad and \quad Ry \subseteq A\} \\ &= \{y | A \neq \emptyset,\ Ry \neq \emptyset \quad and \quad Ry \subseteq A\} \end{aligned}$$

- the squaredirect image of A under R:

$$\begin{aligned} R^{\square}(A) &= \{y | A \cap Ry \neq \emptyset \quad and \quad A = Ry\} \\ &= \{y | A \neq \emptyset,\ Ry \neq \emptyset \quad and \quad A = Ry\} \end{aligned}$$

Finally we mention the most important operation on relations, namely composition or product and its useful extensions introduced by Bandler-Kohout [1] and afterwards modified by De Baets-Kerre [2]–[5].

Let R_1 be a relation from X to Y and R_2 a relation from Y to Z, then:

- the round product of R_1 and R_2 (read: R_1 before R_2, R_1 followed by R_2) is defined as the relation from X to Z given by:

$$\begin{aligned} R_1 \circ R_2 &= \{(x,z)|(\exists y)((x,y) \in R_1 \quad and \quad (y,z) \in R_2)\} \\ &= \{(x,z)|xR_1 \cap R_2 z \neq \emptyset\} \end{aligned}$$

- the subproduct of R_1 and R_2:

$$\begin{aligned} R_1 \triangleleft R_2 &= \{(x,z)|xR_1 \cap R_2 z \neq \emptyset \quad and \quad xR_1 \subseteq R_2 z\} \\ &= \{(x,z)|xR_1 \neq \emptyset,\ R_2 z \neq \emptyset \quad and \quad xR_1 \subseteq R_2 z\} \end{aligned}$$

- the superproduct of R_1 and R_2:

$$\begin{aligned} R_1 \triangleright R_2 &= \{(x,z)|xR_1 \cap R_2 z \neq \emptyset \quad and \quad R_2 z \subseteq xR_1\} \\ &= \{(x,z)|xR_1 \neq \emptyset,\ R_2 z \neq \emptyset \quad and \quad R_2 z \subseteq xR_1\} \end{aligned}$$

- the squareproduct of R_1 and R_2:

$$\begin{aligned} R_1 \square R_2 &= \{(x,z)|xR_1 \cap R_2 z \neq \emptyset \quad and \quad xR_1 = R_2 z\} \\ &= \{(x,z)|xR_1 \neq \emptyset,\ R_2 z \neq \emptyset \quad and \quad xR_1 = R_2 z\} \end{aligned}$$

2 A Brief Outline of Fuzzy Relational Calculus

Since the old Greeks, scientists have recognized that binary or black-or-white logic is not sufficient to model our knowledge which is mostly pervaded with imprecision. We have to wait until 1965 when Lotfi Zadeh introduced the concept of a fuzzy set in his seminal paper entitled "Fuzzy Sets",in order to model imprecise terms as "sets" with unsharp boundaries where the transition from belonging to not belonging is rather gradual than abrupt. In the same spirit Zadeh introduced the concept of a fuzzy relation from a universe X to a universe Y as a fuzzy set R in the cartesian product $X \times Y$ where $R(x,y)$ denotes the strength of relationship between $x \in X$ and $y \in Y$.

More formally a fuzzy relation R from X to Y is a mapping from $X \times Y$ into the unit interval $[0,1]$, attaching to every ordered pair (x,y) in $X \times Y$ a degree of relationship $R(x,y)$ belonging to $[0,1]$.

The basic concepts introduced in section 1 can be generalized or fuzzified as follows. Let R be a fuzzy relation from X to Y, then:

- the domain of R is a fuzzy set in X given by:

$$\mathrm{dom}(R)(x) = \sup\{R(x,y)|y \in Y\}, \forall x \in X$$

- the range of R is a fuzzy set in Y given by:

$$\mathrm{rng}(R)(y) = \sup\{R(x,y)|x \in X\}, \forall y \in Y$$

– the inverse R^{-1} of R is the fuzzy relation from Y to X given by:

$$R^{-1}(y,x) = R(x,y), \forall (y,x) \in Y \times X$$

– the R-afterset of $x \in X$ is the fuzzy set in Y given by:

$$xR(y) = R(x,y), \forall y \in Y$$

– the R-foreset of $y \in Y$ is the fuzzy set in X given by:

$$Ry(x) = R(x,y), \forall x \in X$$

All the set-theoretic operations have been extended in an infinite number of ways to fuzzy sets and a fortiori to fuzzy relations using the concepts of triangular norms T and conorms S introduced by Schweizer-Sklar in the framework of probabilistic metric spaces. The T-intersection (S-union) of two fuzzy relations R_1 and R_2 from X to Y is defined as a fuzzy relation from X to Y given as:

$$R_1 \cap_T R_2(x,y) = T(R_1(x,y), R_2(x,y))$$

$$R_1 \cup_S R_2(x,y) = S(R_1(x,y), R_2(x,y))$$

for all $(x,y) \in X \times Y$.

In order to fuzzify the concepts of images and compositions we need an extension of the classical intersection or conjunction operation and the binary implication to model the inclusion. As said before the intersection of two fuzzy sets in some universe may be modelled by a triangular norm T. In this way we obtain for the T-direct image of a fuzzy set A in X under a fuzzy relation R from X to Y:

$$\begin{aligned} R_T(A) : Y &\to [0,1] \\ y &\mapsto \sup_{x \in X} T(A(x), R(x,y)), \forall y \in Y \end{aligned}$$

Similarly for the T-round composition of a fuzzy relation R_1 from X to Y followed by a fuzzy relation R_2 from Y to Z we obtain:

$$\begin{aligned} R_1 \circ_T R_2 : X \times Z &\to [0,1] \\ (x,z) &\mapsto \sup_{y \in Y} T(R_1(x,y), R_2(y,z)), \forall (x,z) \in X \times Z \end{aligned}$$

A fuzzy implication is defined as a $[0,1]^2 - [0,1]$ mapping I satisfying the boundary conditions: $I(0,0) = I(0,1) = I(1,1)$ and $I(1,0) = 0$. Putting extra conditions such as hybrid monotonicity, neutrality principle and exchange principle leads to more specific implication operators. Some popular operators are $I_{KD}(x,y) = \max(1-x,y)$ (Kleene-Dienes implication), $I_L(x,y) = \min(1, 1-x+y)$ (Lukasiewicz implication) and $I_R(x,y) = 1 - x + xy$ (Reichenbach implication).

For more recent developments on fuzzy implication operators such as S-implications, R-implications and QL-implications we refer to [6,7].

Let R be a fuzzy relation from X to Y, A a fuzzy set in X, T a triangular norm and I a fuzzy implication. Then we define the following extensions of the triangular images:

– the T-I subdirect image of A under R as:

$$R^{\triangleleft}_{T,I}(A) : Y \to [0,1]$$
$$y \mapsto \min(\sup_{x\in X} T(A(x), R(x,y)), \inf_{x\in X} I(A(x), R(x,y))), \forall y \in Y$$

– the T-I superdirect image of A under R as:

$$R^{\triangleright}_{T,I}(A) : Y \to [0,1]$$
$$y \mapsto \min(\sup_{x\in X} T(A(x), R(x,y)), \inf_{x\in X} I(R(x,y), A(x))), \forall y \in Y$$

– the T-I squaredirect image of A under R as:

$$R^{\square}_{T,I}(A) : Y \to [0,1]$$
$$y \mapsto \min(R^{\triangleleft}_{T,I}(A)(y), R^{\triangleright}_{T,I}(A)(y)), \forall y \in Y$$

Finally let us fuzzify the new compositions of fuzzy relations. Let R_1 be a fuzzy relation from X to Y, R_2 a fuzzy relation from Y to Z, T a triangular norm and I a fuzzy implication. Then we define:

– the T-I subproduct of R_1 and R_2 as:

$$R_1 \triangleleft_{T,I} R_2 : X \times Z \to [0,1]$$
$$(x,z) \mapsto \min(\sup_{y\in Y} T(R_1(x,y), R_2(y,z)), \inf_{y\in Y} I(R_1(x,y), R_2(y,z))),$$
$$\forall (x,z) \in X \times Z$$

– the T-I superproduct of R_1 and R_2 as:

$$R_1 \triangleright_{T,I} R_2 : X \times Z \to [0,1]$$
$$(x,z) \mapsto \min(\sup_{y\in Y} T(R_1(x,y), R_2(y,z)), \inf_{y\in Y} I(R_2(y,z), R_1(x,y))),$$
$$\forall (x,z) \in X \times Z$$

– the T-I squareproduct of R_1 and R_2 as:

$$R_1 \square_{T,I} R_2 : X \times Z \to [0,1]$$
$$(x,z) \mapsto \min(R_1 \triangleleft_{T,I} R_2(x,z), R_1 \triangleright_{T,I} R_2(x,z)), \forall (x,z) \in X \times Z$$

For more detailed information about the basics of fuzzy relational calculus and some of its extensions such as to intuitionistic fuzzy set theory and rough set theory we refer to [8]–[18].

3 Application to Image Processing

Also from application point of view the concept of a relation is a very fundamental one. More precisely, relations and their fuzzification have been successfully applied in relational databases, information retrieval, approximate reasoning,

preference modelling, ordering techniques, medical diagnosis, modelling temporal and spatial information. For a short overview of these applications, as well as an extensive list of references for more detailed information, we refer to [19].

In this section we focus on the application of fuzzy relational calculus in image processing. The field of image processing has led to many interesting theories and corresponding practical algorithms that can extract specific information from images (e.g. edges, patterns, ...), improve their quality (e.g. denoising, deblurring, ...) or prepare them for other applications (e.g. compression). Mathematical morphology is one of the theories which offers a wide range of tools for image processing and computer vision. The basic morphological operators dilation and erosion constitute the fundamentals of this theory [20]. A morphological operator transforms an image into another image, using a structuring element (usually chosen by the user).

Binary morphology was developed to process binary images. From a mathematical point of view, such images can be modeled as $X - \{0,1\}$ mappings (i.e., as crisp subsets of X), with X the considered universe (usually a finite subset of $\mathbb{R} \times \mathbb{R}$) and $\{0,1\}$ the set of possible values (0 representing black and 1 representing white). The dilation $D(A,S)$ and erosion $E(A,S)$ of a binary image A using a structuring element S is defined as follows:

$$D(A,S) = \{y | y \in X \text{ and } T_y(S) \cap A \neq \emptyset\},$$
$$E(A,S) = \{y | y \in X \text{ and } T_y(S) \subseteq A\},$$

with $T_y(S) = \{x | x \in X \text{ and } x - y \in S\}$ the translation of S by the point y.

The binary dilation and erosion have a very nice geometrical interpretation. The dilation $D(A,S)$ consists of all points y in X such that the translation $T_y(S)$ of the structuring element has a non-empty intersection with the image A. Consequently, the dilation will typically extend the contours of objects in the image, and fill up small gaps and channels. The erosion $E(A,S)$ consists of all points y in X such that the translation $T_y(S)$ of the structuring element completely fits within the image A. Consequently, the erosion will typically reduce (erode) the contours of objects in the image, and delete small gaps and channels.

The original binary morphology, which was developed for binary images, was soon extended to grayscale morphology for grayscale images. Such images are modeled as $X - [0,1]$ mappings, with $[0,1]$ the interval of possible values (0 representing black, 1 representing white, and intermediate values representing intermediate shades of gray). The first extensions can be divided in two categories: either using the threshold approach [20] (in which grayscale images are transformed using binary structuring elements), or either using the umbra approach [21] (in which grayscale images are transformed using grayscale structuring elements).

Later, different models based on fuzzy set theory were introduced and studied [22]–[27]. These models were inspired by the observation that grayscale images and fuzzy sets can be modeled in the same way, namely as mappings from a universe X into the unit interval [0,1], i.e., as fuzzy subsets of X. This formal resemblance between these two different notions (fuzzy sets versus grayscale images) allows us to use techniques from fuzzy set theory and to apply them in

an image processing context. Using a triangular norm T and a fuzzy implication I, the binary dilation and erosion can be fuzzified as follows:

$$D_T(A,S)(y) = \sup_{x\in X} T(S(x-y), A(x)),$$
$$E_I(A,S)(y) = \inf_{x\in X} I(S(x-y), A(x)),$$

for all y in X. This fuzzification is straightforward, and based on the extension of set-theoretic notions from a binary to a fuzzy context.

At this point the reader surely will detect some similarity between the expressions of the fuzzy dilation $D_T(A,S)$ (with A and S grayscale images, i.e., fuzzy sets) and of the T-direct image $R_T(A)$ (with A a fuzzy set and R a fuzzy relation). This can be formalized as follows: let V denote the subtraction in X, i.e., $V(x,y) = x - y$, for all x and y in X, and let A and S be fuzzy sets in X. We can then define the fuzzy set $R = S \circ V$, which is a binary fuzzy relation in X. It easily follows that $D_T(A,S) = R_T(A)$ with $R = S \circ V$. In other words: the fuzzy dilation of A can be regarded as the T-direct image of A under a specific relation that is defined using the structuring element of the dilation.

A similar statement can be made for the fuzzy erosion, provided that a weak condition is imposed on the structuring element (i.e., on one of the images) and on the fuzzy implication (i.e., on one of the fuzzy logical operators). In many image processing applications it will be required that $S(0) = 1$, i.e. the origin completely belongs to the structuring element. This condition implies that the relation $R = S \circ V$ is a reflexive binary fuzzy relation in X. Furthermore it will be required that the fuzzy implication I is a fuzzy border implication, i.e., I has to satisfy the neutrality principle: $I(1,x) = x$ for all x in $[0,1]$. Under these conditions it can be shown that $E_I(A,S) = R^{\triangleright}_{T,I}(A)$ with again $R = S \circ V$. In other words: the fuzzy erosion of A can be regarded as the T-I superdirect image of A under a specific relation that is defined using the structuring element of the erosion. The condition on the structuring element will cause that the part with the triangular norm in the definition of $R^{\triangleright}_{T,I}$ becomes irrelevant, so that only the part with the fuzzy implication remains.

To summarize we have that:

$$D_T(A,S) = R_T(A),$$
$$E_I(A,S) = R^{\triangleright}_{T,I}(A),$$

with T a triangular norm, I a fuzzy border implication, $R = S \circ V$ and the latter equivalency under the restriction that $S(0) = 1$. Note that the left hand side of these equalities are operators that find their origin in the field of image processing, while the right hand side of these equalities are concepts from fuzzy relational calculus. Although this concerns two different fields, these equivalencies enable us to derive properties in one field (e.g., mathematical morphology in image processing) from properties in the other field (e.g., fuzzy relational calculus). We illustrate this with the properties of monotonicity, expansiveness and restrictiveness, and interaction with union. First we give the properties in a relational context, and next we interpret these properties in the context of mathematical morphology.

Property 1. *Let R be a binary fuzzy relation in X. $R_T(A)$ is increasing w.r.t. T and $R^{\triangleright}_{T,I}(A)$ is increasing w.r.t. T and I; i.e. if $T_1 \leq T_2$ and $I_1 \leq I_2$, then:*

$$R_{T_1}(A) \subseteq R_{T_2}(A)$$
$$R^{\triangleright}_{T_1,I_1}(A) \subseteq R^{\triangleright}_{T_2,I_2}(A).$$

$R_T(A)$ is increasing w.r.t. R; i.e. if $R_1 \subseteq R_2$, then:

$$(R_1)_T(A) \subseteq (R_2)_T(A).$$

$R_T(A)$ and $R^{\triangleright}_{T,I}(A)$ are increasing w.r.t. A; i.e. if $A_1 \subseteq A_2$, then:

$$R_T(A_1) \subseteq R_T(A_2)$$
$$R^{\triangleright}_{T,I}(A_1) \subseteq R^{\triangleright}_{T,I}(A_2).$$

The first property expresses that we can strenghten or weaken the direct images by adjusting the logical operators. In the context of image processing this implies that we can manipulate the effect of the fuzzy dilation and erosion by choosing larger or smaller triangular norms and/or fuzzy implications. For example, if $T_1 \leq T_2$ then we know that the dilated image $D_{T_1}(A, S)$ will be smaller than the dilated image $D_{T_2}(A, S)$. "Smaller" in this context means that contours of objects in the image will be extended to a lesser degree. The two other properties reflect the influence of a larger structuring element or a larger image. For example, if $A_1 \subseteq A_2$ then the latter property leads to the conclusion that the fuzzy erosion $E_I(A_1, S)$ will be smaller than the fuzzy erosion $E_I(A_2, S)$.

Property 2. *Let R be a reflexive binary fuzzy relation in X and I a fuzzy border implication, then:*

$$R^{\triangleright}_{T,I}(A) \subseteq A \subseteq R_T(A)$$

This property shows that the fuzzy T-I superdirect image is smaller (i.e. more specific) than the original fuzzy set, but that the T-direct image is larger (i.e. less specific) than the original fuzzy set. In the context of image processing this property translates to the statement that the erosion of a grayscale image is contained in that grayscale image, and that this grayscale image is contained in its dilation. Note the requirement that R must be a reflexive fuzzy relation. In the context of image processing this is satisfied when $S(0) = 1$ (with $R = S \circ V$).

Property 3. *It holds that:*

$$R_T(A_1 \cup A_2) = R_T(A_1) \cup R_T(A_2)$$
$$R^{\triangleright}_{T,I}(A_1 \cup A_2) \supseteq R^{\triangleright}_{T,I}(A_1) \cap R^{\triangleright}_{T,I}(A_2).$$

The T-direct image of the union of two fuzzy sets can be expressed as the union of the fuzzy direct images of the separate fuzzy sets. For the superdirect image only a containment relation w.r.t. the intersection of the separate fuzzy direct images holds.

In the context of image processing, the above property implies that the dilation of the union of two grayscale images can be computed by considering the two separate dilations of each grayscale image. A similar property holds for the union of structuring elements. Both properties also have computational advantages: one can decompose a grayscale image or a structuring element into different standard objects, calculate the corresponding standard dilation, and then assemble these results to obtain the dilation of the original grayscale image by the original grayscale structuring element.

A more detailed overview and discussion can be found in [28]–[30].

References

1. Bandler, W., Kohout, L.: Fuzzy relational products as a tool for analysis and synthesis of the behavior of complex natural and artificial systems. In: Wang, P.P., Chang, S.K. (eds.) Fuzzy Sets Theory and Applications to Policy Analysis and Information Systems, pp. 341–367. Plenum Press, New York (1980)
2. De Baets, B., Kerre, E.E.: Triangular fuzzy relational compositions revisited. In: Bouchon, B., Valverde, L., Yager, R. (eds.) Uncertainty in Intelligent Systems, pp. 257–267. Elsevier, North Holland (1993)
3. De Baets, B., Kerre, E.E.: Some reflections on fuzzy relational compositions. In: Bouchon, B., Yager, R. (eds.) Proceedings IPMU 1992, Palma de Mallorca, pp. 251–254 (1992)
4. De Baets, B., Kerre, E.E.: A revision of Bandler-Kohout compositions of relations. Mathematica Pannonica 4, 59–78 (1993)
5. De Baets, B., Kerre, E.E.: Fuzzy relational compositions. Fuzzy Sets and Systems 60, 109–120 (1993)
6. Shi, Y., Ruan, D., Kerre, E.E.: On the characterization of fuzzy implications satisfying $I(x, y) = I(x, I(x, y))$. Information Sciences 177, 2954–2970 (2007)
7. Baczynski, M., Jayaram, B.: Fuzzy implications. Springer, Heidelberg (2008)
8. De Baets, B., Kerre, E.E.: On the supercomposition of subcontinuous and fuzzy subcontinuous mappings and related topics. The Journal of Fuzzy Mathematics 2, 17–34 (1994)
9. De Baets, B., Kerre, E.E.: The cutting of compositions. Fuzzy Sets and Systems 62, 295–309 (1994)
10. De Baets, B., Kerre, E.E.: A primer on solving fuzzy relational equations on the unit interval. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2, 205–225 (1994)
11. De Baets, B., Kerre, E.E.: Fuzzy relations and applications. In: Hawkes, P. (ed.) Advances in Electronics and Electron Physics, pp. 255–324. Academic Press, London (1994)
12. Wang, X., De Baets, B., Kerre, E.E.: A comparative study of similarity measures. Fuzzy Sets and Systems 73(2), 259–268 (1995)
13. Nachtegael, M., De Cock, M., Vanderweken, D., Kerre, E.E.: Fuzzy relational images in computer science. In: de Swart, H. (ed.) RelMiCS 2001. LNCS, vol. 2561, pp. 134–151. Springer, Heidelberg (2002)
14. Cornelis, C., Deschrijver, G., De Cock, M., Kerre, E.E.: Intuitionistic fuzzy relational calculus: an overview. In: Proceedings of First International IEEE Symposium on Intelligent Systems, Varna, Bulgaria, vol. I, pp. 340–345 (2002) ISBN 0-7803-7601-3

15. Radzikowska, A.-M., Kerre, E.E.: A fuzzy generalization of information relations. In: Fitting, M., Orlowska, E. (eds.) Beyond Two: Theory and Applications of Multiple-Valued Logic, pp. 287–312. Physica Verlag (2003)
16. Deschrijver, G., Kerre, E.E.: On the composition of intuitionistic fuzzy relations. Fuzzy Sets and Systems 136(3), 333–362 (2003)
17. Deschrijver, G., Kerre, E.E.: On the cuts of intuitionistic fuzzy compositions. Kuwait Journal of Science and Engineering 32(1), 17–38 (2005)
18. De Cock, M., Cornelis, C., Kerre, E.E.: Intuitionistic fuzzy relational images. In: Halgamuge, S.K., Wang, L. (eds.) Computational Intelligence for Modeling and Predictions. Studies in Computational Intelligence, vol. 2, pp. 129–146. Springer, Heidelberg (2005)
19. Kerre, E.E.: An overview of fuzzy relational calculus and its applications. In: Torra, V., Narukawa, Y., Yoshida, Y. (eds.) MDAI 2007. LNCS, vol. 4617, pp. 1–13. Springer, Heidelberg (2007)
20. Serra, J.: Image analysis and mathematical morphology. Academic Press Inc., London (1982)
21. Haralick, R.M., Sternberg, S.R., Zhuang, X.: Image analysis using mathematical morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(4), 532–550 (1987)
22. De Baets, B., Kerre, E.E., Gupta, M.M.: Fundamentals of fuzzy mathematical morphology, part I: basic concepts. International Journal of General Systems 23, 155–171 (1995)
23. De Baets, B., Kerre, E.E., Gupta, M.M.: Fundamentals of fuzzy mathematical morphology, part II: idempotence, convexity and decomposition. International Journal of General Systems 23(4), 307–322 (1995)
24. De Baets, B.: Fuzzy morphology: a logical approach. In: Ayyub, B.M., Gupta, M.M. (eds.) Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach, pp. 53–67. Kluwer Academic Publishers, Boston (1997)
25. Nachtegael, M., Kerre, E.E.: Fuzzy Mathematical Morphology: State of the Art. In: Kerre, E.E., Nachtegael, M. (eds.) Fuzzy Techniques in Image Processing. Studies in Fuzziness and Soft Computing, vol. 52, pp. 3–57. Springer, Heidelberg (2000)
26. Nachtegael, M., Kerre, E.E.: Connections between binary gray-scale and fuzzy mathematical morphologies. Fuzzy Sets and Systems 124, 73–86 (2001)
27. Sussner, P., Valle, M.E.: Classification of Fuzzy Mathematical Morphologies Based on Concepts of Inclusion Measure and Duality. Journal of Mathematical Imaging and Vision (accepted for publication) (2009)
28. De Cock, M., Nachtegael, M., Kerre, E.E.: Images under Fuzzy Relations: a Master-Key to Fuzzy Applications. In: Intelligent Techniques and Soft Computing in Nuclear Science and Engineering - Proceedings of FLINS 2000 (4th International Conference on Fuzzy Logic and Intelligent Technologies for Nuclear Science and Industry), pp. 47–54. World Scientific Publishing, Singapore (2000)
29. Nachtegael, M., De Cock, M., Van der Weken, D., Kerre, E.E.: Fuzzy Relational Images in Computer Science. In: Proceedings of RELMICS 6 (6th International Workshop on Relational Methods in Computer Science), Katholieke Universiteit Brabant, The Netherlands, pp. 156–170 (2001)
30. Nachtegael, M., De Cock, M., Van der Weken, D., Kerre, E.E.: Fuzzy Relational Images in Computer Science. In: de Swart, H. (ed.) RelMiCS 2001. LNCS, vol. 2561, pp. 134–151. Springer, Heidelberg (2002)

A Combined Fuzzy and Probabilistic Data Descriptor for Distributed CBIR

Roberto Gallea, Marco La Cascia, and Marco Morana

Università degli studi di Palermo
DINFO - Dipartimento di Ingegneria Informatica
Viale delle Scienze - Ed.6 - 3° piano - 90128 Palermo (ITALY)
{robertogallea,lacascia,marcomorana}@unipa.it

Abstract. With the wide diffusion of digital image acquisition devices, the cost of managing hundreds of digital images is quickly increasing. Currently, the main way to search digital image libraries is by keywords given by the user. However, users usually add ambiguos keywords for large set of images. A content-based system intended to automatically find a query image, or similar images, within the whole collection is needed. In our work we address the scenario where medical image collections, which nowadays are rapidly expanding in quantity and heterogeneity, are shared in a distributed system to support diagnostic and preventive medicine. Our goal is to produce an efficient content-based description of each image collection in order to perform content-based image retrieval (CBIR) just in the node where the searched images are supposed to be. A novel combined fuzzy and probabilistic data descriptor is presented and experimental results are illustrated.

Keywords: Fuzzy clustering, distributed CBIR, medical images.

1 Introduction

With the wide diffusion of digital image acquisition devices (e.g. digital cameras, MRI scanners, PET scanners), the cost of managing hundreds of digital images is quickly increasing. For instance, medical centers take thousands of datasets everyday, but storing them without a proper organization becomes useless.

Currently, the main way to search digital image libraries is by keywords given by the user. However, this process has been observed to be unsatisfactory since users add few, often subjective, keywords for large set of images, raising ambiguities. In this scenario the need for organizing and finding data in distributed systems represents a further challenging task. A content-based system intended to automatically find a query image, or similar images, within the whole collection is needed.

Content-based image retrieval (CBIR) copes with the difficulties of describing multimedia content in text format. Several works on resource description and selection for text data has been proposed [1,2]. Some works [3,4] rely on keywords frequencies to route a given query to a resource. However, image database management requires *visual* content description to avoid text keywords ambiguity,

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 189–196, 2009.

typical of manual labeling. Due to multimedia capabilities of modern computing, this issue is currently subject of many studies and relevant works were conducted. In [5] a metadatabase records visual content of images per database through image templates and statistical features. Resource selection is done by an histogram-based procedure. A probabilistic model of the database feature space is proposed by [6,7]. In [6] the feature space is fitted by a parametric Gaussian Mixture, while [7] considers features as statistically independent and uses monodimensional histograms for each distribution. Since both have been used to validate our work, more details will be given in Sect. 2.2.

In this paper we address a scenario where medical image collections are shared in a distributed system. Users search through collections of classified data providing a new, unclassified, query image. This can be useful to support doctors' activities, for instance to diagnose already known pathologies. A rough approach is to broadcast the query across the whole network, that is each node has to execute local queries on his data and to return a set of results. Clearly this approach causes query flooding and usually generates network congestion.

Our goal is to produce an efficient content-based description of each image collection in order to perform content-based image retrieval (CBIR) just in the node where the searched image is supposed to be.

This task develops upon two different processes: *resource description* and *resource selection*.

The resource description process aims to the extraction of feature vectors [8] that represent the content information of each image. These vectors are then combined, in some way, to give a resource compact description. The resource description is finally used to support resource selection, that is a score is assigned to all nodes and the query is issued to winning resources only.

The paper will show the following structure: an analysis of the feature space characteristics will be given (Sect. 2.1). The Sect. 2.2 will give an overview of probabilistic methods for data description and retrieval, while the proposed c-means clustering approach is described in Sect. 2.3. Experimental results are shown in Sect. 3 and discussed in Sect. 4. Conclusions will follow in Sect. 5.

2 Methods

In the scenario we supposed to address, a virtual or phisical network is supposed to exist. Each node of such network owns an arbitrary collection of images depicting anatomical structures acquired by means of an MRI scanner. When a new node intends to connect to the network, it constructs a compact summary of all of its data. Then this information is spread across the whole network. When a node wants to submit a query it makes some inference on these descriptors and determines which nodes are most likely to contain the required information. A priority list is created and the queries are then forwarded following this list. When a network node receives the query, it performs a local search and the results are sent back to the node which made the request.

In our work we focus on the first part of the scenario, designing the nodes data descriptor and a procedure to choose how to estabilish the query order among the nodes, so we make no considerations on how the local search is performed.

2.1 Feature Space and Content Description

Each image in the collection can be represented by a finite set of discriminating features. Various features can be used for this purpose, we chose to adopt Haralick's [8] texture-based set of measures such as Mean, Variance, Covariance, Contrast, Energy, Entropy, Homogeneity, etc. In addition to these we added some other application-relevant morphological features like shape perimeter and area.

Each vector represents a point in a k-dimension space where k is the cardinality of the features. If the features are chosen in a proper way, the points (i.e. the images) are expected to be distributed along the whole space to form clusters more or less defined.

2.2 Probabilistic Approaches

In order to obtain a description of the content of a node, an intuitive method is to construct a compact description of its feature space. This can be done by means of probabilistic models fitting, either parametric (e.g. using Gaussian Mixtures [6]), or non-parametric (e.g. histograms [7]). Such methods grant good results but are not practicable using high dimensional feature space. The solution adopted by [7] is to consider each feature statistically independent so that the model is fitted just over k monodimensional single feature distributions. Altough this effectively reduces the complexity of the method, the assumption that features are independent is not always admissible. In all of these methods, the problem of finding the node n most likely to contain the required image is reduced to find the node which maximizes the probability

$$n = \arg\max_{a} \left(P\left(X = x | A = a \right) \right) \tag{1}$$

where x is the query data vector and A is the random variable representing the node containing a given feature vector. In case of statistical independece approximation reduces to

$$n = \arg\max_{a} \left(\prod_{j=1}^{k} P\left(X_j = x_j | A = a \right) \right) \tag{2}$$

2.3 Combined Fuzzy and Probabilistic Data Descriptor

In our approach we abandoned the probabilistic framework and adopted Fuzzy C-means [9] as clustering technique. The original algorithm is based on the minimization of the following objective function:

$$J_s = \sum_{j=1}^{m} \sum_{i=1}^{k} \left(u_{ij} \right)^s d\left(x_i, c_j \right)^2, \quad 1 \leq s \leq \infty \tag{3}$$

where the membership degrees u_{ij} are positive and structured such that $u_{i,1} + u_{i,2} + \ldots + u_{i,m} = 1$.

The method proceeds as an iterative procedure that terminates after a fixed number of iterations or when the improvement of each iteration is substantially small.
The descriptor is constructed as follows:

- Features values are extracted from the images in the whole local database
- Feature space is populated with features vectors
- Fuzzy C-means is applied to find cluster centroids in the feature space
- The descriptor is given by coordinates of centroids:

$$D_F = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1j} \\ c_{21} & c_{22} & \cdots & c_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kj} \end{bmatrix} \tag{4}$$

Using this approach the problem of finding the node most likely to contain a required query image can be chosen in several ways, provided that a score for each node has been assigned. Each query image has some fuzzy membership degree for each of the clusters resulted from C-means. Taking the maximum of the membership degrees, the image is assigned to belong to a single cluster and the score of the node is represented by this membership degree itself. Note that different, perhaps more efficient methods can be used to give a score measure to each node, and this will be subject of further study.

In addition, hybrid fuzzy-probabilistic descriptors can be designed to overcome and mitigate the weakness of both techniques. In such case the descriptors are given by both fuzzy centroids and probabilistic model parameters. We chose to fit 1-D Gaussian Mixtures over single-feature distributions. The model is fitted using an Expectation Maximization [10] algorithm, which gives as output a set of means, variances and reconstruction factors.

$$D_P = \begin{bmatrix} \mu_{11}, \sigma_{11}, \tau_{11} & \mu_{12}, \sigma_{12}, \tau_{12} & \cdots & \mu_{1m}, \sigma_{1m}, \tau_{1m} \\ \mu_{21}, \sigma_{21}, \tau_{21} & \mu_{22}, \sigma_{22}, \tau_{22} & \cdots & \mu_{2m}, \sigma_{2m}, \tau_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k1}, \sigma_{k1}, \tau_{k1} & \mu_{k2}, \sigma_{k2}, \tau_{k2} & \cdots & \mu_{km}, \sigma_{km}, \tau_{km} \end{bmatrix} \tag{5}$$

where m is the number of Gaussian used in each mixture.

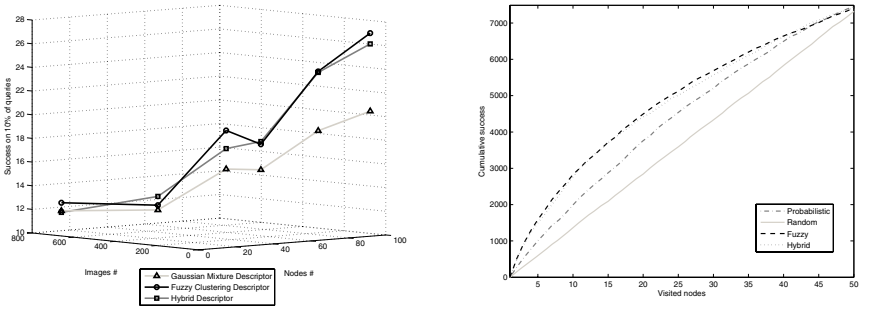
The resulting descriptor is the concatenation of C-means cluster centroids and Gaussian Mixture model parameter vector:

$$D_H = [D_F | D_P] \tag{6}$$

Finally a further adaptive mixing parameter λ can be added to adjust the weight of the two aggregated descriptors.

3 Experimental Tests and Results

To validate our framework, several tests were conducted on simulated networks using real datasets. Each node of such networks has been preloaded with a a random set of 7696 MRI brain images, obtained slicing a subset of the OASIS volumes [11]. OASIS consists of a cross-sectional collection of 416 subjects aged 18 to 96. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. For each node the features have been extracted and feature spaces have been built. The proposed descriptors (C-means based, Gaussian Mixture based and hybrid) have been computed and system simulations have been run. For each simulation 7500 queries have been executed.



(a) (b)

Fig. 1. (a) Rate of success searching the query image in first 10% of network. (b) Cumulative success curve diagram.

The second measure is about efficiency of the system. The concept is explained directly from considerations on the graphical diagram in Fig. 1 (b). Plotting on the y axis the cumulative successes after querying the i-th node, an efficiency curve is obtained. The slope of such curve indicates the system quality. The steeper the curve, the better the system performs. For reference purposes, an approximately straight line resulting from the random querying scheme is depicted too.

Both described tests search for an exact match of the query image, additional tests have been performed to simulate the complete CBIR system, that is searching not just for the very same image, but even *similar* ones as well. Similarity between two images refers to their visual and structural differences; in order to measure it, SSIM [12] has been used. The SSIM index is a value between 0 (i.e. mean zero correlation with the original image) and 1 (i.e. same image). The system has been evaluated using Fuzzy descriptors and the same network configuration of the first test. The number of images found in each node whose SSIM index exceeds a given threshold is recorded. Table 2 summarizes the means of images found for best five nodes of priority list after 7500 queries.

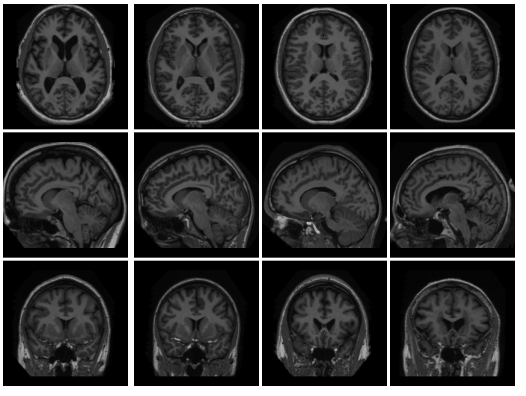


Fig. 2. Examples of query images (first column) and relative results (right) for three different views of MRI images

Table 1. Rate of success searching the query image in first 10% of network using different number of nodes, images and clusters. Results (%) for our Fuzzy method, GM model and Hybrid approach are shown.

	Nodes	Images	Fuzzy	GM	Hybrid
9 Clusters	100	75	18.49	20.55	18.68
	75	100	18.46	19.25	17.15
	50	150	14.51	15.66	14.70
	37	200	14.81	15.46	14.12
	20	375	12.51	13.49	12.39
	10	750	11.63	11.69	11.51
16 Clusters	100	75	26.93	20.33	26.02
	75	100	23.96	18.95	23.87
	50	150	18.02	15.87	18.25
	37	200	19.27	16.00	17.73
	20	375	12.85	12.45	13.60
	10	750	12.52	11.84	11.70
24 Clusters	100	75	38.33	20.05	38.16
	75	100	32.13	17.74	32.64
	50	150	23.44	16.54	24.23
	37	200	23.15	17.53	23.00
	20	375	15.81	13.53	16.80
	10	750	13.21	11.89	12.47

4 Discussion

Experimental results showed that our Fuzzy descriptor performs generally better than Gaussian Mixture model descriptor. Even though using the same descriptor size leads to similar results, our descriptor scales better. While Fuzzy descriptor performance can be improved increasing the number of clusters, increasing Gaussian Mixture model performance is not as easy: in the linearly independent features case, adding more Gaussians does not improve the distributions fitting (i.e. performance does not change). Instead, fitting the whole

Table 2. Means of images found for best five nodes of priority list, after 7500 queries, using different number of nodes, images and clusters

	Nodes	Images	1st	2nd	3rd	4th	5th
9 Clusters	100	75	7.12	6.01	5.23	5.31	4.83
	75	100	9.31	8.25	7.45	7.03	6.38
	50	150	14.01	13.26	13.12	12.67	11.78
	37	200	18.22	17.46	16.28	15.55	15.18
	20	375	35.51	34.39	32.73	31.15	30.68
	10	750	71.63	70.29	69.51	67.45	67.11
16 Clusters	100	75	9.51	9.02	8.32	7.71	7.13
	75	100	11.64	10.82	10.05	9.34	9.03
	50	150	16.72	15.96	15.21	14.67	13.97
	37	200	21.24	20.86	19.78	19.25	18.53
	20	375	37.14	35.23	35.07	34.64	33.98
	10	750	74.31	73.92	72.52	72.15	71.77
24 Clusters	100	75	10.31	10.01	9.62	8.41	7.73
	75	100	13.42	12.89	12.01	11.53	10.68
	50	150	17.28	16.66	15.84	15.13	14.31
	37	200	23.02	22.46	22.08	21.65	21.03
	20	375	39.01	38.37	37.77	37.01	36.12
	10	750	76.23	75.02	75.21	74.12	73.22

feature space with more Gaussians would actually improve the overall system performance, but is too hard to achieve using an high number of features. Gaussian Mixture approach results more effective when few images are available in the node. Using an hybrid approach can represent a good choice in this case. The λ parameter can be set to automatically adapt to the local collection size, weighting more Gaussian Mixture approach in small image databases, Fuzzy one otherwise.

Obviously this adaptive solution is not for free, since the descriptor size increases to the sum of both descriptors.

5 Conclusions

In this work was addressed the task of searching medical images shared in a distributed system. In this scenario, an efficient content-based description of each collection is needed in order to perform content-based image retrieval (CBIR) just in the nodes where the searched image is supposed to be. Probabilistic approaches are usually based on models fitting, either parametric or non-parametric. However, these methods are not practicable using high dimensional correlated-feature space. A novel Fuzzy content-based image descriptor has been presented. Firstly, features values are extracted from the images contained in each node, then Fuzzy C-means clustering is applied to find cluster centroids in the feature space. The descriptor is given by coordinates of centroids. In addition, a combined Fuzzy and Probabilistic descriptor has been designed to overcome and mitigate the weakness of both techniques. An adaptive mixing parameter λ can be added to adjust the weight of the two aggregated descriptors and this will be subject of further study.

References

1. Schwartz, M.F., Emtage, A., Kahle, B., Neuman, B.C.: A comparison of internet resource discovery approaches (1993)
2. Obraczka, K., Danzig, P.B., Li, S.h.: Internet resource discovery services. IEEE Computer 26, 8–22 (1993)
3. Fuhr, N.: Optimum database selection in networked ir. In: Proceedings of the SIGIR 1996 Workshop Networked Information Retrieval, p. 7 (1996)
4. Callan, J.P., Lu, Z., Croft, W.B.: Searching distributed collections with inference networks. In: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 21–28. ACM Press, New York (1995)
5. Chang, W., Sheikholeslami, G., Zhang, A., Syeda-mahmood, T.F.: Efficient resource selection in distributed visual information systems. In: ACM Multimedia 1997, pp. 926–946 (1997)
6. Miguel, N., De Pinho, B., De Vasconcelos, C.: Bayesian models for visual information retrieval. PhD thesis, Supervisor-Andrew B. Lippman (2000).
7. Müller, W., Henrich, A.: Fast retrieval of high-dimensional feature vectors in p2p networks using compact peer data summaries. In: MIR 2003: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, Berkeley, California, pp. 79–86. ACM, New York (2003)
8. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3(6), 610–621 (1973)
9. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms (Advanced Applications in Pattern Recognition). Springer, Heidelberg (1981)
10. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(1), 1–38 (1977)
11. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. J. Cognitive Neuroscience 19(9), 1498–1507 (2007)
12. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)

A Fuzzy Approach to the Role of Symmetry in Shape Formation: The Illusion of the Scalene Triangle

Baingio Pinna[1] and Marco E. Tabacchi[2,3]

[1] Dipartimento di Scienze dei Linguaggi, University of Sassari
[2] DMA, Università degli Studi di Palermo
[3] Istituto Nazionale di Ricerche Demopolis, Italy
baingio@uniss.it, tabacchi@math.unipa.it

Abstract. The main purposes of this work are to demonstrate the role of directional symmetry as a second order principle that polarizes the perception of the shape and to show how this preference can be easily encoded in an algorithm using a fuzzy operator for symmetry detection.

The role of grouping in influencing shape perception and the role of directional symmetry was demonstrated through small triangles that create a large triangle. The specific questions answered in the psychophysical experiments were the following: Can the grouping by similarity influence both the pointing and the shape of the small and the large isosceles triangles? Conversely, can the shape of the large triangle influence the perceptual strength and direction of the grouping of the inner elements? As for the algorithm implementation, we wanted show that a fuzzy operator, usually employed in symmetry detection, can as well represent such a preference by obtaining a human-like performance.

The results demonstrated that the grouping principles influence not only the way elements in the visual field "go together" to form an integrated, holistic (Gestalt) percept, but also the local and the whole shape perception: the grouping by similarity influences both the pointing and the shape of both the small and the large isosceles triangles. This creates the illusion of a scalene triangle: small and large isosceles triangles appear as scalene. Conversely, the shape variation of the large triangle induced by the bevelling of the large triangle influences the perceived strength and direction of the grouping of the inner small triangles.

We suggest that the preference for such principles can be emulated using a fuzzy algorithm that capture the gist of humans' preferences for symmetry.

Keywords: symmetry, fuzzy operators, shape formation, perception.

1 Visual Organisation and Shape Perception

1.1 The Form of Grouping

The problem of perceptual organization, first studied by Gestalt psychologists, is one of the central issues of Vision Science. It is related to the perception of

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 197–204, 2009.

a world made up of phenomenal objects but not differences of luminance and edges. The main question [11,12,13,14] asked is: how do the elements in the visual field 'go together' to form an integrated, holistic percept (Gestalt)? The answer to this question brought Wertheimer to discover the well-known 'principles of grouping'.

In Figs. 1a-b, two large square shapes made up of rows (a) or columns (b) of small squares are perceived. In Fig. 1c, the small squares do not show any preferential direction of the inner organization, i.e. neither rows nor columns, but a lattice of small squares is globally perceived. Under these conditions, the row or column organization can be induced through the visual attention but it appear not as clearly as in Figs. 1a-b and easily reversible when the attention is switched in one or in the other result.

The inner local organization of the small squares in rows or columns is due to the Gestalt grouping principle of similarity [14], stating that, all else being equal, the most similar elements (in color, brightness, size, empty/filled, shapes, etc.) are grouped together. The outer global organization in large square shapes is due to the principle of similarity of shape and to the principle of exhaustiveness, according to which, all else being equal, all the components of a stimulus pattern tend not to be left outside but included as parts in a grouped whole.

1.2 The Form of Shape: The Rectangle Illusion

On a closer observation of Fig. 1, nave subjects agreed to report more subtle and precise properties of the perceived shapes [8,9].

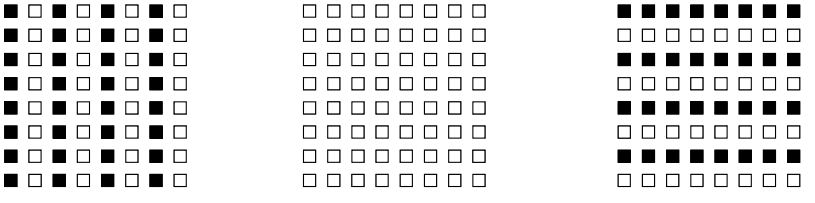


Fig. 1. Two large square shapes made up of rows (a) or columns (b) of small squares. In (c), rows and columns are not perceived as clearly and stably as in (a), (b), but they appear reversible. Both small and large squares appear similar to vertical (a) and horizontal (b) rectangles.

The shapes of small and large squares are not perceived isotropic (directional invariant) but with a clear directional symmetry. These results emerge more clearly by comparing them with those of the control illustrated in Fig. 1c. The inner organization in rows or column orients and elongates the shape of both the small and the large squares in the same direction as the one of the perceptual grouping. In other words, the perception of the rows distorts by widening the base of both the small and the large squares that appear like horizontal rectangles. On the contrary, the column organization induces a perceptual lengthening of

the height of both small and large squares that appear similar to vertically rectangles.

These apparent deformations of the whole and local geometric shapes persist or are even stronger by zooming the focus of attention only on a small array of squares, e.g. the extreme 3X3 squares in the left or right upper side of Figs. 1a-b. Similarly to the whole deformation, each small square appears distorted like a rectangle.

This effect can be related to Oppel-Kundt's group of illusions, according to which an empty (unfilled) space looks wider than a space filled by some objects [2,4] and to Helmholtz's square illusion, where a square appears wider when it is filled with vertical lines and higher when filled with horizontal lines [3]. However, our effects show several important differences: (i) the whole shape distortion is induced by grouping and not by filled vs. unfilled space; (ii) the direction of the illusory distortion is the opposite of the one perceived in both Oppel-Kundt's and Helmholtz's square illusions, and (iii) the shape distortion involves both the small squares and the whole square shape. The previous descriptions suggest that the form of grouping can influence the form of shape: squares "become" rectangles, and this might depend on the directional symmetry derived from the grouping by similarity. They also demonstrate a relation, to be studied in depth, between the problem of grouping and the process of shape perception. In Fig. 1, the perceptual results show how individual elements group into wholes separated from others. Grouping *per s*e does not make any prediction about shape. The role of the gestalt principles is to define the rules of "what is or stay with what" and, then, the grouping and not the shape. The notion of whole due to grouping is phenomenally different from the one depending on shape and represents the groups of elements that assume the role of "parts" of a holistic percept. The form of shape is instead the result of a global perceptual process emerging likely parallel to or after the form of grouping and giving to the whole a unitary form along the boundary contours. Nevertheless, grouping and shape formation can be considered as two complementary integrated processes of perceptual organization. This is not a literal or a fictitious distinction but a phenomenal necessity that can have consequences in terms of neural circuitry. [8] (in press) and [9] (under revision) have recently distinguished a further kind of perceptual organization called "form of meaning" to be added to the other two forms.

The main purposes of this work are (i) to demonstrate the role of directional symmetry as a second order principle that polarizes the perception of the shape and (ii) to show how this preference can be encoded in an algorithm using a fuzzy operator for symmetry detection.

2 The Illusion of the Scalene Triangle: Phenomenology, Psychophysics and Fuzzy Modelling

The role of grouping in influencing shape perception and the role of directional symmetry can be demonstrated by using small triangles that create a large triangle. The purpose of the next experiments and of the following adaptation

of a symmetry detection fuzzy model is to answer the following questions: Can the grouping by similarity influence both the pointing and the shape of the small and the large isosceles triangles? Conversely, can the shape of the large triangle influence the perceptual strength and direction of the grouping of the inner elements? What is the role of symmetry in determining grouping and shape perception? Is this role easily implementable using a fuzzy model?

2.1 Methods

Subjects. Independent groups of 14 nave undergraduate students participated to the experiments. All subjects had normal or corrected-to-normal acuity.

Stimuli. The stimuli were composed of small isosceles triangles connected in their vertexes as to create a large isosceles triangle (see Fig. 2).

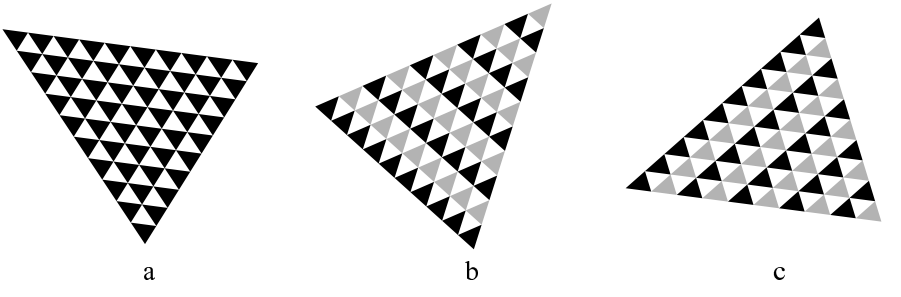


Fig. 2. (a) Small isosceles triangles create a large isosceles triangle pointing toward the top left-hand corner. (b) The grouping of the triangles on the basis of similarity of lightness makes the large elongated isosceles shape to appear more pronounced than the one of (a) and more strongly determining the pointing of both large and small rectangles toward the top right-hand corner. (c) When the grouping of isosceles triangles is parallel to one of the two equal sides of each small triangle, the three sides of both the small and large triangles appear unequal, i.e. scalene, pointing preferentially toward the bottom left-hand corner.

Four variations were created: 3 directions of grouping by similarity and 1 where the grouping is absent and then all the triangles are black. The two sides of the isosceles small and large triangles were respectively: 48.12 arcmin and 57.74 arcmin. The stimuli were presented on a computer screen with ambient illumination from a Osram Daylight fluorescent light (250 lux, 5600°K). The luminance of the white background was 88.3 cd/m2. Black and gray triangles had a luminance contrast of 0.97 and 0.45 respectively (luminance value of 2.6 and 51.3 cd/m^2). Stimuli were presented in a frontoparallel plane and viewed binocularly at a distance of 50 cm. During the experiment the orientation of the large triangle was randomly varied from one subject to another. The head position of the observer was stabilized by a chin rest.

Procedure. The method of forces choice and adjustment were used respectively in the complete and bevelled psychophysical conditions. Before starting the experiment the subjects familiarized with the notion of isosceles and scalene triangles seeing some examples of both complete or bevelled triangles on a computer screen. Then, for practicing, the subjects were invited to press one of the two buttons labelled "I" for "isosceles" or "S" for "scalene" and, by adjusting the position of an arrow, to define its perceived pointing.

By pressing a button two independent groups of 14 subjects chose the perceived shapes, isosceles or scalene, of the small or large triangles. By positioning an arrow, two independent groups of 14 subjects chose the direction of the perceived pointing of the small or large triangles. Each type of measurement was randomized and repeated three times.

At the end of the previous experimental sessions, the task of the subjects was to scale the relative strength of the perceived grouping effect due to similarity, i.e. how easy, immediate and direct is to perceive the grouping by similarity with respect to the perceived shape and pointing of the small and large triangles. The range values were from 1 to 6, the same as the number of stimuli with black and gray triangles. The upper value "6" was defined by the perceived strongest grouping among the six stimuli, i.e. by the one most easily and immediately perceived in relation to the shape and pointing. Whereas the value "1" was defined as the minimum grouping perceived, i.e. the least easy, immediate and direct. The stimuli were present all at the same time to each observer in a random order.

2.2 Replicating the Scalene Triangle Illusion Using a Fuzzy Model

In order to replicate human performance in the judgment of directionality we have chosen to use FiST [10], a fuzzy model we have already used at the purpose of exposing preferential biases regarding vertical axes in human judgment when detecting symmetry in single objects.

The Field Symmetry Transform (FiST for short) is an algorithmic approach to the problem of symmetry detection in digital objects. FiST takes as input a digital raster image, and outputs an histogram of the symmetry distribution in the image itself; analysis of maxima and minima in the histogram reveals how the main symmetric axes are angled.

The process employed by the algorithm is as follows: FiST treats the input image as a bidimensional plane, in which each pixel $p(x, y)$ is a virtual charged particle at continue plane coordinates (x, y), with positive intensity proportional to the fuzzy grey intensity of p. Once all of the pixel have been represented as virtual charged particles, an equal-spaced orthogonal grid is superimposed on the bidimensional plane, and the vector field resulting from the contribution of all the virtual charges is computed at each crossing point of the grid. The contribution given by the charged particle p lying at coordinates (x, y) to the field at grid point h of coordinates (i, j), according to Coulomb's law, is:

$$E_p^{(h)} = \frac{1}{4}\pi\frac{Q_x}{r^2}\hat{r}$$

where Q is the charge of the particle p, and r is the unit vector pointing form the particle p to the evaluation point h (or from the point (x, y) to the point (i, j)). Due to the superimposition principle of the charges, the total field vector E in h from all the points in grid G is given by:

$$E_p = \sum_{h \in G} E_p^{(h)}$$

The vectors obtained through this process are then clustered according to their direction, and a histogram is obtained by counting the number of vectors in each cluster. The resulting histogram is the FiST of the original image, and by appropriate rescaling it is interpreted as a fuzzy degree of membership to the symmetric set: when only one main symmetry is found, the result is directly interpretable as a fuzzy number centered on the maximum symmetric axis, while when different symmetry axis are found, the obtained histogram can be decomposed as the union of different fuzzy numbers, with their relative degree of membership rescaled for comparative purposes. Choice of parameters for FiST (resolution of the grid, width of the clusters, linearity of the charge in function of the pixel intensity) depends strictly on the type of image taken in input, its resolution and bit depth. Due to the theoretical approach used, FiST has some interesting peculiarities: among them the fact that, being based on a field metaphor, specific preferences for directions (such as the vertical bias usually observed in humans) can easily be encoded by directly manipulate the field.

Stimuli. Triangles shown in Fig. 2 has been digitalised using standard 4×4 antialiasing techniques, at 512×512 pixels.

Procedure. FiST has been applied as follows: the vector grid has been sized at 128×128, using only the vectors lying inside a digital circle with 256 pixels radius co-centered with the original image, in order to balance the geometrical contribution of the original image. The interval used for histogram building has been chosen as 6°, in order to minimize interferences due to the discrete nature of the image and the ensuing noise. The algorithm has been applied using custom code developed in MATLAB.

2.3 Results and Conclusions

In Fig. 3a, the mean ratings of choices of the isosceles triangle for the small and large triangles are plotted as a function of each stimulus (black triangles, gray/black synergistic, gray/black antagonistic and gray/black antagonistic). In Fig. 3b, the rates of choices of the pointing in the direction of the most acute angle of the isosceles triangle are plotted as a function of each stimulus. When the grouping is absent and all the triangles are black (see Fig. 3a), the subjects perceived small isosceles triangles creating a large isosceles triangle all pointing toward a univocal direction – the geometrical one. When the grouping by similarity of lightness is synergistic with or parallel to the smallest side of both small and large triangles (see Fig. 2b), the small and large isosceles triangles appeared more pronounced, elongated and pointed than the one of Fig. 2a. When

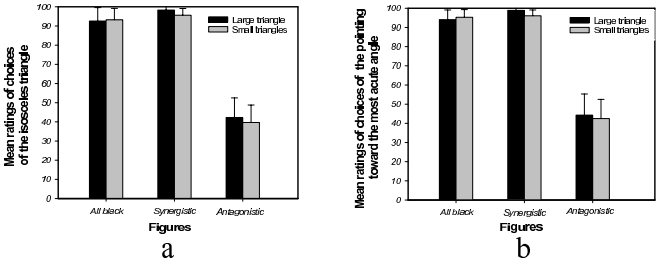


Fig. 3. (a) Rates of choices of the isosceles triangle for both small and large triangles plotted as a function of each stimulus (black triangles, gray/black synergistic, gray/black antagonistic and gray/black antagonistic). (b) Rates of choices of the pointing in the direction of the most acute angle of the large isosceles triangle plotted as a function of each stimulus.

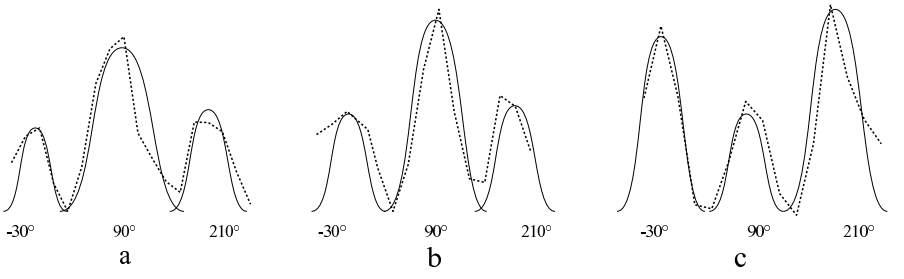


Fig. 4. Results form FiST. In Synergistic (a) or Uniform (b) condition, the direction of the figure is clearly determined by its predominant axis. In Contrasting condition (c), a different axis emerges. In this graph, angles are normalised as if shown in horizontal position.

the grouping of isosceles triangles is parallel to one of the two equal sides of both small and large triangles (Fig. 2c), a local and global scalene shape illusion emerged: the small and large triangles appear scalene, pointing preferably toward the direction opposite to the direction of the grouping.

As shown in Figure 4, results form the application of FiST share the same underlying meaning with results obtained from the psycho physical experiment. When there is no explicit grouping (Fig. 4a), or internal and general directions agrees (Synergistic condition, Fig. 4b) the direction is univocally determined by this agreement. In the contrasting condition (Fig. 4c), a symmetry axis, corresponding to the opposite direction respect to the original grouping, absent in the other conditions, emerges. We call this effect "the illusion of the scalene triangle". These results are phenomenally corroborated by the spontaneous descriptions of the subjects. Results of a two-way mixed factorial ANOVA (figure variation – within-subjects variable, small and the large triangles – between-subjects variable) was significant only for the difference between the figures in both choices (isosceles triangle –$F_{3,24}$=6.6, P<0.01; pointing –$F_{3,24}$=5.4, P<0.01).

References

1. Attneave, F.: Triangles as ambiguous figures. American Journal of Psychology 81, 447–453 (1968)
2. Da Pos, O., Zambianchi, E.: Visual illusions and effects. Milano, Guerini (1996)
3. von Helmholtz, H.: Handbuch der Physiologischen Optik, Part III. Voss, Leipzig (1866); Kristof, W.: Versuche mit der Waagrechten Strecke-Punkt-Figur. Acta Psychologica 18, 17–28 (1961)
4. Oppel, J.J.: Über Geometrisch-optische Täuschungen. Jahresbericht des Physikalischen Vereins zu Frankfurt am Main, 37–47 (1854-1855)
5. Palmer, S.E.: What makes triangles point: Local and global effects in configurations of ambiguous triangles. Cognitive Psychology 12, 285–305 (1980)
6. Palmer, S.E.: Reference frames in the perception of shape and orientation. In: Shepp, B.E., Ballesteros, S. (eds.) Object perception: Structure and process, Hillsdale, NJ, pp. 121–163. Erlbaum, Hillsdale (1989)
7. Palmer, S., Bucher, N.M.: Textural effect in perceiving pointing of ambiguous triangle. Journal of Experimental Psychology: Human Perception & Performance 8(5), 693–708 (1981)
8. Pinna, B., Albertazzi, L.: From grouping to visual meanings: A new theory of perceptual organization (under revision)
9. Pinna, B., Reeves, A.: From Perception to Art: How the brain creates meanings. Spatial Vision (in press)
10. Tabacchi, M.E.: FiST, Preferenze Cognitive nella scelta delle direzioni apparenti, In-Cognito Quaderni Romanzi di Scienze Cognitive (in press)
11. Wertheimer, M.: Über das Denken der Naturvölker. Zeitschrift für Psychologie 60, 321–378 (1912a)
12. Wertheimer, M.: Untersuchungen über das Sehen von Bewegung. Zeitschrift für Psychologie 61, 161–265 (1912b)
13. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt. I. Psychologische Forschung 1, 47–58 (1922)
14. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt. II. Psychologische Forschung 4, 301–350 (1923)

A Unified Algebraic Framework for Fuzzy Image Compression and Mathematical Morphology

Ciro Russo

Dipartimento di Matematica e Informatica – University of Salerno, Italy
cirusso@unisa.it

Abstract. In this paper we show how certain techniques of image processing, having different scopes, can be joined together under a common "algebraic roof".

Introduction

In the last years, fuzzy logics and fuzzy set theory have been widely applied to image processing tasks. In particular, the theory of *fuzzy relation equations*, deeply investigated in [2], is involved in many algorithms for compression and reconstruction of digital images (see, for example, [1,5,7]).

In such techniques, however, the approach is mainly experimental and the algebraic context is seldom clearly defined. Basically, most of the fuzzy algorithms for image compression, make use of join-product operators; after all, a complete lattice order and a multiplication that is residuated w.r.t. the lattice-order are the fundamental ingredients of these operators.

On the other hand, there exists another class of operators acting on digital images that, although having a completely different scope, has the same algebraic form: *mathematical morphological operators*. Mathematical Morphology is a technique for image processing and analysis whose birth can be traced back to the book [6] by G. Matheron and whose establishment is due mainly to the work of Heijmans and Serra (e.g. [3,4,10]). The basic problem in mathematical morphology is to design nonlinear operators that extract relevant topological or geometric information from images. This requires the development of a mathematical model for images and a rigorous theory that describes fundamental properties of the desirable image operators.

Essentially, mathematical morphological operators analyse the objects in an image by "probing" them with a small geometric "model-shape" (e.g., line segment, disc, square) called the *structuring element*. These operators are defined on spaces having both a complete lattice order (set or fuzzy set inclusion, in concrete applications) and an external action from another ordered structure (the set of translations); more, they are usually coupled in adjoint pairs.

Regarding these different classes operators, what is really outstanding from an algebraic point of view is the fact that they can both be expressed in terms of a complete lattice order and a residuated product. Our aim is to show how all these operators can be joined together in a common mathematical context: the

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 205–212, 2009.

categories of *quantale modules* and the operators called $\mathcal{Q}$*-module transforms*. We will also show that such operators

- are precisely the $\mathcal{Q}$-module homomorphisms between free modules,
- are completely determined by the mathematical counterpart of the coder (for compression algorithms) and of the structuring element (in the case of mathematical morphology).

Throughout the paper, due to space constraint, we will omit all the proofs of propositions and theorems; however they can all be found in [9].

1 Preliminaries

In this section we will briefly recall some basic notions on several ordered algebraic structures.

Definition 1. Let $\langle X, \leq \rangle$ and $\langle Y, \leq \rangle$ be two posets. A map $f : X \longrightarrow Y$ is said to be *residuated* iff there exists a map $g : Y \longrightarrow X$ such that, for all $x \in X$ and for all $y \in Y$, $f(x) \leq y \iff x \leq g(y)$.

It is immediate to verify that the map g is uniquely determined; we will call it the *residual map* or the *residuum* of f, and denote it by f_*. The pair (f, f_*) is called *adjoint*; a residuated map preserves all existing joins and its residuum preserves all existing meets.

The category $\mathcal{SL}$ of sup-lattices is the one whose objects are complete lattices and morphisms are maps preserving arbitrary joins or, that is the same, residuated maps. For a sup-lattice $\mathbf{L}$, we will use the notation $\mathbf{L} = \langle L, \vee, \bot \rangle$. For any sup-lattice $\mathbf{L} = \langle L, \vee, \bot \rangle$, it is possible to define a dual sup-lattice in an obvious way: if we consider the opposite partial order $\geq$, then $\mathbf{L}^{\mathrm{op}} = \langle L, \wedge, \top \rangle$ is a sup-lattice and, clearly, $(\mathbf{L}^{\mathrm{op}})^{\mathrm{op}} = \mathbf{L}$.

Proposition 2. *Let $\langle X, \leq \rangle$ and $\langle Y, \leq \rangle$ be posets, and let (f, f_*) be an adjoint pair, with $f : X \longrightarrow Y$. Then the following hold:*

$$\begin{array}{lccccc} (i) & f \text{ is surjective} & \iff & f_* \text{ is injective} & \iff & f \circ f_* = \mathrm{id}_Y; \\ (ii) & f \text{ is injective} & \iff & f_* \text{ is surjective} & \iff & f_* \circ f = \mathrm{id}_X. \end{array}$$

A binary operation $\cdot$ on a partially ordered set $\langle P, \leq \rangle$ is said to be *residuated* iff there exist binary operations $\backslash$ and $/$ on P such that for all $x, y, z \in P$, $x \cdot y \leq z$ iff $x \leq z/y$ iff $y \leq x \backslash z$. The operations $\backslash$ and $/$ are referred to as the left and right *residua* of $\cdot$, respectively. In other words, a residuated binary operation over $\langle P, \leq \rangle$ is a map from $P \times P$ to P that is residuated in both arguments. In the situations where $\cdot$ is a monoid operation with a unit element e and the partial order is a lattice order, we can add the monoid unit and the lattice operations to the similarity type to get an algebraic structure $\mathbf{R} = \langle R, \vee, \wedge, \cdot, \backslash, /, e \rangle$ called *residuated lattice*.

In the category $\mathcal{Q}$ of quantales, $\mathrm{Obj}(\mathcal{Q})$ is the class of complete residuated lattices and the morphisms are the maps preserving products, the unit, arbitrary joins and the bottom element. An alternative, yet equivalent, definition of quantale is the following.

Definition 3. A *quantale* is an algebraic structure $\mathbf{Q} = \langle Q, \vee, \cdot, \bot, e\rangle$ such that

($Q1$) $\langle Q, \vee, \bot\rangle$ is a sup-lattice,
($Q2$) $\langle Q, \cdot, e\rangle$ is a monoid,
($Q3$) $x \cdot \bigvee_{i\in I} y_i = \bigvee_{i\in I}(x \cdot y_i)$ and $\left(\bigvee_{i\in I} y_i\right) \cdot x = \bigvee_{i\in I}(y_i \cdot x)$ for all $x \in Q$, $\{y_i\}_{i\in I} \subseteq Q$.

$\mathbf{Q}$ is said to be *commutative* if so is the multiplication. Obviously, if $\mathbf{Q}$ is commutative then the two residua coincide and $x/y = y\backslash x$ is denoted by $y \to x$.

Before giving examples of quantale structures interesting for the scope of this paper, we recall that a binary operation $* : [0,1]^2 \longrightarrow [0,1]$ is called a *triangular norm*, *t-norm* for short, provided it is associative, commutative, monotone in both arguments and has 1 as the neutral element. A t-norm $*$ is called *left-continuous* if, for all $\{x_n\}_{n\in\mathbb{N}}, \{y_n\}_{n\in\mathbb{N}} \in [0,1]^{\mathbb{N}}$,

$$\left(\bigvee_{n\in\mathbb{N}} x_n\right) * \left(\bigvee_{n\in\mathbb{N}} y_n\right) = \bigvee_{n\in\mathbb{N}}(x_n * y_n).$$

In this case, clearly, $*$ is residuated and its residuum (unique, since $*$ is commutative) is given by $x \to y = \bigvee\{z \in [0,1] \mid z * x \le y\}$.

Example 4. If $*$ is any left-continuous t-norm on the real unit interval, then $\langle [0,1], \vee, *, 0, 1\rangle$ is a commutative quantale.

2 Join-Product Operators in Image Processing

2.1 Fuzzy Algorithms for Image Compression and Reconstruction

In the literature of image compression, the fuzzy approach is based essentially on the theory of fuzzy relation equations. The underlying idea is the following: a grey-scale image is a matrix in which every element represents a pixel and its value, included in the set $\{0, \ldots, 255\}$ in the case of a 256-bit encoding, is the grey-level. Then, if we normalize the set $\{0, \ldots, 255\}$ by dividing each element by 255, grey-scale images can be modeled equivalently as fuzzy relations, fuzzy functions (i.e. $[0,1]$-valued maps), fuzzy subsets of a given set or $[0,1]$-valued matrices. A similar model is used for RGB colour images, where each image is represented by three fuzzy relations (respectively: functions, sets or matrices).

So, if we consider a grey-scale image I of sizes $m \times n$ ($m, n \in \mathbb{N}$), we can see it as an $m \times n$ matrix I_{ij} whose values are in $[0,1]$. Now we consider two natural numbers $a \le m$ and $b \le n$ and fix a $[0,1]$-valued map in four variables $C \in [0,1]^{m\times n\times a\times b}$ — usually called *coder* or *codebook*; then we compress the image I into an image $I' = I'_{hk}$ of sizes $a \times b$ by setting

$$I'_{hk} = \bigvee_{i,j} I_{ij} * C_{ijhk}, \tag{1}$$

where $*$ is any left-continuous t-norm on $[0,1]$. The reconstructed image $I'' = I''_{ij}$ is defined by

$$I''_{ij} = \bigwedge_{h,k} C_{ijhk} \to_* I'_{hk}, \tag{2}$$

where $\to_*$ is the residuum of $*$.

Even if some fuzzy algorithms for image compression may look different at a first glance, most of them can be rewritten in a form similar to (1), with a reconstruction process that will consequently look like (2).

2.2 Dilation and Erosion in Mathematical Morphology

In Mathematical Morphology binary images are modeled, in the wake of tradition and intuition, as subspaces or subsets of a suitable space E, which is assumed to possess some additional structure (topological space, metric space, graph, etc.), usually depending on the kind of task at hand.

Concretely, the class of n-dimensional binary images is represented as $\mathscr{P}(E)$, where E is, in general, $\mathbb{R}^n$ or $\mathbb{Z}^n$. In the first case we have continuous binary images, otherwise we are dealing with discrete binary images. The basic relations and operations between images of this type are essentially those between sets, namely set inclusion, union, intersection and complementation. It is intuitively clear, then, that complete lattices are the algebraic structures required for abstracting the ideas introduced so far.

Definition 5. Let $\mathbf{L}$, $\mathbf{M}$ be complete lattices. A map $\delta : L \longrightarrow M$ is called a *dilation* if it distributes over arbitrary joins: $\delta\left({}^{\mathbf{L}}\bigvee X\right) = {}^{\mathbf{M}}\bigvee \delta(X)$, for every $X \subseteq L$. A map $\varepsilon : M \longrightarrow L$ is called an *erosion* if it distributes over arbitrary meets: $\varepsilon\left({}^{\mathbf{M}}\bigwedge Y\right) = {}^{\mathbf{L}}\bigwedge \varepsilon(Y)$, for every $Y \subseteq M$.

Two maps $\delta : L \longrightarrow M$ and $\varepsilon : M \longrightarrow L$ are said to form an *adjunction*, (δ, ε), between $\mathbf{L}$ and $\mathbf{M}$ if $\delta(x) \leq y \Longleftrightarrow x \leq \varepsilon(y)$, for all $x \in L$ and $y \in M$.[1]

Assume that $\delta : \mathbf{L} \longrightarrow \mathbf{M}$ is a dilation. For $x \in L$, we can write $\delta(x) = \bigvee_{y \leq x} \delta(y)$, where we have used the fact that δ distributes over join. Every dilation defined on $\mathbf{L}$ is of the form above, and the adjoint erosion is given by $\varepsilon(y) = \bigvee_{\delta(x) \leq y} x$.

Now, keeping in mind the models $\mathbb{R}^n$ and $\mathbb{Z}^n$, it is possible to introduce the concepts of *translation* of an image and *translation invariance* of an operator, by means of the algebraic operation of sum. Indeed, let E be $\mathbb{R}^n$ or $\mathbb{Z}^n$ and consider the complete lattice $\mathscr{P}(E)$; given an element $h \in E$, we define the *h-translation* τ_h on $\mathscr{P}(E)$ by setting, for all $X \in \mathscr{P}(E)$, $\tau_h(X) = X + h = \{x + h \mid x \in X\}$, where the sum is intended to be defined coordinatewise.

Next, we consider the case of operators that are translation invariant; here the sets $\delta(\{x\})$ are translates of a fixed set, called the *structuring element*, by $\{x\}$. An operator $f : \mathscr{P}(E) \longrightarrow \mathscr{P}(E)$ is called *translation invariant*, *T-invariant* for short, if $\tau_h \circ f = f \circ \tau_h$ for all $h \in E$. It is proved in [4] that every T-invariant dilation on $\mathscr{P}(E)$ is given by $\delta_A(X) = \bigcup_{x \in X} A + x$, and every T-invariant erosion is given by $\varepsilon_A(X) = \{y \in E \mid A + y \subseteq X\} = \{y \in E \mid y \in X + \check{A}\}$, where A is

[1] Notice that the notation used in the literature of Mathematical Morphology is slightly different. Indeed, an adjunction is presented with a dilation in the second coordinate and an erosion in the first. Here we use such a reversed notation because, as we will see, adjunctions are adjoint pairs in the sense of Definition 1.

an element of $\mathscr{P}(E)$, called the *structuring element*, and $\breve{A} = \{-a \mid a \in A\}$ is the reflection of A around the origin.

Now we observe that the above expressions for erosion and dilation can also be written, respectively, as

$$\delta_A(X)(y) = \bigvee_{x \in E} A(y-x) \wedge X(x), \qquad \varepsilon_A(Y)(x) = \bigwedge_{y \in E} A(y-x) \to Y(y), \quad (3)$$

where each subset X of E is identified with its (Boolean) membership function and $X \to Y =: X^c \vee Y$. Moving from these expressions, and recalling that $\wedge$ is a residuated commutative operation (that is, a continuous t-norm) whose residuum is $\to$, it is possible to extend these operations from the complete lattice of sets $\mathscr{P}(E) = \{0,1\}^E$ to the complete lattice of fuzzy sets $[0,1]^E$, by means of left-continuous t-norms and their residua. What we do, concretely, is to extend the morphological image operators of dilation and erosion, from the case of binary images, to the case of grey-scale images.

So let $*$ be a left-continuous t-norm and $\to$ be its residuum; a grey-scale image X is a fuzzy subset of E. Given a fuzzy subset $A \in [0,1]^E$, called a *fuzzy structuring element*, the operators

$$\delta_A(X)(y) = \textstyle\bigvee_{x \in E} A(y-x) * X(x), \qquad \varepsilon_A(X)(x) = \bigwedge_{y \in E} A(y-x) \to X(y)$$

are, respectively, a translation invariant dilation and erosion on $[0,1]^E$.

3 Quantale Modules

Definition 6. Let $\mathbf{Q}$ be a quantale and $\mathbf{M} = \langle M, \vee, \bot \rangle$ a sup-lattice. $\mathbf{M}$ is a (left) $\mathbf{Q}$-*module* if there exists an external binary operation, called *scalar multiplication*, $\star : (q,m) \in Q \times M \longmapsto q \star m \in M$, such that

($M1$) $(q_1 \cdot q_2) \star m = q_1 \star (q_2 \star m)$, for all $q_1, q_2 \in Q$ and $m \in M$;
($M2$) the external product is distributive with respect to arbitrary joins in both coordinates or — that is the same — it is residuated;
($M3$) $e \star m = m$.

From ($M2$) it follows that, for all $q \in Q$, there exists the residual map $(q^\star)_*$ of $q^\star$, and for all $m \in M$ there exists the residual map $({}^\star m)_*$ of ${}^\star m$. In particular it is possible to define another external operation over M:

$$\backslash_\star : (q,m) \in Q \times M \longmapsto q \backslash_\star m = (q^\star)_*(m) \in M.$$

Example 7. Let $\mathbf{Q}$ be a quantale and X be an arbitrary non-empty set. We can consider the sup-lattice $\mathbf{Q}^X = \langle Q^X, \vee^X, \bot^X \rangle$, where $\bot^X$ is the $\bot$-constant function from X to Q and the join and the scalar multiplication $\star$ are defined pointwisely from those in $\mathbf{Q}$.

Then $\mathbf{Q}^X$ is a left $\mathbf{Q}$-module and, for all $q \in Q$, $f \in Q^X$ and $x \in X$, $(q \backslash_\star f)(x) = q \backslash f(x)$.

It is easy to show that the module in the previous example is the free $\mathbf{Q}$-module over the set of generators X.

Definition, and properties, of right $\mathbf{Q}$-modules are completely analogous. If $\mathbf{Q}$ is commutative, right and left $\mathbf{Q}$-modules coincide and we will say simply $\mathbf{Q}$-modules. If $\mathbf{Q}$ is a quantale and $\mathbf{M}$ is a left $\mathbf{Q}$-module, the dual sup-lattice $\mathbf{M}^{\mathrm{op}}$ is a right $\mathbf{Q}$-module (and vice versa) with the external multiplication $\backslash_\star$.[2]

Let $\mathbf{Q}$ be a quantale and $\mathbf{M}_1, \mathbf{M}_2$ be two $\mathbf{Q}$-modules. A map $f : M_1 \longrightarrow M_2$ is a $\mathbf{Q}$-module homomorphism if $f\left({}^{\mathbf{M}_1}\bigvee_{i\in I} m_i\right) = {}^{\mathbf{M}_2}\bigvee_{i\in I} f(m_i)$ for any family $\{m_i\}_{i\in I} \subseteq \mathbf{M}_1$, and $f(q \star_1 m) = q \star_2 f(m)$, for all $q \in Q$ and $m \in M_1$, where $\star_i$ is the external product of $\mathbf{M}_i$, for $i = 1, 2$.

Proposition 8. *Let $\mathbf{Q}$ be a quantale, $\mathbf{M}_1$, $\mathbf{M}_2$ be two $\mathbf{Q}$-modules and $f : \mathbf{M}_1 \longrightarrow \mathbf{M}_2$ be a homomorphism. Then f is a residuated map and the residual map $f_* : M_2 \longrightarrow M_1$ is a $\mathbf{Q}$-module homomorphism between $\mathbf{M}_2^{\mathrm{op}}$ and $\mathbf{M}_1^{\mathrm{op}}$.*

Definition 9. Let $\mathbf{M}$ and $\mathbf{N}$ be two $\mathcal{Q}$-modules and $\mathrm{Hom}_{\mathbf{Q}}(\mathbf{M}, \mathbf{N})$, the set of all the homomorphisms from $\mathbf{M}$ to $\mathbf{N}$. Then the structure $\mathbf{Hom}_{\mathbf{Q}}(\mathbf{M}, \mathbf{N}) = \langle \mathrm{Hom}_{\mathbf{Q}}(\mathbf{M}, \mathbf{N}), \sqcup, \bot^{\bot} \rangle$, with the pointwise join and the $\bot$-constant homomorphism as bottom element, is a sup-lattice; moreover, if $\mathbf{Q}$ is a commutative quantale, $\mathbf{Hom}_{\mathbf{Q}}(\mathbf{M}, \mathbf{N})$ is a $\mathbf{Q}$-module with the scalar multiplication $\diamond$ defined, again, pointwisely: for all $q \in Q$ and $h \in \mathrm{Hom}_{\mathbf{Q}}(\mathbf{M}, \mathbf{N})$, $q \diamond h$ is the homomorphism defined by $(q \diamond h)(x) = q \star h(x) = h(q \star x)$, for all $x \in M$.

If $\mathbf{N} = \mathbf{M}$, $\mathbf{Hom}_{\mathbf{Q}}(\mathbf{M}, \mathbf{M})$ is denoted by $\mathbf{End}_{\mathbf{Q}}(\mathbf{M}) = \langle \mathrm{End}_{\mathbf{Q}}(\mathbf{M}), \sqcup, \bot^{\bot} \rangle$.

4 $\mathcal{Q}$-Module Transforms

In this section we introduce the $\mathcal{Q}$-module transforms and we list some results about them. Then we present a classification of these operators that have interesting theoretical and concrete consequences. For an extensive treatment of $\mathcal{Q}$-module transforms, the reader may refer to [9].

Definition 10. Let $\mathbf{Q} \in \mathcal{Q}$ and X, Y be non-empty sets and let us consider the free $\mathbf{Q}$-modules $\mathbf{Q}^X$ and $\mathbf{Q}^Y$. We will call a *$\mathcal{Q}$-module transform* between $\mathbf{Q}^X$ and $\mathbf{Q}^Y$, with *kernel* p, the operator $H_p : Q^X \longrightarrow Q^Y$ defined by

$$H_p f(y) = \bigvee_{x\in X} f(x) \cdot p(x, y) \quad \text{for all } y \in Y,$$

where $p \in Q^{X\times Y}$. Its *inverse transform* $\Lambda_p : Q^Y \longrightarrow Q^X$ is the map defined by

$$\Lambda_p g(x) = \bigwedge_{y\in Y} g(y)/p(x, y) \quad \text{for all } x \in X.$$

Theorem 11. *Let $\mathbf{Q} \in \mathcal{Q}$, X, Y be two non-empty sets and $p \in Q^{X\times Y}$. If H_p is the $\mathcal{Q}$-module transform, with kernel p, between $\mathbf{Q}^X$ and $\mathbf{Q}^Y$, and Λ_p is its inverse transform, then the following hold:*

[2] In what follows, in all the definitions and results that can be stated both for left and right modules, we will refer generically to "modules" — without specifying left or right — and we will use the notations of left modules.

(*i*) (H_p, Λ_p) *is an adjoint pair, i.e.* H_p *is a residuated map and* $\Lambda_p = H_{p_*}$;
(*ii*) $H_p \in \mathrm{Hom}_{\mathbf{Q}}\left(\mathbf{Q}^X, \mathbf{Q}^Y\right)$ *and* $\Lambda_p \in \mathrm{Hom}_{\mathbf{Q}}\left(\left(\mathbf{Q}^Y\right)^{\mathrm{op}}, \left(\mathbf{Q}^X\right)^{\mathrm{op}}\right)$.

The following classification of the kernels has a few interesting theoretical implications but it is important for applications to image processing. We refer to [1] (where an orthonormal transform is presented), [8] and [9] for details.

Definition 12. Let $\mathbf{Q} \in \mathcal{Q}$, and X, Y be non-empty sets. Let us consider a map $p \in Q^{X \times Y}$; we set the following definitions:

(*i*) p is called a *coder* iff there exists an injective map $\varepsilon : Y \longrightarrow X$ such that $e \leq p(\varepsilon(y), y)$ for all $y \in Y$;
(*ii*) p is said to be *normal* iff there exists an injective map $\varepsilon : Y \longrightarrow X$ such that $p(\varepsilon(y), y) = e$ for all $y \in Y$;
(*iii*) p is said to be *strong* iff it is normal and $p(\varepsilon(y_1), y_2) = \bot$ for all $y_1 \neq y_2 \in Y$;
(*iv*) p is said to be *orthogonal* iff $p(x, y_1) \cdot p(x, y_2) = \bot$ for all $y_1, y_2 \in Y$ such that $y_1 \neq y_2$ and for all $x \in X$;
(*v*) p is said to be *orthonormal* iff it is orthogonal and normal.

If p is a coder, the $\mathcal{Q}$-module transform H_p is called *faithful* and, if p is normal, strong, orthogonal or orthonormal, the corresponding transform will have the same adjective. Also, we observe that $(v) \Longrightarrow (iii) \Longrightarrow (ii) \Longrightarrow (i)$.

Theorem 13. *Let* $\mathbf{Q} \in \mathcal{Q}$ *and let* H_p *be a* $\mathcal{Q}$*-module strong transform, by the coder* $p \in Q^{X \times Y}$*, with inverse transform* Λ_p*. Then* $H_p \circ \Lambda_p = \mathrm{id}_{Q^Y}$*; thus* H_p *is onto and, by Proposition 2,* Λ_p *is one-one.*

Lemma 14. *Let* $\mathbf{Q} \in \mathcal{Q}$*,* X *be a non-empty set,* Y *be a non-empty subset of* X *and* $p, p' \in Q^{X \times Y}$ *be two maps. Then* $H_p = H_{p'}$ *if and only if* $p = p'$.

The previous result ensures us that a $\mathcal{Q}$-module transform H_p is completely determined by its kernel p, while next result is the converse of Theorem 11(*ii*); it proves that all the homomorphisms between free modules are transforms.

Theorem 15. *The sup-lattices* $\mathbf{Hom}_{\mathbf{Q}}(\mathbf{Q}^X, \mathbf{Q}^Y)$ *and* $\mathbf{Q}^{X \times Y}$ *are isomorphic. And, if* $\mathbf{Q}$ *is commutative, they are isomorphic also as* $\mathbf{Q}$*-modules. In particular* $\mathbf{End}_{\mathbf{Q}}(\mathbf{Q}^X) \cong \mathbf{Q}^{X \times X}$.

Here we just "scratched the surface" of $\mathcal{Q}$-module transforms, especially in order to focus the attention on the main thesis of the present paper, i.e. the fact that fuzzy image compression and mathematical morphological operators fall within the same class of operators under an algebraic point of view. We, again, refer to [8] or [9] the reader who may be interested in $\mathcal{Q}$-module transforms.

5 Conclusion

As the reader may have already noticed, the operators in Section 2 are all special cases of $[0, 1]$-module transforms. We will now analyse them in detail.

Let us consider the compression operator defined in Subsection 2.1. Its domain is $[0,1]^{m\times n}$ and its codomain is $[0,1]^{a\times b}$; in the light of the definitions and results presented so far, we get immediately that (1) is the $[0,1]$-module transform $H_C : [0,1]^{m\times n} \longrightarrow [0,1]^{a\times b}$ with kernel $C \in [0,1]^{m\times n\times a\times b}$ and the reconstruction (2) is its inverse transform $\Lambda_C : [0,1]^{a\times b} \longrightarrow [0,1]^{m\times n}$.

We already observed in Subsection 2.2 that dilations are precisely the suplattice homomorphisms while erosions are their residua. In order to faithfully represent dilations and erosions that are translation invariant as $\mathcal{Q}$-module transforms from a free $[0,1]$-module to itself, we make the further assumption that the set over which the free module is defined has the additional structure of Abelian group. So let $\mathbf{X} = \langle X, +, -, 0\rangle$ be an Abelian group, $*$ a t-norm on $[0,1]$ and consider the free $[0,1]$-module $[0,1]^X$. For any element $k \in [0,1]^X$, we define the two variable map $\overline{k} : (x,y) \in X \times X \longmapsto k(y-x) \in [0,1]$. Then, for all $k \in [0,1]^X$, the translation invariant dilation, on $[0,1]^X$, whose structuring element is k, is precisely the $\mathcal{Q}$-module transform $H_{\overline{k}} : [0,1]^X \longrightarrow [0,1]^X$, with the kernel $\overline{k}$ defined above. Obviously, the translation invariant erosion whose structuring element is k is $\Lambda_{\overline{k}}$.

Then the representation of both classes of operators as $\mathcal{Q}$-module transforms is rather trivial. Actually, what we want to point out here is that, if we drop the assumption that our quantale is defined on $[0,1]$, the classes of transforms defined in this section become much wider. The purpose of this consideration is not to suggest purely speculative abstractions but, rather, to underline that suitable generalizations of these operators exist already and they may be useful provided their underlying ideas are extended to tasks involving other quantales.

References

1. Di Nola, A., Russo, C.: Łukasiewicz Transform and its application to compression and reconstruction of digital images. Informat. Sci. 177(6), 1481–1498 (2007)
2. Di Nola, A., Sessa, S., Pedrycz, W., Sanchez, E.: Fuzzy relation equations and their applications to knowledge engineering. Kluwer, Dordrecht (1989)
3. Goutsias, J., Heijmans, H.J.A.M.: Fundamenta Morphologicae Mathematicae. Fundamenta Informaticae 41, 1–31 (2000)
4. Heijmans, H.J.A.M.: Morphological Image Operators. Ac. Press, Boston (1994)
5. Hirota, K., Pedrycz, W.: Fuzzy relational compression. IEEE Trans. Syst. Man Cyber. – Part B 29(3), 407–415 (1999)
6. Matheron, G.: Random Sets and Integral Geometry. J. Wiley & Sons, N.Y (1975)
7. Nobuhara, H., Takama, Y., Hirota, K.: Image compression/reconstruction based on various types of fuzzy relational equations. Trans. Inst. Elec. l Eng. Japan 121-C(6), 1102–1113 (2001)
8. Russo, C.: Quantale Modules, with Applications to Logic and Image Processing, Ph.D. Thesis, University of Salerno, Italy (2007)
9. Russo, C.: Quantale Modules and their Operators, with Applications. J. Logic Comput. (to appear, 2009) doi:10.1093/logcom/exn088
10. Serra, J.: Image Analysis and Mathematical Morphology. Ac. Press, London (1982)

Adaptive Image Watermarking Approach Based on Kernel Clustering and HVS

Hong Peng[1,2], Jun Wang[3], and Weixing Wang[1]

[1] School of Electronic Engineering, University of Electronic Science & Technology of China, Chengdu, Sichuan, 610054, China
[2] School of Mathematics & Computer Engineering, Xihua University, Chengdu, Sichuan, 610039, China
[3] School of Electrical Information Engineering, Xihua University, Chengdu, Sichuan, 610039, China
ph66@tom.com

Abstract. In this paper, an adaptive image watermarking approach is introduced, which consists of kernel fuzzy c-means (KFCM) clustering algorithm and human visual system (HVS). Firstly, the host image is divided into image blocks and block-wise DCT transform is accomplished. Then, three local features of image blocks are extracted from its DCT coefficients, and these features are used to train KFCM in order to select the embedding position and determine the embedding strength of image blocks adaptively. The experimental results show the proposed algorithm is robust to common attacks such as JPEG, filtering, noise addition, scaling, sharpen, etc.

1 Introduction

Digital image watermarking technique provides copyright protection of image data by hiding appropriate information in the original image. The embedded watermark should be robust and imperceptible, but the ways of pursuing imperceptibility and robustness are conflict. It is an important issue to find a fair balance between imperceptibility and robustness. In recent years, a number of adaptive watermarking methods for digital image have been proposed. Especially, the adaptive watermarking technique for incorporating the features of the human visual system (HVS) model can provide an excellent solution.

Recently, intelligent algorithms, such as neural networks, support vector machines and fuzzy methods, are introduced into digital watermarking technique, and can simultaneously improve robustness and visual quality of the watermarked image [1,2,3,4,5,6,7]. In [3], Lou et al. proposed an adaptive watermarking method using fuzzy logic technique. Wu et al. [4] proposed a digital watermarking scheme in DCT domain based on fuzzy clustering technique, which can adaptively control the embedding strength of different blocks. Sakr et al. [5] proposed an adaptive image watermarking algorithm based on dynamic fuzzy inference system. In [6], Chang et al. proposed a Fuzzy-ART based adaptive

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 213–220, 2009.

digital watermarking scheme in DCT domain. In [7], Wang et al. presented a DWT-based robust watermarking scheme with Fuzzy-ART.

In this paper, we propose a robust adaptive watermarking approach, which incorporates the features of the human visual system (HVS) and kernel fuzzy c-means (KFCM) clustering algorithm. Firstly, we divide image into non-overlapping image blocks with size 8×8. And we extract the local features of image blocks, that is, local luminance feature, local texture feature and the frequency feature of image blocks. Secondly, By running KFCM on these features, we obtain the maximal fuzzy membership degree for each image block. According the fuzzy membership degree, we can select the embedding position and determine the embedding strength of image block adaptively. The proposed approach can simultaneously improve the robustness and visual quality of the the watermarked image.

2 Preliminaries

2.1 Kernel Fuzzy C-Means Algorithm

Given an unlabeled data set $X = \{x_1, x_2, \ldots, x_m\} \subseteq R^d$, and a mapping $\phi : R^d \to F$, the kernel fuzzy c-means algorithm in the feature space F by a mapping ϕ minimizes the function J_r [9,10]:

$$J_r(X) = \sum_{i=1}^{c}\sum_{j=1}^{m}(\mu_{ij})^r \|\phi(x_j) - v_i^{\phi}\|^2, \tag{1}$$

where μ_{ij} is the membership degree of data point x_j to the ith fuzzy cluster, and r is a fuzziness coefficient. The ith cluster centroid is $v_i^{\phi} = n_i^{-1}\sum_{j=1}^{m}(\mu_{ij})^r\phi(x_j)$ and $n_i = \sum_{j=1}^{m}(\mu_{ij})^r$. The key notion in the kernel fuzzy c-means algorithm lies in the calculation of the distance in the feature space. The distance between $\phi(x_j)$ and v_i^{ϕ} in the feature space is calculated through the kernel in the input space:

$$\begin{aligned}
\|\phi(x_j) - v_i^{\phi}\|^2 &= \phi(x_j)\cdot\phi(x_j) - 2\phi(x_j)\cdot\frac{\sum_{k=1}^{m}(\mu_{ik})^r\phi(x_k)}{\sum_{k=1}^{m}(\mu_{ik})^r} \\
&\quad + \frac{\sum_{k=1}^{m}(\mu_{ik})^r\phi(x_k)}{\sum_{k=1}^{m}(\mu_{ik})^r}\cdot\frac{\sum_{l=1}^{m}(\mu_{il})^r\phi(x_l)}{\sum_{l=1}^{m}(\mu_{il})^r} \\
&= \phi(x_j)\cdot\phi(x_j) - \frac{2\sum_{k=1}^{m}(\mu_{ik})^r\phi(x_k)\cdot\phi(x_j)}{\sum_{k=1}^{m}(\mu_{ik})^r} \\
&\quad + \frac{\sum_{k=1}^{m}\sum_{l=1}^{m}(\mu_{ik})^r(\mu_{il})^r\phi(x_k)\cdot\phi(x_l)}{(\sum_{k=1}^{m}(\mu_{ik})^r)^2} \\
&= K(x_j, x_j) - \frac{2\sum_{k=1}^{m}(\mu_{ik})^r K(x_k, x_j)}{\sum_{k=1}^{m}(\mu_{ik})^r} \\
&\quad + \frac{\sum_{k=1}^{m}\sum_{l=1}^{m}(\mu_{ik})^r(\mu_{il})^r K(x_k, x_l)}{(\sum_{k=1}^{m}(\mu_{ik})^r)^2},
\end{aligned} \tag{2}$$

where $K(x_k, x_l) = \phi(x_k) \cdot \phi(x_l)$. By using $n_i = \sum_{j=1}^{m} (\mu_{ij})^r$, we have

$$\|\phi(x_j) - v_i^{\phi}\|^2 = K(x_j, x_j) - \frac{2}{n_i} \sum_{k=1}^{m} (\mu_{ik})^r K(x_k, x_j) + \frac{1}{n_i^2} \sum_{k=1}^{m} \sum_{l=1}^{m} (\mu_{ik})^r (\mu_{il})^r K(x_k, x_l). \tag{3}$$

Therefore, the objective function can be rewritten as follows:

$$J_r(X) = \sum_{i=1}^{c} \sum_{j=1}^{m} (\mu_{ij})^r \left(K(x_j, x_j) - \frac{2}{n_i} \sum_{k=1}^{m} (\mu_{ik})^r K(x_k, x_j) = \frac{1}{n_i^2} \sum_{k=1}^{m} \sum_{l=1}^{m} (\mu_{ik})^r (\mu_{il})^r K(x_k, x_l) \right). \tag{4}$$

The kernel fuzzy c-means algorithm iteratively updates the new membership degree μ_{ij} at each iteration. The update of μ_{ij} in the feature space is defined through the kernel in the input space as follows:

$$\mu_{ij} = \left(\sum_{k=1}^{c} \left(\frac{\|\phi(x_j) - v_i^{\phi}\|^2}{\|\phi(x_j) - v_k^{\phi}\|^2} \right)^{1/(r-1)} \right)^{-1}. \tag{5}$$

From Eq.(3), the kernel fuzzy c-means algorithm does not need to calculate the cluster centroids because the centroid information is considered in updating the membership degree μ_{ij}.

The proposed kernel fuzzy c-means algorithm can be summarized in the following steps:

Step 1: Fix c, t_{max}, $r > 1$ and $\varepsilon > 0$ for some positive constant.
Step 2: Initialize the memberships μ_{ij}^0.
Step 3: For $t = 1, 2, \ldots, t_{max}$, do:
(a) Update all memberships μ_{ij}^t with Eq.(5);
(b) Compute $E_t = max_{ij} |\mu_{ij}^t - \mu_{ij}^{t-1}|$, if $E_t \leq \varepsilon$, stop;
end;

In this paper, we will use kernel fuzzy c-means technique to classify image blocks, and select more suitable image blocks which have larger the fuzzy membership degree μ_{ij} to embed the watermark.

2.2 The HVS Model

Generally, the watermark is a weak signal which is embedded into a strong background, and if it is lower than JND (Just Noticeable Difference), the watermark is unable to sense by human eye. So, the watermark strength can be determined by HVS. In HVS model, a number of factors affect the noise sensitively of the eye like luminance, texture and frequency. In order to better fit the behavior of

the HVS to the watermarking problem, the following considerations have been taken into account [5].

The first term takes into account the local luminance of image. Human eye is less sensitive to the areas of image, where the brightness is the too high or too low. The luminance sensitivity can be taken into account as follows:

$$L_k = \left(\frac{C_{DC,k}}{\bar{C}_{DC}} \right)^{\gamma}, \tag{6}$$

where $C_{DC,k}$ is the DC coefficient of the DCT of the kth image block, $\bar{C}_{DC}$ is the mean value of all $C_{DC,k}$ coefficients of a image, and γ is set to 0.649 to control the degree of luminance sensitivity as in [5].

The second term takes into account the local texture of the image. The stronger the texture, the lower the visibility of the embedded signal. Thus, a stronger signal can be embedded. Texture sensitivity can be estimated by quantizing the DCT coefficients of an image block using the JPEG quantization table. The result is the rounded to the the nearest integers. The texture sensitivity can be calculated by the following formula (as in [5]):

$$T_k = \sum_{i,j=1}^{m} cond\left(\left[\frac{C_k(i,j)}{Q(i,j)} \right] \right), \tag{7}$$

where $cond(x)$ takes the rounded value of x, and return "1" if the value is not equal to zero, "0" otherwise.

The third term takes into account the frequency of the image. The frequency sensitivity can be defined as the response of a human perceptual system to differences in the frequencies of stimuli. High frequencies are less visible to the human eye, while in the basic visual model the variance of low frequencies is more sensitive. The frequency sensitivity F_k is represented by the JPEG quantization table (luminance).

In this paper, for each image block, we extract its three features, luminance feature L_k, texture feature T_k and frequency F_k, and form a vector $x_k = (L_k, T_k, L_k)$ which is used as the input vector of kernel fuzzy c-means method.

3 Proposed Watermarking Approach

3.1 Watermark Embedding

Let host image I be a gray-scale image with size $M_1 \times M_2$, and watermark W be a binary logo image with size $N_1 \times N_2$. The watermark embedding process is described as follows:

Step 1. Dividing image.

The host image I is divided into non-overlapping image blocks with size 8×8. Let I_k notes k-th image block, $k = 1, 2, \ldots, m$, where $m = [(M_1 \times M_2)/(8 \times 8)]$. Then, each image block I_k is transformed by DCT independently, and its coefficient block is denoted by C_k.

Step 2. Calculating the local features of image blocks.

For each image block I_k, we calculate its three features, luminance feature L_k, texture feature T_k and frequency F_k, and form a vector $x_k = (L_k, T_k, L_k)$ which is used as the input vector of kernel fuzzy c-means method.

Step 3. Preprocessing by kernel fuzzy c-means clustering.

After calculating the local features of all image blocks, we obtain a data set $D = \{x_1, x_2, \ldots, x_m\}$. Let c is the number of clustering. We run the kernel fuzzy c-means algorithm, and then obtain the fuzzy membership degrees μ_{ik}, $1 \leq i \leq c, 1 \leq k \leq m$. Thus, for each image block I_k, we obtain the corresponding fuzzy membership degrees $\mu_k = \max\{\mu_{ik} \mid 1 \leq i \leq c\}$, $1 \leq k \leq m$.

Step 4. Selecting the embedding position.

Firstly, we select m_1 image blocks from the m image blocks, which has larger fuzzy membership degrees μ_k. Then, for each selected image block, we select two low-middle frequency coefficients as the embedding position, which is larger AC coefficient. These selected coefficients are denoted by $V = \{c_i \mid i = 1, 2, \ldots, n\}$, where $n = 2 \times m_1$, and $n \geq N_1 \times N_2$.

Step 5. Embedding the watermark.

For each coefficient in V, watermark embedding is accomplished by using following rule:

$$c_i' = c_i(1 + \mu_k \cdot \beta \cdot w_i) \tag{8}$$

where c_i is original coefficient, and c_i' is the coefficient of watermarked image. β is a constant, and μ_k is the fuzzy membership degrees of kth image block where c_i is the coefficient of kth image block.

Step 6. Lastly, each image block is reconstructed by applying the inverse DCT transform respectively, and then all image blocks are combined into the final watermarked image I'.

3.2 Watermark Extraction

The watermark extraction process is described as follows:

Step 1. Dividing image.

The input image I' is partitioned into non-overlapping image blocks with size 8×8. Let I_k' notes k-th image block, $k = 1, 2, \ldots, m$. Then, each image block I_k' is transformed by DCT independently, and its coefficient block is denoted by C_k'.

Step 2. Calculating the local features of image blocks.

For each image block I_k', we calculate its three features, luminance feature L_k', texture feature T_k' and frequency F_k', and form a vector $x_k' = (L_k', T_k', L_k')$.

Step 3. Preprocessing by kernel fuzzy c-means clustering.

After calculating the local features of all image blocks, we obtain a data set $D' = \{x_1', x_2', \ldots, x_m'\}$. By use the kernel fuzzy c-means algorithm in the same manner as in the embedding process, we obtain the fuzzy membership degrees μ_k of each image block I_k, $1 \leq k \leq m$.

Step 4. Extracting the watermark.

We select the embedding position of the watermark in the same manner as in the embedding process, and the selected coefficients are denoted by $V' = \{c'_i \mid i = 1, 2, \ldots, n\}$. The watermark extraction can be described by the following equation:

$$w_i = \frac{(c'_i - c_i)}{(\mu_k \cdot \beta \cdot c_i)} \tag{9}$$

where $i = 1, 2, \ldots, N_1 \times N_2$.

Step 5. Lastly, the one-dimensional sequence $w_1 w_2 \ldots w_{N_1 \times N_2}$ is converted into the two-dimensional logo watermark image W'.

4 Experimental Results

In our experiment, the proposed method has been extensively tested on various standard images under different kinds of attack. Due to the limitation of space, here we only demonstrate the experimental results on "Lena" image with size 256×256, shown in Fig.1(a). The watermark is a binary "Beijing Olympic" logo image with size 32×32, shown in Fig.1(b). Firstly, we should decide some parameters in the experiment. For KFCM, we select Gaussian kernel as the kernel function of KFCM, and its the width parameter $\sigma = 200$. Set the number of clustering $c = 256$, and the constant $\beta = 0.3$.

Fig.1(c) depicts the watermarked image, which has PSNR value of 43.635dB. If the original and the watermarked images are observed, we cannot find any perceptual degradation. When there is no attack, the watermark can be extracted without a bit errors, shown in Fig.1(d).

Yi *et al.* [8] has proposed that the original image is decomposed into non overlapping blocks, and all blocks are classified into four categories by using the human visual system. Then, according to the classification, the watermark components with different strength are embedded into the original image. The comparative results with Yi's method against different attacks are detailedly shown as follows. When the image is median filtered, the watermark extracted by our method has only 6 bit errors, shown in Fig.2(a), whereas the watermark extracted by Yi's method has totally about 19 bit errors (Fig.2(b)). After average filtering, our method can extract the watermark with bit errors 65 (Fig.2(c)), and the watermark extracted by Yi's method has about 144 bit errors (Fig.2(d)). When the watermarked image is rotated to the left 75^o, the images are greatly degraded, but our method can extract the watermark with bit errors 138 (Fig.2(e)), whereas the watermark extracted by Yi's method has about 211 bit errors, which shown in Fig.2(f). To test the robustness against adding noise, adding 5% salt and pepper noise randomly degrades the watermarked image, the performance of the propose method has low satisfaction, but the watermark can be recognized, shown in Fig.2(g). However, the watermark extracted by Yi's method is difficult to be recognized (Fig.2(h)). The proposed method has been extensively tested under other various attacks. Results of comparing with Yi's are illustrated in Table 1, in which MAE (the Mean Absolute Error) is measurement of the difference between an original watermark and the

Fig. 1. (a)Original image; (b)Original watermark; (c)Watermarked image; (d)Extracted watermark

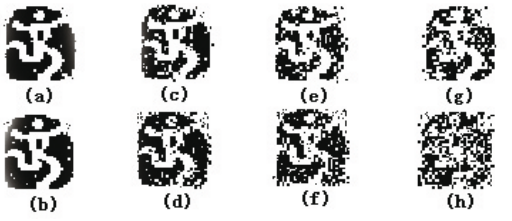


Fig. 2. The extracted binary watermark image after (a) median filtering by the proposed method, (b) median filtering by Yi's method (c) average filtering by the proposed method, (d)average filtering by Yi's method, (e) rotating to the left 75^o by the proposed method,(f) rotating to the left $75^o$by Yi's method, (g) 5% noise adding by the proposed method, (h) 5% noise adding by Yi's method

Table 1. The experimental results comparison under different attacks

Attacks	PSNR (dB)	The proposed method MAE	Yi's method MAE
Bluring	39.062	0	0.0872
Sharpen	33.853	0.0385	0.0703
Scaling(75%)	39.109	0	0.0171
Scaling(50%)	33.986	0	0.0279
gaussian noise(5%)	24.584	0.4794	0.1224
gaussian filter(3×3)	41.761	0.0108	0.0430
JPEG (QF=80)	38.235	0.0061	0.0725
JPEG (QF=60)	36.792	0.2235	0.3423

corresponding extracted watermark. These extensive experiments show that our technique possesses significant robustness against the various attacks.

5 Conclusion

In this paper, we proposed an adaptive image watermarking approach based on kernel fuzzy c-means (KFCM) clustering algorithm and human visual system (HVS). We extract three local image features from image blocks, which can well reflect the visual characteristic of image. According to these image features,

KFCM is used to select the embedding position and determine the embedding strength of image block adaptively. The proposed approach can simultaneously improve robustness and visual quality of the watermarked images. The experimental results show that the proposed method possess significant robustness against the various attacks.

Acknowledgements

This work is supported by the importance project foundation of the education department of Sichuan province (No.2007ZA112), the training fund of science and technology of Sichuan province (No.08209057), China, and the importance project foundation of Xihua University (No.ZG0722603).

References

1. Yu, P.T., Tsai, H.H., Lin, J.S.: Digital Watermarking Based on Neural Networks for Color Images. Signal Processing, 663–671 (2001)
2. Tsai, H.H., Sun, D.W.: Color Image Watermark Extraction Based on Support Vector Machines. Information Sciences 177(2), 550–569 (2007)
3. Lou, D.C., Yin, T.L.: Adaptive Digital Watermarking Using Fuzzy Logic Techniques. Optical Engineering 41, 2675–2687 (2002)
4. Wu, J.-Z., Xie, J.-Y.: Adaptive Image Watermarking Scheme Based on HVS and Fuzzy Clustering Theory. In: IEEE Int. Conf. Neural Networks & Signal Processing 2003, vol. 2, pp. 1493–1496 (2003)
5. Sakr, N., Zhao, J.-Y., Croza, V.: A Dynamic Fuzzy Logic Approach to Adaptive HVS-based Watermarking. In: IEEE International Workshop on Haptic Audio Visual Environments and Their Appplications 2005, pp. 121–126 (2005)
6. Chang, C.-H., Ye, Z., Zhang, M.-Y.: Fuzzy-ART Based Adaptive Digital Watermarking Schemen. IEEE Trans. on Circuits and Systems for Video Technology 25(1), 65–81 (2005)
7. Wang, H.-J., Chang, C.Y., Pan, S.-W.: A DWT-based Robust Watermarking Scheme with Fuzzy-ART. In: 2006 International Joint Conference on Neural Networks, pp. 16–21 (2006)
8. Yi, K.X., Shi, J.Y.: Adaptive 2-Dimension Image Watermarking Algorithm. Journal of Image and Graphics 6, 444–449 (2001)
9. Kima, D.-W., Leeb, K.Y., Leea, D., Leea, K.H.: Evaluation of The Performance of Clustering Algorithms in Kernel-induced Feature Space. Pattern Recognition 38, 607–611 (2005)
10. Liu, J.W., Xu, M.Z.: Kernelized Fuzzy Attribute C-means Clustering Algorithm. Fuzzy Sets and Systems 159, 2428–2445 (2008)

An Automatic Three-Dimensional Fuzzy Edge Detector

Marco Cipolla, Fabio Bellavia, and Cesare Valenti

Dipartimento di Matematica e Applicazioni
Università degli Studi di Palermo, Italy
{mcipolla,fbellavia,valenti}@unipa.it

Abstract. Three-dimensional object analysis is of particular interest in many research fields. In this context, the most common data representation is boundary mesh, namely, 2D surface embedded in 3D space. We will investigate the problem of 3D edge extraction, that is, salient surface regions characterized by high flexure. Our automatic edge detection method assigns a value, proportional to the local bending of the surface, to the elements of the mesh. Moreover, a proper scanning window, centered on each element, is used to discriminate between smooth zones of the surface and its edges. The algorithm does not require input parameters and returns a set of elements that represent the salient features on the model surface. This method is general enough, returns representative structures of the object, as edges, and can be considered as a pre-processing step for further applications, such as 3D compact representation, matching and recognition.

Keywords: Surface segmentation, automatic three-dimensional edge detection.

1 Introduction

Edge detection is one of the most important tasks of any vision system since it reduces significantly the amount of data, still preserving the relevant structural properties, and provides strong visual clues that can help the recognition process. This problem is still challenging in the three-dimensional space, since it requires geometrical information.

There is a wide range of 3D acquisition modalities, such as laser scan, multiple view and shape from shading that require different data representations to handle properly the object surface [1]. A triangular mesh is defined as $M = \{V, E, F\}$, where V, E, and F represent the sets of *vertices*, *edges* and *facets* of the model surface, respectively. 3D model analysis is a general concept and the final segmentation depends on the specific application; a description of the issue and general techniques can be found in [2], where the segmentation problem is formulated as a constrained optimization task.

This paper describes a novel three-dimensional edge detector, based on a fuzzy approach. The proposed algorithm computes the so called *saliency* of each arc on the surface, through a fuzzy membership defined on a continuous domain. This domain is automatically generated and the *fuzzification* process infers a natural segmentation of the surface, useful for locating contours and edge-type features to represent the objects.

Section 2 briefly reviews the state of the art algorithms for edge detection and illustrates our methodology. Section 3 reports conclusions on the experimental results and introduces possible future works.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 221–228, 2009.

2 3D Edge Detection

Mesh segmentation algorithms can help in locating the edges on the surface of the objects: these contours should lie on the zones that separate different regular surfaces. A region growing strategy is reported in [3], while two watershed approaches are proposed in [4,5]. Unfortunately, these methods require thresholding and are not fully automatic.

The algorithm presented in [6] locates line-type features representing edges and ridges, which can be used as a starting point for further segmentation procedures. This technique requires several control parameters to be set and the selection of the operator that must be applied on the mesh. The algorithm performs two phases: the *classification* to assign a curvature weight to each arc and the *selection* of the candidate edges by a standard or an hysteresis threshold method. The result are sets of edges (called *patches*), that are reduced to line-type features by a successive *thinning* procedure.

Mean curvature flow and anisotropic diffusion are used in [7], where the authors give a theoretic formulation for fairing and segmenting 3D surfaces. Nevertheless, this method, in order to perform the segmentation stage, needs the tuning of a threshold parameter and a pre-filtering step for a more reliable computation of the shape operator.

According to the related literature, we propose here an heuristic approach, able to extract salient features from the surface of an object. Our algorithm relies on a weighting function that assigns a measure to the elements of the mesh, proportional to the local bending degree of the surface. Different weighting functions, namely *Second Order Difference Operator*, *Extended Second Order Difference Operator* [6] and *tensor and normal voting* [5], have been introduced and a comparison of their performance/complexity has been already reported in [5,6]. These operators require the setting of parameters (e.g. the number of elements within the weighting window) that are very dependent on the input mesh and that must be manually set. Obviously, this process is time-consuming, subjective and error-prone. Moreover, real objects can be acquired at different samplings by laser scanner devices and the resulting mesh can be modified by using simplification techniques [8]. From this point of view, an automatic methodology, able to return stable three-dimensional segmentations, is highly desirable.

Given an arc $a \in E$, let $L_i(a) = \{a' : \text{such that the length of the minimum path from } a \text{ to } a' \text{ is } i\}$, that is $L_0(a) = a$ and $L_0(a)$, $L_1(a)$, $L_2(a)$,... are disjointed layers. For the sake of simplicity, each $L_i(a)$ is the i-th layer of the breadth first search tree with root a. We define the neighborhood $N_r(a)$ with radius r of a as $N_r(a) \triangleq \bigcup_{i=0}^{r} L_i(a)$, as in figure 1. Later on, the area covered by $N_r(a)$ (i.e. the sum of the areas of the triangles in $N_r(a)$) will be indicated by $A_r(a)$.

According to the second order difference operator, the weight $w(a)$ of the arc a, shared by two facets f_1 and f_2, with normal versors n_1 and n_2 respectively, (see figure 2a) is equal to the dihedral angle between them:

$$w(a) = \arccos(n_1 \cdot n_2)$$

where $(\cdot)$ indicates the usual inner product.

In the case of the extended second order difference operator, the normal versors n_1 and n_2 are computed as average normal vectors to the facets that share the vertices, $v_1 \in f_1$ and $v_2 \in f_2$, opposite to a (see figure 2b).

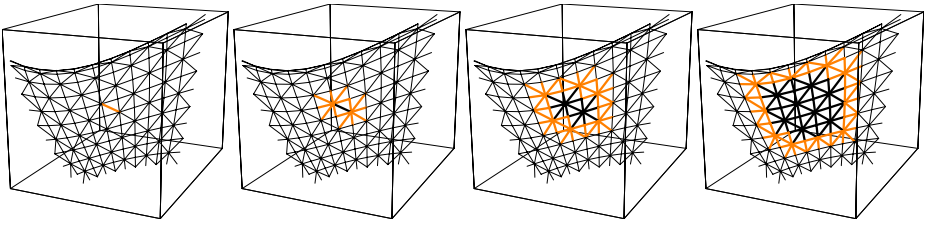


Fig. 1. The first 4 neighborhoods $N_0(a)$, $N_1(a)$, $N_2(a)$ and $N_3(a)$, obtained by visiting the arcs according to a breadth first search. New layers have been marked in orange

Our method classifies a, according to the distribution of w, with respect to the neighborhoods $N_r(a)$, of increasing radius r: from an ideal point of view, an edge separates different smooth surfaces which, on average, show a decreasing trend for w. Vice versa, arcs close to edges, namely *ramps*, have an increasing trend for w. Smooth surfaces, which are zones without edges, are characterized by a constant trend of w.

Let $\nu_r(a)$ be the variance of the weights w relative to the arcs that fall into the window $N_r(a)$, centered on a and with radius r. We define the saliency $s(a)$, to be assigned to a, as:

$$s(a) = \sum_{i>0} \phi_i(a) \quad \text{with} \quad \phi_i(a) = \eta_i(a) \frac{\partial \nu_i(a)}{\partial r} |\nu_1(a) - \nu_i(a)|$$

where $\eta_i(a) = e^{A_{\min} - A_i(a)}$ is used to lower the resulting values ϕ_i, when moving far from a, with respect to the smallest window of radius $r = 1$, that is $A_{\min} = \min_{a \in E} \{A_1(a)\}$.

Though the whole mesh should be scanned to compute s, it can be seen that, due to the effect of η, the contribution of large radii is close to zero and can be neglected. The normalization coefficient η avoids to consider structures too far from the center a of the neighborhood and ensures a coherent saliency even in the case of meshes with variable samplings which results in facets with different dimensions. Often, this is due to techniques applied to decimate the facets, while preserving the overall appearance of the object [8]. Moreover, it must be noted that the resolution of a mesh can be *normalized* by properly setting the length of its arcs [9].

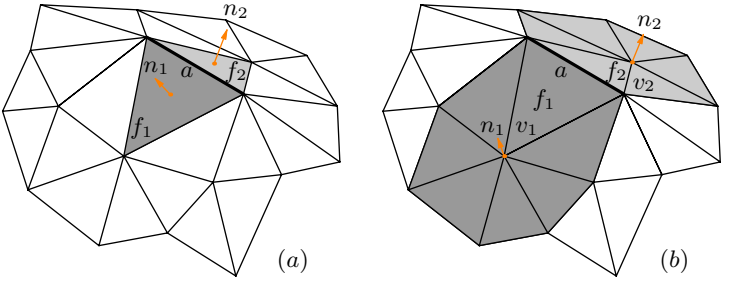


Fig. 2. (a) SOD: facets f_1 and f_2, with versors n_1 and n_2, share the bolded arc a. (b) ESOD: versors n_1 and n_2 are computed as average normals of the *other* facets adjacent to f_1 and f_2.

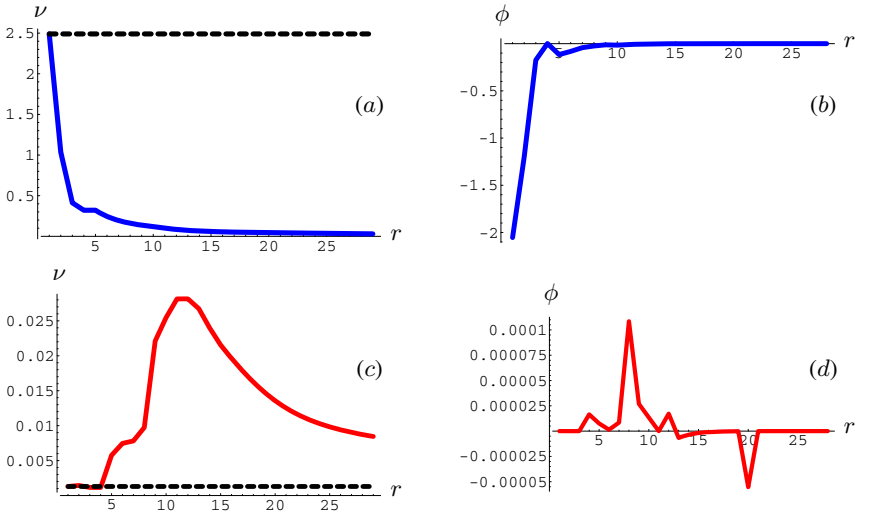


Fig. 3. Typical trends of variance ν_i (a,c) and score ϕ_i (b,d) for an arc, classified as edge (a,b) or ramp (c,d), respectively. The dashed line indicates the value ν_1, used as reference mark by ϕ_i.

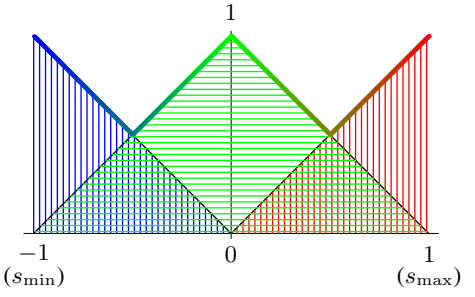


Fig. 4. Color representation of the fuzzy sets S_{E} (edges, in blue), S_{S} (smooth surfaces, in green) and S_{R} (ramps, in red). Membership to the sets has been represented by color gradients (line in bold). Author's note: coloured original image.

The derivative $\frac{\partial \nu_i(a)}{\partial r} = \nu_i(a) - \nu_{i-1}(a)$ measures how fast the variance changes. In particular, sharp edges are characterized by a fast decreasing f and, vice versa, smooth details have slower decreasing f. The term $|\nu_1(a) – \nu_i(a)|$ further exalts the changes of f, with respect to the initial value ν_1, since we are interested in the comparison between the smallest neighborhood $N_1(a)$ and the arcs in a variety of $N_{r>1}(a)$. This term is needed because it is possible for two given f trends to have locally similar derivatives, though different values of f should suggest different values of s (see figure 3).

Actually, negative values of s correspond to edges, positive values indicate ramps, while values close to zero denote smooth surfaces. Therefore, the saliency formula s assigns a score to each arc a, thus to discriminate among edges, ramps and smooth surfaces. In order to classify the arcs into these three categories, we have implemented a fuzzy set approach. The fuzzy sets $(S_{\mathrm{E}}, m_{\mathrm{E}})$, $(S_{\mathrm{R}}, m_{\mathrm{R}})$ and $(S_{\mathrm{S}}, m_{\mathrm{S}})$ of figure 4

identify edges, ramps and smooth surfaces, respectively. The extrema of S_{E} and S_{R} are bounded by the minimum value $s_{\min}$ and the maximum value $s_{\max}$ of s, normalized in the interval $[-1, 1]$, so that $S_{\mathrm{E}} \triangleq [s_{\min}, 0] \equiv [-1, 0]$ and $S_{\mathrm{R}} \triangleq [0, s_{\max}] \equiv [0, 1]$. Though the user has the opportunity to choose the bounds of S_{S}, thus to select structures that belong to both edges and ramps, we set $S_{\mathrm{S}} \triangleq [s_{\min}, s_{\max}] \equiv [-1, 1]$.

The $[0, 1]$-valued membership $m(a)$ of a given arc a was defined as the maximum value of its memberships $m_{\mathrm{E}}(a)$, $m_{\mathrm{R}}(a)$ and $m_{\mathrm{S}}(a)$. For the sake of clarity, we indicate in the following figures (the online version of this paper contains coloured images) edges in blue, ramps in red and smooth zones in green. Hue transitions represent the membership degree to a particular set, that is, the membership value is mapped into the RGB color space (coded by m_{R}, m_{S} and m_{E}, in this order).

3 Experimental Results and Conclusions

We presented a methodology for adaptive mesh surfaces segmentation based on mesh elements weighting. The analysis is performed at local level and the segmentation simplifies the object representation for inspection, recognition, matching, etc. We tested the soundness of our algorithm on freely available objects [10] that have different shapes, resolutions and surface complexities, without any data pre-filtering for noise reduction.

The normalization coefficient η ensures a coherent weighting when the resolution of the input object changes according to some decimation algorithm. Otherwise, a set of scanning windows with a variety of sizes would be required to correctly segment not uniformly sampled surfaces. Vice versa, our approach lets obtain a robust segmentation

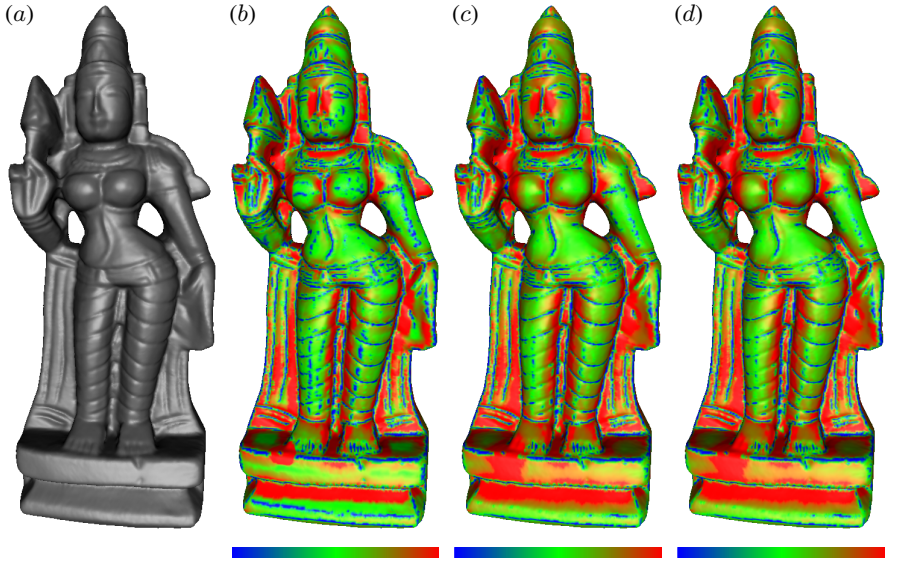


Fig. 5. A three-dimensional object (a) and the effect of different scanning radii (b,c,d:$r = 10$, $r = 20$, d:$r = 30$). Of course, some small detail may vary, but the overall result is quite stable. Colors refer to the interpretation given in figure 4. Author's note: coloured original image.

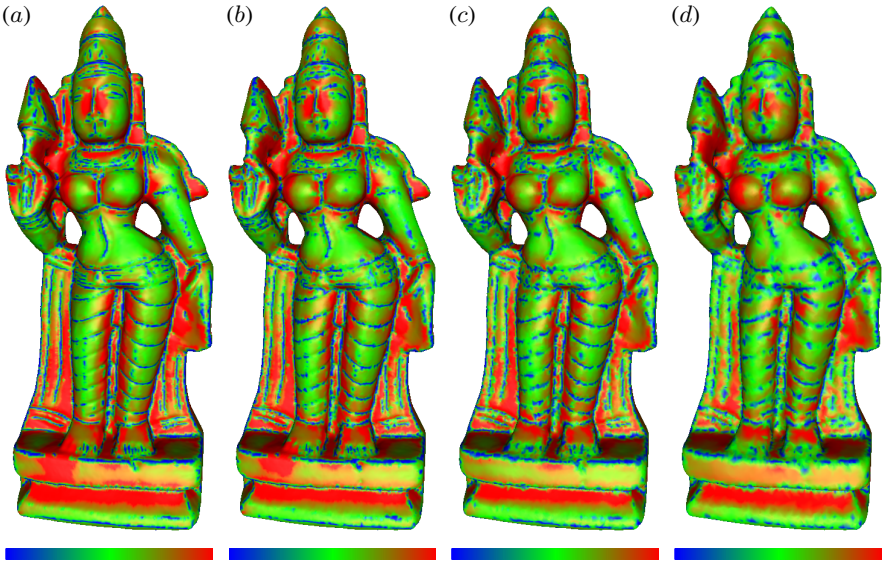


Fig. 6. The proposed algorithm is almost invariant with respect to scale. This example shows the segmentation computed on the full resolution mesh (a) and its reduced versions (b:75% of the original vertices, c:50% and d:25%). Author's note: coloured original image.

and avoids the definition of a maximum scanning radius (see figure 5). The example of figure 6 shows the same object at different resolutions, obtained by reducing the original number of vertices down to 25%. Obviously, many details are missing on this rough surface, but nonetheless our algorithm is almost able to detect the same most evident features, as for any other resolution.

In this paper, we compared the second order difference operator and its extended version, because they do not need any parameter. Figures 7–8 reports a qualitative comparison between objects segmented through these operators. The use of the former has to be preferred because it detects small, but usually important details. Indeed, this means also that this function is more sensible to noise, while the latter tends to erase small components because it performs an averaging on the normal vectors.

The algorithm is resolution independent and automatic, since it accepts just the mesh to analyze and returns a reasonable segmentation without predefined or user-tuned parameters. Ad hoc aggregation procedures group together the output into distinctive features. In our case, to isolate the three-dimensional edges, we introduced a fuzzy measure, based on the saliency degree. This approach takes advantage of the ability to locate in an intuitive fashion important zones on the surface of the objects (e.g. its edges and ramps).

Future work will focus on the definition of geometric descriptors of the features, for recognition and matching tasks. Lastly, we will test also the goodness of set S_{E}, obtained by our method, as a sub-sample of the object's surface to solve problems as object alignment and symmetry detection.

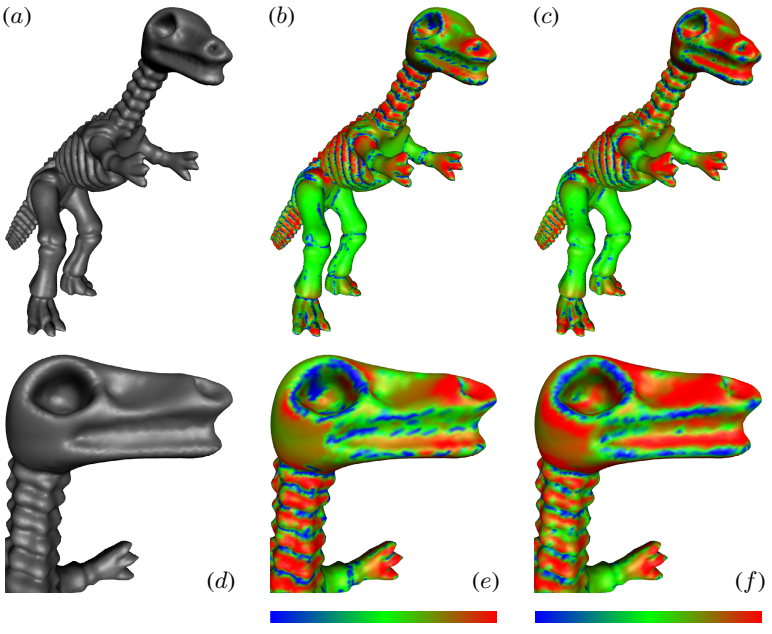


Fig. 7. Dinosaur (a) with SOD (b) and ESOD (c) results with respective zooms (d,e,f). Author's note: coloured original image.

The computational time of the method is strictly related to the *breadth first search* algorithm. For an efficient elaboration of each neighborhood $N_r(a)$, memory accesses must be optimized in order to reduce cache misses. This issue has been handled by scanning the mesh according to a *locality principle*, that is, arcs near to each other are processed *sequentially* for trying to access previously created windows, already available in the cache. This locality principle was implemented by sorting the elements of the mesh (i.e. its sets V, E and F) through an *octree* subdivision and by processing the arcs included in each single cube. For a computed window $N_r(a)$, the time complexity to perform the variance analysis is linear with respect to $|N_r(a)|$. Table 1 illustrates an average execution time, obtained with our C language implementation on a 2.8GHz dual core CPU with 2GB RAM. The number of facets and vertices is also reported.

For the sake of completeness, we provide a graphical user interface, downloadable from http://www.math.unipa.it/~mcipolla/edges.html, to view the results presented here.

Table 1. Time analysis versus size of the mesh

Object	Facets	Vertices	Time
Indian Goddess	274822	137406	40′02″
Dinosaur	80354	112384	17′37″
Venus	268686	134345	36′27″

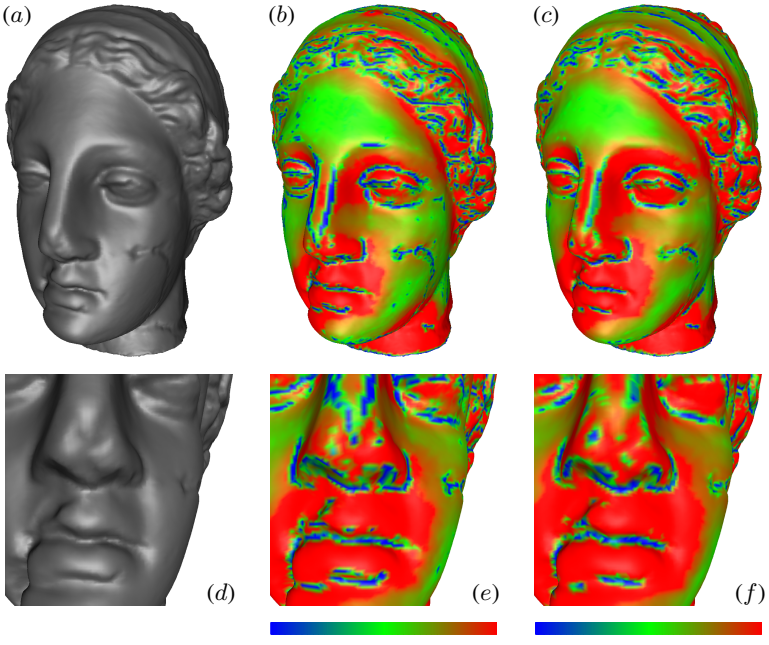


Fig. 8. Venus (a) with SOD (b) and ESOD (c) results with respective zooms (d,e,f). Author's note: coloured original image.

References

1. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004)
2. Shamir, A.: A Formulation of Boundary Mesh Segmentation. In: 2nd International Symposium on 3D Data Processing, Visualization and Transmission, pp. 82–89 (2004)
3. Papaioannou, G., Karabassi, E.-A., Theoharis, T.: Segmentation and Surface Characterization of Arbitrary 3D Meshes for Object Reconstruction and Recognition. In: 15th International Conference on Pattern Recognition, vol. 1, pp. 734–737 (2000)
4. Chen, L., Georganas, N.D.: An efficient and robust algorithm for 3d mesh segmentation. Multimedia Tools and Applications 29(2), 109–125 (2006)
5. Sun, Y., Page, D.L., Paik, J.K., Koschan, A., Abidi, M.A.: Triangle Mesh-Based Edge Detection and its Application to Surface Segmentation and Adaptive Surface Smoothing. In: Proceedings of the International Conference on Image Processing, vol. III, pp. 825–828 (2002)
6. Hubeli, A., Meyer, K., Gross, M.: Multiresolution Feature Extraction for Unstructured Meshes. In: Conference on Visualization, pp. 287–294 (2001)
7. Clarenz, U., Dziuk, G., Rumpf, M.: On generalized mean curvature flow in surface processing. In: Geometric Analysis and Nonlinear Partial Differential Equations, pp. 217–248 (2003)
8. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: International Conference on Computer Graphics and Interactive Techniques, pp. 209–216 (1997)
9. Johnson, A.E., Hebert, M.: Control of polygonal mesh resolution for 3-d computer vision. Graphical Models and Image Processing, 261–285 (1998)
10. The Cyberware Sample Gallery, http://www.cyberware.com

Fuzzy Sets for Image Texture Modelling Based on Human Distinguishability of Coarseness*

Jesús Chamorro-Martínez and Pedro Martínez-Jiménez

Department of Computer Science and Artificial Intelligence, University of Granada,
C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
{jesus,pedromartinez}@decsai.ugr.es

Abstract. In this paper, the "coarseness" texture property is modelled by means of fuzzy sets, relating representative coarseness measures (our reference set) with the human perception of this type of feature. In our study, a wide variety of measures are analyzed, and the coarseness human perception are collected from polls filled by subjects. The capability of each measure to discriminate different coarseness degrees is analyzed, taking into account this capability for defining the membership function.

Keywords: Image features, texture features, fuzzy texture, visual coarseness.

1 Introduction

Texture is, together with color and shape, one of the most used feature for image analysis. It is usual for humans to describe visual textures according to some vague "textural concepts" like *coarseness/fineness*, *orientation* or *regularity* [1,2]. From all of them, the *coarseness/fineness* is the most popular one, being common to associate the presence of fineness with the presence of texture. In this sense, a *fine* texture corresponds to small texture primitives (e.g. the image in figure 1(A)), whereas a *coarse* texture corresponds to bigger primitives (e.g. the image in figure 1(I)).

There are many measures in 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 [3]. However, given a certain measure value, 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). In this sense, the majority of the approaches in the literature are crisp proposals where uncertainty is not properly taken into account.

To face this problem, fuzzy logic has been recently employed for representing the imprecision related to texture. However, in many of these approaches, fuzzy logic is usually applied just during the process, but the output do not habitually model the imprecision (being often a crisp one) [4]. Other interesting approaches

* This work has been partially supported by the Andalusian Government under the TIC1570 and TIC249.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 229–236, 2009.

Fig. 1. Some examples of images with different degrees of fineness

emerge from the content-based image retrieval scope, where semantic data is managed by means of fuzzy sets [5]. However, these fuzzy sets are not obtained by considering the relationship between the feature and the human perception of texture.

In our approach, we propose to model fineness by means of fuzzy sets relating representative coarseness measures (our reference set) with the human perception of this type of feature. For this purpose, pools are used for collecting data about the human perception of fineness. This data is used to analyze the capability of each measure to discriminate different coarseness degrees, and to define the membership function on the basis of this capability.

The rest of the paper is organized as follows. In section 2 we introduce our methodology to obtain the fuzzy sets. Results are shown in section 3, and the main conclusions and future work are sumarized in section 4.

2 Fineness Modelling

As it was pointed, there is not a clear perceptual interpretation of the value given by a fineness measure. To face this problem, we propose to model the fineness perception as a fuzzy set defined on the domain of a given measure[1]. Let $\mathcal{P} = \{P_1, \ldots, P_K\}$ be a set of measures of fineness 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[2]

$$\mathcal{T}_k : \mathbb{R} \rightarrow [0, 1] \tag{1}$$

where a value of 1 will mean fineness presence while a value of 0 will mean no fineness presence (i.e. coarseness presence).

[1] Let us remark that "coarseness" and "fineness" are opposite but related properties.
[2] 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.

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 $\Psi_k = \{(m_k^1, v^1), \ldots, (m_k^N, v^N)\}$, the 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 from the image set $\mathcal{I} = \{I_1, \ldots, I_N\}$ will be described.

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.

The Poll. In order to obtain assessments about the perception of fineness, L subjects will be asked to assign images from $\mathcal{I}$ to classes, so that each class has associated a perception degree of fineness. In particular, $L = 20$ 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). 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. For each image in $\mathcal{I}$, one assessment v^i that summarizes the Θ^i values is needed. To aggregate opinions, an OWA operator guided by a quantifier have been used. Concretely, the quantifier "the most" has been used, which allows to represent the opinion of the majority of the subjects [6].

2.2 Fineness Measures

In this paper, we propose to use the 17 measures shown in table 1 (that includes classical statistical measures well known in the literature, measures in the frequency domain, etc.). As it was expected, some of them have better ability to represent fineness than the others. Thus, to study the ability of each measure to discriminate different degrees of fineness (i.e. how many classes can P_k actually discriminate), 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

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: $\widetilde{Part}_k = C_1, C_2, \ldots, C_{n-k}$

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 $\Upsilon_k = C_1^k, C_2^k, \ldots, C_{NC_k}^k$ 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 representative 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.

Table 1 shows the parameters obtained by applying the proposed algorithm 1 with the different measures considered in this paper. The second column of this table shows how the initial classes have been grouped. The columns from third to eighth show the representative values $\bar{c}_r^k$ and $\bar{v}_r^k$ associated to each cluster.

2.3 Obtaining the Membership Function

In this section we will deal with the problem of obtaining the membership function for the fuzzy set $\mathcal{T}_k$. In our proposal, the following properties will be considered in order to define this function:

Table 1. Clusters, representatives of each class and RMSE found by applying the estimated membership function

Measures	Classes	$(\bar{c}_5, \bar{v}_5)$	$(\bar{c}_4, \bar{v}_4)$	$(\bar{c}_3, \bar{v}_3)$	$(\bar{c}_2, \bar{v}_2)$	$(\bar{c}_1, \bar{v}_1)$	RMSE
Correlation [3]	{1,2-4,5-6,7-8,9}	(0.122,1)	(0.403,0.812)	(0.495,0.562)	(0.607,0.219)	(0.769,0)	0.278
Abbadeni [8]	{1,2-8,9}	-	-	(0.089,1)	(0.166,0.566)	(0.404,0)	0.287
Amadasun [1]	{1,2-6,7-8,9}	-	(5.621,1)	(8.945,0.812)	(11.63,0.391)	(26.94,0)	0.293
ED [9]	{1,2,3-5,6-8,9}	(0.348,1)	(0.282,0.719)	(0.261,0.344)	(0.238,0.125)	(0.165,0)	0.332
Tamura [2]	{1,2-6,7-8,9}	-	(1.540,1)	(1.864,0.812)	(2.125,0.242)	(3.045,0)	0.366
SRE [10]	{1,2-8,9}	-	-	(0.995,1)	(0.987,0.477)	(0.966,0)	0.370
LH [3]	{1,2-8,9}	-	-	(0.023,1)	(0.052,0.621)	(0.127,0)	0.390
FD [11]	{1,2,3-8,9}	-	(3.383,1)	(3.174,0.539)	(2.991,0.125)	(2.559,0)	0.393
DGD [12]	{1,2-8,9}	-	-	(0.020,1)	(0.038,0.621)	(0.091,0)	0.396
Weszka [13]	{1,2-6,7-8,9}	-	(0.153,1)	(0.113,0.812)	(0.099,0.258)	(0.051,0)	0.398
SNE [14]	{1,2-8,9}	-	-	(0.879,1)	(0.775,0.477)	(0.570,0)	0.401
Contrast [3]	{1,2-5,6-8,9}	-	(3312,1)	(2529,0.781)	(1863,0.234)	(790.8,0)	0.420
Newsam [15], Entropy [3], Uniformity [3], FMPS [16], Variance [3]: {1-9}							

- $\mathcal{T}_k$ should be a monotonic function
- $\mathcal{T}_k(\bar{c}_r^k) = \bar{v}_r^k \ \forall r = 1, \ldots, NC_k$, i.e., for the representative values $\bar{c}_r^k$ of each class $C_r^k \in \Upsilon_k$, the membership function $\mathcal{T}_k$ should return the mean assessment given by the subjects to that class.
- The values $\mathcal{T}_k(x) = 0$ and $\mathcal{T}_k(x) = 1$ should be achieved from a certain value

To take into account the above considerations, in this paper we propose to define $\mathcal{T}_k$ as a function that associates the values given by a certain measure with the assessments given by the human subjects[3], i.e.:

$$\mathcal{T}_k(x) = \begin{cases} 0 & x \leq x_1 \\ f^1(x) & x \in (x_1, x_2] \\ f^2(x) & x \in (x_2, x_3] \\ \vdots & \vdots \\ 1 & x > x_{NC_k} \end{cases} \tag{2}$$

with $f^r(x)$ being a line defined as $f^r(x) = a_1^r x + a_0^r$.

To define the knots x_r of equation 2, the representatives of the classes C_r^k, with $r = 1, \ldots, NC_k$, will be used (recall that these classes were obtained considering the ability of the measure P_k to discriminate the perception of fineness following the algorithm 1). Thus, we propose to define $x_r = \bar{c}_r^k$, with $r = 1, \ldots, NC_k$, and with $\bar{c}_r^k$ being the centroid of C_r^k. Note that the way the function is defined allows to ensure that, for the representative values $\bar{c}_r^k$ of each class, the membership function returns the mean assessment given by the subjects to that class (i.e., $\mathcal{T}_k(\bar{c}_r^k) = \bar{v}_r^k$ according with the second constraint). From this point of view,

[3] Note that this function is defined for measures that increases according to the perception of fineness. For those that decreases, the function needs to be changed appropriately.

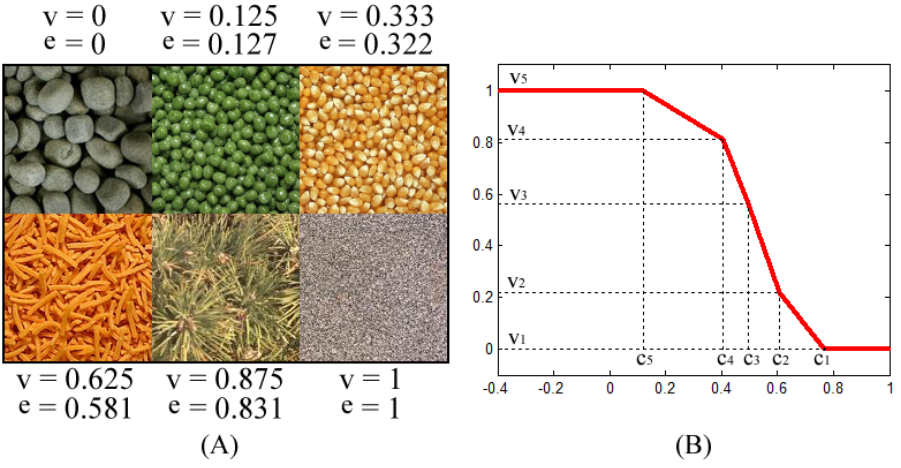


Fig. 2. Results for a mosaic image. (A) Original mosaic image with the assessment values given by users -v- and the value estimated with our model -e- (B)Membership function associated to the proposed fuzzy set.

$f^r(x)$ may be considered as a function that represents the transition between the classes C_r^k and C_{r+1}^k.

The function $f^r(x)$ will be obtained as the 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 above approach has been used to define the membership functions $\mathcal{T}_k$ associated to the 17 measures studied in this paper. 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 (it has been sorted according to the RMSE value).

3 Results

In this section, the membership function $\mathcal{T}_k$ with least fitting error (obtained for the measures Correlation and defined by the parameter values shown in Table 1) will be applied in order to analyze the performance of the proposed model. Let's consider Figure 2(A) corresponding to a mosaic made by several images, each one with a different increasing perception degree of fineness. The perception degree of fineness for each subimage has been calculated using the proposed model and the numerical results are shown in Figure 2(A): v is the value of the human assessment and e is the computed degree. It can be noticed that our model

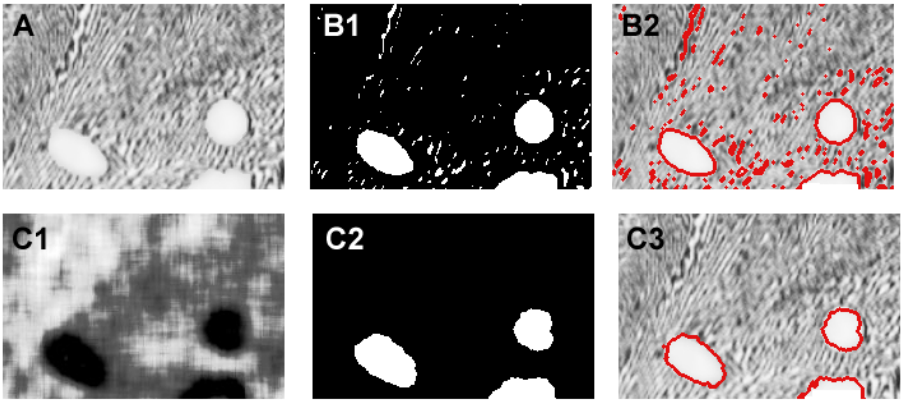


Fig. 3. Pattern recognition (A) Original image (B1) Binary image obtained by thresholding the original one (B2) Region outlines of B1 superimposed on original image (C1) Fineness membership degrees obtained with our model from the original image (C2) Binary image obtained by thresholding C1 (C3) Region outlines of C2 superimposed on original image

captures the evolution of the perception degrees of fineness. Figure 2(B) shows the membership function associated to the fuzzy set used in this experiment.

Figure 3 presents an example where the proposed fuzzy set has been employed for pattern recognition. In this case, the Figure shows a microscopy image (Figure 3(A)) corresponding to the microstructure of a metal sample. The lamellae indicates islands of eutectic, which are to be separated from the uniform light regions. The brightness values in regions of the original image are not distinct, so texture information is needed for extracting the uniform areas. This fact is showed in Figure 3(B1,B2), where a thersholding on the original image is displayed (homogeneous regions cannot be separated from the textured ones as they "share" brightness values). Figure 3(C1) shows a mapping from the original image to its "fineness" membership degree. In this case, we use a window of size 20×20. Thus, Figure 3(C1) represents the degree in which the human perceives the texture and it can be noticed that uniform regions correspond to areas with low degrees of fineness (i.e., high coarseness), so if only the pixels with fineness degree lower than 0.1 are selected (which it is equivalent to a coarseness degree upper than 0.9), the uniform light regions emerge with ease (Figure 3(C2,C3)).

4 Conclusions and Future Works

In this paper, fuzzy sets for representing the fineness/coarseness concept have been proposed. The memberships functions have been defined relating fineness measures with the human perception of this texture property. Pools have been used for collecting data about the human perception of fineness, and the capability of each measure to discriminate different coarseness degrees has been

analyzed. The results given by our approach show a high level of connection with the assessments given by subjects. As future work, the performance of the fineness functions will be analyzed in applications like textural classification or segmentation.

References

1. Amadasun, M., King, R.: Textural features corresponding to textural properties. IEEE Transactions on Systems, Man and Cybernetics 19(5), 1264–1274 (1989)
2. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetics 8, 460–473 (1978)
3. Haralick, R.: Statistical and structural approaches to texture. Proceedings IEEE 67(5), 786–804 (1979)
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, vol. 2, pp. 1173–1176 (2004)
5. Lin, H., Chiu, C., Yang, S.: Finding textures by textual descriptions, visual examples, and relevance feedbacks. Pattern Recognition Letters 24(14), 2255–2267 (2003)
6. Yager, R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man and Cybernetics 18(1), 183–190 (1988)
7. Hochberg, Y., Tamhane, A.: Multiple Comparison Procedures. Wiley, Chichester (1987)
8. 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., vol. 3, pp. 901–904 (2000)
9. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
10. Galloway, M.: Texture analysis using gray level run lengths. Computer Graphics and Image Processing 4, 172–179 (1975)
11. Peleg, S., Naor, J., Hartley, R., Avnir, D.: Multiple resolution texture analysis and classification. IEEE Transactions on Pattern Analysis and Machine Intelligence (4), 518–523 (1984)
12. Kim, S., Choi, K., Lee, D.: Texture classification using run difference matrix. In: Proc. of IEEE 1991 Ultrasonics Symposium, vol. 2, pp. 1097–1100 (1991)
13. Weszka, J., Dyer, C., Rosenfeld, A.: A comparative study of texture measures for terrain classification. IEEE Transactions on Systems, Man and Cybernetics 6, 269–285 (1976)
14. Sun, C., Wee, W.: Neighboring gray level dependence matrix for texture classification. Computer Vision, Graphics and Image Processing 23, 341–352 (1983)
15. 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)
16. 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, 3735–3753 (2003)

Geometry of Spatial Bipolar Fuzzy Sets Based on Bipolar Fuzzy Numbers and Mathematical Morphology

Isabelle Bloch

Télécom ParisTech (ENST) - CNRS UMR 5141 LTCI - Paris, France
Isabelle.Bloch@enst.fr

Abstract. We propose in this paper new tools for dealing with bipolar fuzzy spatial information: particular geometrical objects are defined, as well as measures such as cardinality and perimeter, represented as bipolar fuzzy numbers. A definition of distance from a point to a bipolar fuzzy set is introduced as well. These definitions are based on mathematical morphology operators, recently proposed in the framework of bipolar fuzzy sets.

1 Introduction

Bipolarity has not been much exploited in the spatial domain yet, although it has many features to manage imprecise and incomplete information that could be interesting in this domain. As highlighted e.g. in [1,2], positive information represents what is guaranteed to be possible, for instance because it has already been observed or experienced, while negative information represents what is impossible or forbidden, or surely false. This paper aims at introducing new geometrical tools for dealing with bipolar fuzzy spatial information. After recalling some definitions in Section 2, we propose to extend the notion of bipolar fuzzy number to define particular geometrical objects such as points, disks and rectangles in Section 3. Geometrical measures such as cardinality (Section 4) and perimeter based on gradient (Section 5) are then proposed. They are defined as bipolar fuzzy numbers, in order to reflect the bipolar and fuzzy nature of the objects. Finally, we introduce a definition of distance from a point to a bipolar fuzzy set in Section 6. These definitions are based on mathematical morphology operators, recently proposed in the framework of bipolar fuzzy sets [3,4].

2 Background: Bipolar Fuzzy Numbers and Spatial Bipolar Fuzzy Sets

Interval-valued fuzzy numbers and intervals have been defined in [5,6]. Similarly, we define a bipolar fuzzy number as a pair of fuzzy sets μ and ν such that μ and $1-\nu$ are fuzzy numbers and $\forall \alpha \in \mathbb{R}\,(or\ \mathbb{N}), \mu(\alpha)+\nu(\alpha) \leq 1$. This definition can be relaxed by allowing $1-\nu$ to be a fuzzy interval (i.e. its core is an interval).

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 237–245, 2009.

If both μ and $1-\nu$ are fuzzy intervals, then (μ, ν) will be called bipolar fuzzy interval.

Let us now move to spatial information and let $\mathcal{S}$ denote the spatial domain. It could typically be $\mathbb{R}^n$ or $\mathbb{Z}^n$. Here we consider a finite bounded domain $\mathcal{S}$. A bipolar fuzzy set on $\mathcal{S}$ is defined by a pair of functions (μ, ν) such that $\forall x \in \mathcal{S}, \mu(x) + \nu(x) \leq 1$. Note that a bipolar fuzzy set is formally equivalent to an intuitionistic fuzzy set [7] or an interval-valued fuzzy set [8], where the interval at each point x is $[\mu(x), 1-\nu(x)]$. For each point x, $\mu(x)$ defines the degree to which x belongs to the bipolar fuzzy set (positive information) and $\nu(x)$ the non-membership degree (negative information). This formalism allows representing both bipolarity and fuzziness. Semantically, a bipolar fuzzy set is not one physical object, but may represent information coming from different sources: for instance the positive part may represent observed or preferred positions for a spatial object, while the negative part may represent constraints on this position, and $1-\mu-\nu$ indifferent or neutral positions.

Let $\mathcal{B}$ denote the set of bipolar fuzzy sets on $\mathcal{S}$, $(\mathcal{B}, \preceq)$ is a complete lattice for the partial order defined as: $(\mu_1, \nu_1) \preceq (\mu_2, \nu_2)$ iff $\forall x \in \mathcal{S}, \mu_1(x) \leq \mu_2(x)$ and $\nu_1(x) \geq \nu_2(x)$. The supremum and the infimum are denoted by $\wedge$ and $\vee$, respectively.

An equivalent of the extension principle writes [5,6] $((\mu_1, \nu_1) \circledast (\mu_2, \nu_2))(\gamma) = \vee_{\gamma=\alpha\circledast\beta}(\mu_1, \nu_1)(\alpha) \wedge (\mu_2, \nu_2)(\beta)$, where $\circledast$ denotes any operation. This principle can in particular be applied to define operations on fuzzy numbers or intervals.

On the lattice $(\mathcal{B}, \preceq)$, a dilation is defined as an operation that commutes with the supremum and an erosion as an operation that commutes with the infimum, as shown in our previous work [3,4]. Particular forms, invariant under translation and involving a bipolar fuzzy structuring element, have also been detailed in this work, along with their properties.

3 Bipolar Fuzzy Points, Spheres and Parallelepipeds

We propose to use the notion of bipolar fuzzy number to derive particular geometrical bipolar fuzzy sets on $\mathcal{S}$.

Bipolar fuzzy points, disks (in 2D) or spheres (in 3D) can be defined by applying a rotation invariance principle, while rectangles or parallelepipeds can be defined based on a Cartesian product.

Definition 1. *Let (μ_1, ν_1) be a bipolar fuzzy number, and let $d(x_0, x)$ denote the distance between two points x_0 and x in $\mathcal{S}$ (Euclidean distance, or a discrete version of it for instance). A bipolar fuzzy point is defined as: $\forall x \in \mathcal{S}, (\mu_{x_0}, \nu_{x_0})(x) = (\mu_1(d(x_0, x)), \nu_1(d(x_0, x)))$.*

Definition 2. *Let (μ_1, ν_1) be a bipolar fuzzy interval, and let $d(x_0, x)$ denote the distance between two points x_0 and x in $\mathcal{S}$. A bipolar fuzzy disk (in 2D) or sphere (in 3D) is defined as: $\forall x \in \mathcal{S}, (\mu_D, \nu_D)(x) = (\mu_1(d(x_0, x)), \nu_1(d(x_0, x)))$.*

Definition 3. *Let (μ_i, ν_i) be bipolar fuzzy intervals defined on each axis of the coordinate frame. A bipolar fuzzy rectangle (in 2D) or parallelepiped*

(in 3D) is defined as the Cartesian product of these bipolar fuzzy intervals: $\forall x \in \mathcal{S}, (\mu_R, \nu_R)(x) = \wedge_i((\mu_i, \nu_i)(x_i)) = (\min_i(\mu_i(x_i)), \max_i(\nu_i(x_i)))$ *where* $x_1, x_2(, x_3)$ *denote the coordinates of* x $(x = (x_1, x_2(, x_3))^t)$ *and* $\wedge$ *denotes the conjunction of bipolar fuzzy sets.*

These definitions extend the notion of fuzzy point [9], fuzzy disk [10] and fuzzy rectangle [10] to the bipolar case. Note that μ_D is a convex fuzzy disk, and ν_D the complement of a convex fuzzy disk. Relaxing the convexity assumption would lead to more general bipolar fuzzy disks and spheres. As for rectangles, the conjunction is expressed as a minimum of the positive parts and the maximum of the negative parts. The positive part is exactly a fuzzy rectangle.

Proposition 1. *Definitions 1, 2 and 3 actually provide bipolar fuzzy sets in* $\mathcal{S}$.

Proposition 2. *If the bipolar fuzzy numbers or intervals involved in Definitions 1-3 are not bipolar (i.e.* $\nu_i = 1 - \mu_i$*), then these definitions provide non-bipolar fuzzy sets and are consistent with the existing definitions in the fuzzy case. If moreover* μ_i *is crisp, then the defined sets are crisp and are points, disks and rectangles in the classical sense.*

An example of bipolar fuzzy disk is shown in Figure 1.

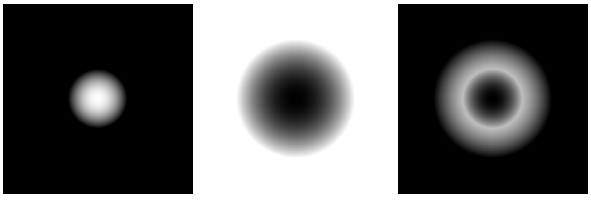


Fig. 1. A bipolar fuzzy disk with its positive part μ (left) and its negative part ν (middle). The indetermination (or neutral area) $\pi = 1 - \mu - \nu$ is shown on the right.

4 Cardinality, Surface and Volume

Let $(\mu, \nu) \in \mathcal{B}$ be a bipolar fuzzy set defined in the spatial domain $\mathcal{S}$. The cardinality of intuitionistic or interval-valued fuzzy sets has been introduced e.g. in [11] as an interval: $[\sum_{x \in \mathcal{S}} \mu(x), \sum_{x \in \mathcal{S}} (1 - \nu(x))]$, with the lower bound representing the classical cardinality of the fuzzy set defining the positive part (the least certain cardinality), and the upper bound the cardinality of the complement of the negative part (i.e. the whole not impossible region is considered, leading to the largest possible cardinality). The length of the interval reflects the indetermination encoded by the bipolar representation. Several authors have used a similar approach, based on interval representations of the cardinality.

When dealing with fuzzy sets, it may be more interesting to consider the cardinality as a fuzzy number, instead as a crisp number, for instance using the extension principle: $|\mu|(n) = \sup\{\alpha \in [0, 1] : |\mu_\alpha| = n\}$, where μ_α denotes α-cuts, defining the degree to which the cardinality of μ is equal to n.

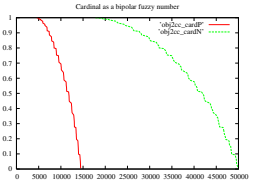


Fig. 2. Cardinality of the bipolar fuzzy set of Figure 3 represented as a bipolar fuzzy number (the negative part, represented by the upper curve is inverted)

Here we propose a similar approach for defining the cardinality of a bipolar fuzzy set as a bipolar fuzzy number, which contrasts with the previously interval-based approaches [4].

Definition 4. *Let* $(\mu, \nu) \in \mathcal{B}$*. Its cardinality is defined as:* $\forall n, |(\mu, \nu)|(n) = (|\mu|(n), 1 - |1 - \nu|(n))$.

Proposition 3. *The cardinality introduced in Definition 4 is a bipolar fuzzy number on* $\mathbb{N}$ *(with* $\forall n, |\mu|(n) + (1 - |1 - \nu|(n)) \leq 1$*).*

In the spatial domain, the cardinality can be interpreted as the surface (in 2D) or the volume (in 3D) of the considered bipolar fuzzy set.

An example is shown in Figure 2, for the bipolar fuzzy object displayed in Figure 3. For this example, the cardinality computed as an interval would provide $[11000, 40000]$, which approximately corresponds to the 0.5-level of the bipolar fuzzy number.

5 Gradient and Perimeter

A direct application of erosion and dilation is the morphological gradient, which extracts boundaries of objects by computing the difference between dilation and erosion, as introduced in [4] for bipolar fuzzy sets.

Definition 5. *Let* (μ, ν) *be a bipolar fuzzy set. We denote its dilation by a bipolar fuzzy structuring element by* (δ^+, δ^-) *and its erosion by* $(\varepsilon^+, \varepsilon^-)$*. We define the bipolar fuzzy gradient as:* $\nabla(\mu, \nu) = (\min(\delta^+, \varepsilon^-), \max(\delta^-, \varepsilon^+))$ *which is the set difference, expressed as the conjunction between* (δ^+, δ^-) *and the negation* $(\varepsilon^-, \varepsilon^+)$ *of* $(\varepsilon^+, \varepsilon^-)$*.*

Proposition 4. *The bipolar fuzzy gradient has the following properties: (i) Definition 5 defines a bipolar fuzzy set; (ii) If the dilation and erosion are defined using t-representable bipolar t-norms and t-conorms (see [3,4] for details), we have:* $\nabla(\mu, \nu) = (\min(\delta_{\mu_B}(\mu), \delta_{\mu_B}(\nu)), \max(\varepsilon_{1-\nu_B}(\nu), \varepsilon_{1-\nu_B}(\mu)))$*. Moreover, if* (μ, ν) *is not bipolar (i.e.* $\nu = 1 - \mu$*), then the positive part of the gradient is equal to* $\min(\delta_{\mu_B}(\mu), 1 - \varepsilon_{\mu_B}(\mu))$*, which is exactly the morphological gradient in the fuzzy case.*

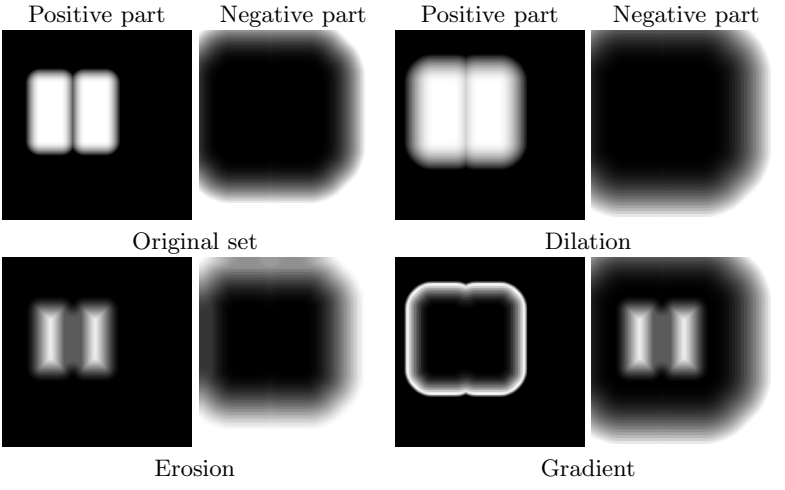


Fig. 3. Gradient of a bipolar fuzzy object

An illustration is displayed in Figure 3. It illustrates both the imprecision (through the fuzziness of the gradient) and the indetermination (through the indetermination between the positive and the negative parts). The object is here somewhat complex, and exhibits two different parts, that can be considered as two connected components to some degree. The positive part of the gradient provides a good account of the boundaries of the union of the two components, which amounts to consider that the region between the two components, which has lower membership degrees, actually belongs to the object. The positive part has the expected interpretation as a guaranteed position and extension of the contours. The negative part shows the level of indetermination in the gradient: the gradient could be larger as well, and it could also include the region between the two components.

Now the perimeter (in 2D) or surface (in 3D) of a bipolar fuzzy set can be derived from the notions of cardinality and gradient. It is then a bipolar fuzzy number. As for the cardinality, this representation is suitable to account for both fuzziness and indetermination. This is a richer representation than a

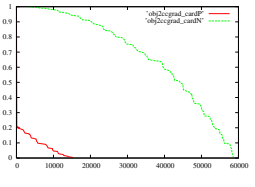


Fig. 4. Perimeter of the bipolar fuzzy set shown in Figure 3 represented as a bipolar fuzzy number (the negative part is inverted), and computed as the cardinality of the gradient

simple number, that could be obtained for instance as a weighted sum of the cardinalities of the cuts, as done in the fuzzy case in [10].

Definition 6. *Let (μ, ν) be a bipolar fuzzy set. Its perimeter (or surface) is defined as the bipolar fuzzy number $|\nabla(\mu, \nu)|$, where the gradient $\nabla(\mu, \nu)$ is given in Definition 5 and the cardinality $|.|$ in Definition 4.*

An example is shown in Figure 4.

6 Distance from a Point to a Bipolar Fuzzy Set

While there is a lot of work on distances and similarity between interval-valued fuzzy sets or between intuitionistic fuzzy sets (see e.g. [12,13]), none of the existing definitions addresses the question of the distance from a point to a bipolar fuzzy set, nor includes the spatial distance in the proposed definitions. As in the fuzzy case [14], we propose to define the distance from a point to a bipolar fuzzy set using a morphological approach. In the crisp case, the distance from a point x to a set X is equal to n iff x belongs to the dilation of size n of X (the dilation of size 0 being the identity), but not to dilations of smaller size (it is sufficient to test this condition for $n-1$ in the discrete case). The transposition of this property to the bipolar fuzzy case leads to the following novel definition, using bipolar fuzzy dilations introduced in [3].

Definition 7. *The distance from a point x of $\mathcal{S}$ to a bipolar fuzzy set (μ, ν) $(\in \mathcal{B})$ is defined as: $d(x, (\mu, \nu))(0) = (\mu(x), \nu(x))$ and $\forall n \in \mathbb{N}^*, d(x, (\mu, \nu))(n) = \delta^n_{(\mu_B, \nu_B)}(x) \wedge c(\delta^{n-1}_{(\mu_B, \nu_B)}(x))$, where c is a complementation (typically the standard negation $c(a, b) = (b, a)$ is used) and $\delta^n_{(\mu_B, \nu_B)}$ denotes n iterations of the dilation, using the bipolar fuzzy set (μ_B, ν_B) as structuring element.*

In order to clarify the meaning of this definition, let us consider the case where the structuring element is not bipolar, i.e. $\nu_B = 1 - \mu_B$. Then the dilation writes (see [3,4] for details): $\delta_{(\mu_B, 1-\mu_B)}(\mu, \nu) = (\delta_{\mu_B}(\mu), \varepsilon_{\mu_B}(\nu))$, where $\delta_{\mu_B}(\mu)$ is the fuzzy dilation of μ by μ_B and $\varepsilon_{\mu_B}(\nu)$ is the fuzzy erosion of ν by μ_B (see [15] for fuzzy mathematical morphology). The bipolar degree to which the distance from x to (μ, ν) is equal to n then writes: $d(x, (\mu, \nu))(n) = (\delta^n_{\mu_B}(\mu) \wedge \varepsilon^{n-1}_{\mu_B}(\nu), \varepsilon^n_{\mu_B}(\nu) \vee \delta^{n-1}_{\mu_B}(\mu))$, i.e. the positive part is the conjunction of the positive part of the dilation of size n (i.e. a dilation of the positive part of the bipolar fuzzy object) and the negative part of the dilation of size $n-1$ (i.e. an erosion of the negative part of the bipolar fuzzy object), and the negative part is the disjunction of the negative part of the dilation of size n (erosion of ν) and the positive part of the dilation of size $n-1$ (dilation of μ).

Proposition 5. *The distance introduced in Definition 7 has the following properties: (i) it is a bipolar fuzzy set on $\mathbb{N}$; (ii) it reduces to the distance from a point to a fuzzy set, as defined in [14], if (μ, ν) and (μ_B, ν_B) are not bipolar (hence the consistency with the classical definition of the distance from a*

point to a set is achieved as well); (iii) the distance is strictly equal to 0 (i.e. $d(x,(\mu,\nu))(0) = (1,0)$ *and* $\forall n \neq 0, d(x,(\mu,\nu))(n) = (0,1)$*) iff* $\mu(x) = 1$ *and* $\nu(x) = 0$*, i.e. x completely belongs to the bipolar fuzzy set.*

An example is shown in Figure 5. The results are in agreement with what would be intuitively expected. The positive part of the bipolar fuzzy number is put towards higher values of distances when the point is moved to the right of the object. After a number n of dilations, the point completely belongs to the dilated object, and the value to which the distance is equal to n', with $n' > n$, becomes $(0,1)$. Note that the indetermination in the membership or non-membership to the object (which is truly bipolar in this example) is also reflected in the distances.

These distances can be easily compared using the extension principle given in Section 2, providing a bipolar degree $d_{\leq}$ to which a distance is less than another one. For the examples in Figure 5, we obtain for instance : $d_{\leq}[d(x_1,(\mu,\nu)) \leq d(x_2,(\mu,\nu))] = [0.69, 0.20]$ where x_i denotes the ith point from left to right in the figure. In this case, since x_1 completely belongs to (μ,ν), the degree

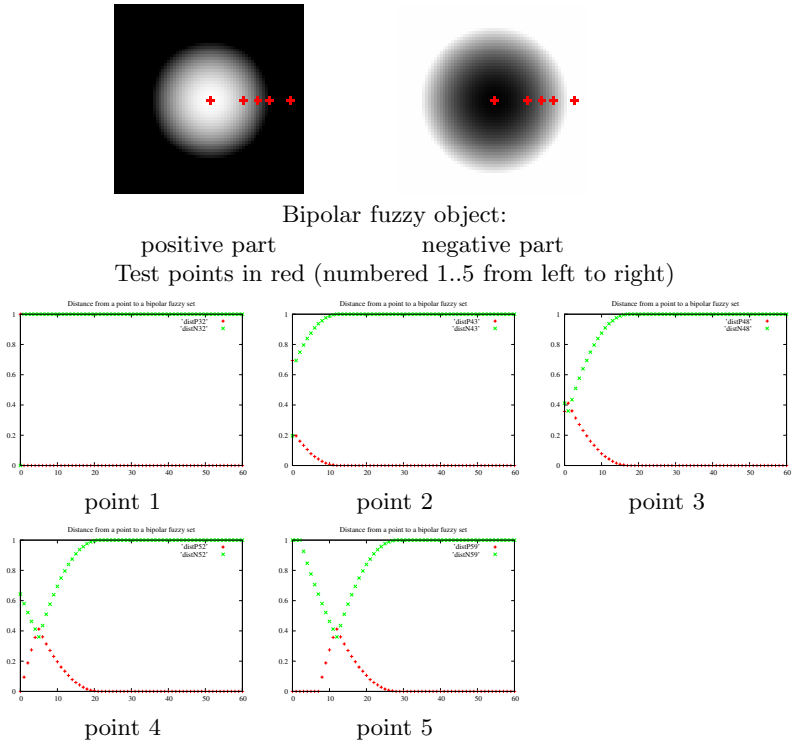


Fig. 5. A bipolar fuzzy set and the distances from 5 different points to it, represented as bipolar fuzzy numbers (positive part: lower curve, negative part: upper curve)

to which its distance is less than the distance from x_2 to (μ, ν) is equal to $[\sup_a d^+(a), \inf_a d^-(a)]$, where d^+ and d^- denote the positive and negative parts of $d(x_2, (\mu, \nu))$. As another example, we have $d_{\leq}[d(x_5, (\mu, \nu)) \leq d(x_2, (\mu, \nu))] = [0.03, 0.85]$, reflecting that x_5 is clearly not closer to the bipolar fuzzy set (μ, ν) than x_2.

7 Conclusion

We have shown in this paper how the set of operations on bipolar fuzzy objects (or equivalently interval-valued or intuitionistic fuzzy sets) can be enhanced with new geometrical features, having nice properties. This enriches existing tools developed for image thresholding or edge detection based on intuitionistic or interval-valued fuzzy sets [16,17,13] or for mathematical morphology [3,4,18]. Extensions to other types of operations will be the aim of future work.

References

1. Dubois, D., Kaci, S., Prade, H.: Bipolarity in Reasoning and Decision, an Introduction. In: International Conference on Information Processing and Management of Uncertainty, IPMU 2004, Perugia, Italy, pp. 959–966 (2004)
2. Benferhat, S., Dubois, D., Kaci, S., Prade, H.: Bipolar Possibility Theory in Preference Modeling: Representation, Fusion and Optimal Solutions. Information Fusion 7, 135–150 (2006)
3. Bloch, I.: Dilation and Erosion of Spatial Bipolar Fuzzy Sets. In: Masulli, F., Mitra, S., Pasi, G. (eds.) WILF 2007. LNCS (LNAI), vol. 4578, pp. 385–393. Springer, Heidelberg (2007)
4. Bloch, I.: A Contribution to the Representation and Manipulation of Fuzzy Bipolar Spatial Information: Geometry and Morphology. In: Workshop on Soft Methods in Statistical and Fuzzy Spatial Information, Toulouse, France, pp. 7–25 (2008)
5. Wang, G., Li, X.: The Applications of Interval-Valued Fuzzy Numbers and Interval-Distribution Numbers. Fuzzy Sets and Systems 98, 331–335 (1998)
6. Hong, D., Lee, S.: Some Algebraic Properties and a Distance Measure for Interval-Valued Fuzzy Numbers. Information Sciences 148, 1–10 (2002)
7. Atanassov, K.T.: Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 20, 87–96 (1986)
8. Zadeh, L.A.: The Concept of a Linguistic Variable and its Application to Approximate Reasoning. Information Sciences 8, 199–249 (1975)
9. Buckley, J.J., Eslami, E.: Fuzzy Plane Geometry I: Points and Lines. Fuzzy Sets and Systems 86, 179–187 (1997)
10. Rosenfeld, A.: The Fuzzy Geometry of Image Subsets. Pattern Recognition Letters 2, 311–317 (1984)
11. Szmidt, E., Kacprzyk, J.: Entropy for Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 118, 467–477 (2001)
12. Szmidt, E., Kacprzyk, J.: Distances between Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 114, 505–518 (2000)
13. Vlachos, I., Sergiadis, G.: Intuitionistic Fuzzy Information – Applications to Pattern Recognition. Pattern Recognition Letters 28, 197–206 (2007)

14. Bloch, I.: On Fuzzy Distances and their Use in Image Processing under Imprecision. Pattern Recognition 32, 1873–1895 (1999)
15. Bloch, I., Maître, H.: Fuzzy Mathematical Morphologies: A Comparative Study. Pattern Recognition 28, 1341–1387 (1995)
16. Chaira, T., Ray, A.: A New Measure using Intuitionistic Fuzzy Set Theory and its Application to Edge Detection. Applied Soft Computing Journal 8, 919–927 (2008)
17. Couto, P., Bustince, H., Melo-Pinto, P., Pagola, M., Barrenechea, E.: Image Segmentation using A-IFSs. In: IPMU 2008, Malaga, Spain, pp. 1620–1627 (2008)
18. Nachtegael, M., Sussner, P., Mélange, T., Kerre, E.: Some Aspects of Interval-Valued and Intuitionistic Fuzzy Mathematical Morphology. In: IPCV 2008 (2008)

Interactive Image Retrieval in a Fuzzy Framework

Malay K. Kundu, Minakshi Banerjee, and Priyank Bagrecha

Machine Intelligence Unit, Center for Soft Computing Research
Indian Statistical Institute
203, B. T. Road, Kolkata 700 108, India
{minakshi_r,malay,b.priyank_t}@isical.ac.in

Abstract. In this paper, an interactive image retrieval scheme using MPEG-7 visual descriptors is proposed. The performance of image retrieval systems is still limited due to semantic gap, which is created from the discrepancies between the computed low-level features (color, texture, shape, etc.) and user's conception of an image. As a result, more interest has been created towards development of efficient learning mechanism other than designing sophisticated low-level feature extraction algorithms. A simple relevance feedback mechanism is proposed, that learns user's interest and updates feature weights based on a fuzzy feature evaluation measure. This has an advantage of handling comparatively small number of samples over those using standard classifiers involving large number of training samples and having more complexity. Extensive experiments have been performed to test to what extent the performance of an image retrieval system can be enhanced further using MPEG-7 standard visual features at minimum cost.

Efficient image retrieval techniques from a large database have become an active field of research with the advent of the World-Wide Web. Content-Based Image Retrieval(CBIR) techniques are becoming more important with this basic requirement [1]. It is aimed at retrieving relevant images from an image database by measuring similarity between the automatically derived low-level features (color, texture, shape, etc.) of the query image and the images stored in the database. Although different image characterization methods [2],[3] have been explored to represent images with basic low-level features but their usefulness is limited by the gap, between low-level features and high-level concepts known as semantic gap. Performance of CBIR is still far from user's expectations owing to semantic gap.

To facilitate effective use of audio, visual(color, texture, shape, etc.) and motion descriptors, ISO/IEC has launched MPEG-7 to address multimedia retrieval. It provides a collection of specific, standard descriptors [4] used as a benchmark for evaluation of new schemes for image retrieval [5]. Among various state-of-the-art techniques in narrowing down the semantic gap, relevance feedback mechanism has been identified as an essential tool to provide significant performance boost in CBIR systems [6], [7], [8], through continuous learning and interaction with

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 246–253, 2009.

end-users. The system provides initial retrieval results through query-by-example, sketch, etc., based on which the user judges the retrieved results as to whether and to what degree, they are relevant (positive examples)/irrelevant (negative examples) to the query. Machine learning algorithm is then applied to learn the user's feedback and improve the results iteratively till user's satisfaction.

Most of the relevance feedback methods employ two approaches [6] namely, query vector moving technique and feature re-weighting technique to improve retrieval results. In the first approach, the query is reformulated by moving the vector towards positive examples and away from the negative examples, assuming that all positive examples will cluster in the feature space. Feature re-weighting method is used to enhance the importance of those components of a feature vector, that help in retrieving relevant images, while reducing the importance of the features that does not help. However in such cases, the selection of positive and negative examples, from a small number of samples having large number of features, still remain as a problem.

Relevance feedback techniques in CBIR, have mostly utilized information of the relevant images but have not made use of the information from irrelevant images. Zin et al., [9] have proposed a feature re-weighting technique by using both the relevant and the irrelevant information, to obtain more effective results.

Recently, relevance feedback has been considered as a learning and classification process, using classifiers like Bayesian classifiers [10], neural network [11], etc. However trained classifiers become less effective when the training samples are insufficient in number. To overcome such problems, active learning methods have been used in [12].

Our contributions in this paper deal with studying the performance of a CBIR system using MPEG-7 visual descriptors. We present a simple relevance feedback method which uses the concept of combining information from both relevant and irrelevant images, from a small number of retrieved images. The motivation is to test how far it can enhance the performance of a CBIR system already using MPEG-7 visual features by using a simple feedback mechanism requiring less computational time and managing small samples opposed to different classifiers [11], [10] used for relevance feedback. We also present a comparison of the system against a CBIR system which uses moments of significant points (corners and around) [13] of a color image involving 6 feature components only.

1 Evaluating Importance of a Feature from Relevance Rating

In conventional CBIR approaches an image I is usually represented by a set of features, $F = \{f_q\}_{q=1}^{N}$, where f_q is the qth feature component in the N dimensional feature space. The commonly used decision function for measuring similarity between the query image I_{qr} and other images I, is represented as,

$$D_{is}(I, I_{qr}) = \sum_{q=1}^{N} w_q ||f_q(I) - f_q(I_{qr})|| \tag{1}$$

where $||f_q(I)-f_q(I_{qr})||$ is the Euclidean distance between the qth component and w_q is the weight assigned to the qth feature component. The weights should be adjusted such that, the features have small variation over the relevant images and large variation over the irrelevant images. Let k similar images I_s={I_1,I_2,...,I_k} where, $I_k \in I_s$, are returned to the user. Let I_r be the set of relevant images and I_{ir} be the set of irrelevant images as marked by the user. $I_r = \{I_j \mid I_j$ relevant, for $I_j \in I_s\}$ and $I_{ir} = \{I_j \mid I_j$ irrelevant, for $I_j \in I_s\}$. The information from I_r and I_{ir} are combined to compute the relative importance of the individual featues, from fuzzy feature evaluation index (FEI) [14] in pattern classification problems.

The (FEI) is defined from interclass and intraclass ambiguities as follows : Let C_1, C_2,.... C_j ... C_m be the m pattern classes in an N dimensional ($f_1, f_2, f_q, ...f_N$) feature space where class C_j contains, n_j number of samples. It can be shown that entropy of a fuzzy set [15] gives a measure of 'intraset ambiguity ' along the qth co-ordinate axis in C_j is computed as,

$$H(A) = (\frac{1}{n_j \ln 2}) \sum_i S_n(\mu_(f_{iqj})); i = 1, 2...n_j \tag{2}$$

where the Shannon's function, $S_n(\mu(f_{iqj}))$=-$\mu(f_{iqj})$ln$\mu(f_{iqj})$-{1-$\mu(f_{iqj})$}ln {1-$\mu(f_{iqj})$} Entropy is dependent on the absolute values of membership (μ) $H_{min} = 0$ for μ=0 or 1, $H_{max} = 1$ for μ=0.5 Entropy (H) of C_j along qth component can be computed using a standard S-type membership function as shown in (3).

$$\begin{aligned} S(x; a, b, c) &= 0 && x \le a \\ &= 2 \times \{\tfrac{(x-a)}{(c-a)}\}^2 && a \le x \le b \\ &= 1 - 2 \times \{\tfrac{(x-c)}{(c-a)}\}^2 && b \le x \le c \\ &= 1 && x \ge c \end{aligned} \tag{3}$$

where,
b = $(f_{qj})av$
c = $b + max\{|(f_{qj})av - (f_{qj})max|, |(f_{qj})av - (f_{qj})min|\}$
a = $2b - c$

$(f_{qj})av$, $(f_{qj})max$, (f_{qj})min denote the mean, maximum and minimum values respectively computed along the qth co-ordinate axis over all the n_j samples in c_j. Since $\mu(b) = \mu(f_{qj})av = 0.5$, the values of H are 1.0 at $b = (f_{qj})av$ and would tend to zero when moved away from b towards either c or a of the S function. Selecting b= $(f_{qj})av$ indicates that, the cross over point is near the query feature component. Higher value of H,indicates more number of samples having $\mu(f)$ equal to 0.5. with a tendency of the samples to cluster around the mean value, resulting in less internal scatter within the class.

After combining the classes C_j and C_k the mean, maximum and minimum values (f_{qkj})av, (f_{qjk})max, (f_{qjk})min respectively of qth dimension over the samples $(n_j + n_k)$ are computed similarly, where n_k are the samples in class C_k. The criteria of a good feature is that, it should be nearly invariant within class,

while emphasizing differences between patterns of different classes [14]. The value of H would therefore decrease, after combining C_j and C_k as the goodness of the qth feature in discriminating pattern classes C_j and C_k increases. The measure denoted as H_{qjk} is called "interset ambiguity" along q^{th} dimension between classes C_j and C_k. Considering the two types of ambiguities, the proposed Feature evaluation index (FEI) for the qth feature is,

$$(FEI_q) = \frac{H_{qjk}}{H_{qj} + H_{qk}} \tag{4}$$

Lower value of FEI_q, indicates better quality of importance of the qth feature in recognizing and discriminating different classes. The precision of retrieval can be improved with these values.

In the proposed algorithm, the number of classes are two. The relevant/ positive images constitute the (intraclass) and the irrelevant/negative images constitute the (interclass) image features. To evaluate the importance of the *qth* feature, the *qth* component of the retrieved images is considered. i.e., $I^{(q)} = \{I_1^{(q)}, I_2^{(q)}, I_3^{(q)},, ...I_k^{(q)}\}$

H_{qj} is computed from $I_r^{(q)} = \{I_{r1}^{(q)}, I_{r2}^{(q)}, I_{r3}^{(q)},, ...I_{rk}^{(q)}\}$. Similarly H_{qk} is computed from the set of images, $I_{ir}^{(q)} = \{I_{ir1}^{(q)}, I_{ir2}^{(q)}, I_{ir3}^{(q)},, ...I_{irk}^{(q)}\}$. H_{qkj} is computed combining both the sets. Images are ranked according to Euclidean distance. The user marks the relevant and irrelevant set from 20 returned images, for automatic evaluation of (FEI). The weight w_q is a function of the evaluated (FEI_q) as shown in eqn. 5

$$w_q = F_q(FEI_q) \tag{5}$$

Now the problem is, what would be weight updation function for the automatically evaluated important features, when all feature elements are merged into a big overall feature vector. Owing to such situations, different feature updation functions like, $[w_q = FEI_q^2, \frac{1}{FEI_q^2}, exp(FEI_q)]$ could be tested. Whichever is the best strategy can be decided by selecting a better performing w_q for majority of the queries in the database. For the query feature vector F, the individual components of relevant images are expected to vary within a smaller range say (ϵ) and may be represented as.

$$I_r = \{I_j \in I_s : \frac{\delta f_q}{|F|} \leq \epsilon\} \tag{6}$$

In the first pass, all features are considered to be equally important. Hence w_1=w_2,...=w_q=1. The feature spaces of the relevant images are therefore altered in a similar fashion after updating the components with w_q. As a result, the ranks of the relevant images are not affected much. For irrelevant images, one feature component may be very close to the query, whereas other feature component may be far away from the query feature. But the magnitude of the similarity vector may be close to the relevant ones. These images may be characterized as,

$$I_{ir} = \{I_j \in I_s : \frac{\delta f_{ql}}{|F|} \gg \epsilon \; and \frac{\delta f_{qm}}{|F|} \ll \epsilon\} \tag{7}$$

where, $l \in 1, 2....N$ and $m{=}l^c$. For example, multiplying by FEi_q^2 will in effect decrease the component feature separation such that, the relevant component i.e., the term $\frac{\delta f_{qm}}{|F|} \ll \epsilon$ of (6) will be more closer to the query than the irrelevant component and due to combined distance within the similarity metric, the relevant images may come up. Multiplying by $\frac{1}{FEI_q^2}$ increases the feature separation between the irrelevant component, such that due to the combined effect the irrelevant image may be pulled down.

2 Experimentation

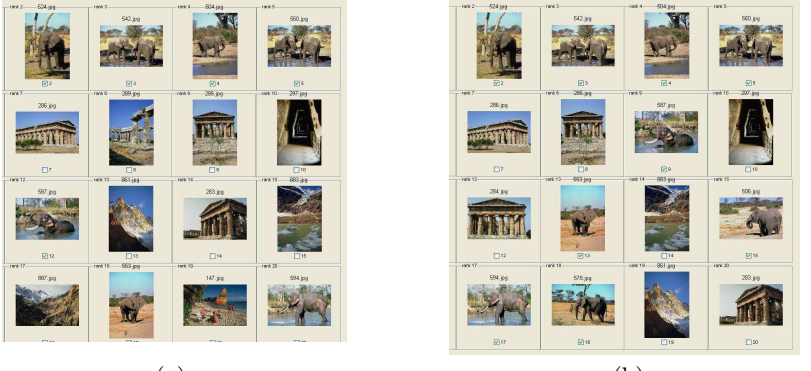
To prove the effectiveness of the proposed relevance feedback mechanism, extensive experiments have been performed on MPEG-7 standard visual features downloaded from http://standards.iso.org/ittf/licence.html, upon two databases namely, (A) SIMPLIcity images consisting of 1000 images from 10 different categories (B) Corel 10000 miscellaneous database which is labeled into 79 semantic categories and downloaded from http://wang.ist.psu.edu/IMAGE. The updation formula $w_q{=}FEI_q^2$ is used in each iteration as it generated better results in majority of the cases.

Among the different representation schemes used as the MPEG-7 core features, CSD(Color Structure Descriptor) and EHD (Edge Histogram Descriptor) [5] are chosen as color and texture features, to evaluate overall similarity between images. Shape could be more important if region based properties were extracted. *The color structure descriptor* histogram aims at identifying localized color distribution using a small sliding structuring window, which is constructed in the hue-min-max-difference (HMMD). CSD is defined by non linear quantization of the color space, and represented by a 184 bin histogram. A 184 bin of CSD can be mapped to lower number of bins (120, 64, or 32) by coarsely quantizing the color space and L2 (Euclidean) distance is used for similarity evaluation. In *Edge Histogram Descriptor* the original image is divided into 16 subimages, each subimage is divided into a fixed number of blocks. Each image block is partitioned into 2x2 block of pixels. The edge detector operators are then applied to 2X2 blocks, treating each block as a pixel and the average intensity as the corresponding block intensity value. A total of 80 bins, 3 bits/bin, are used for representing the edge histogram while L2 distance is used as the metric.

As representation schemes associated with the traditional color histograms, suffer from problems associated with high-dimensional indexing, the performance is also evaluated against a CBIR system, using color moments at significant spatial locations (corners and around) with a feature vector size of six components only [13].

MPEG-7 CSD feature extraction approximately takes 1-2 secs, whereas EHD takes approximately 500ms. Such a difference can be accounted to the use of sliding structuring window depending upon the image size in the case of CSD descriptors. The proposed relevance feedback has been tested independently using the features (CSD, EHD and moment based [13]) and retrieval performance measured in terms of precision as shown in eqn. (8).

As low-level features are not always powerful in representing the semantic concepts, the images similar in semantic contents are selected as positive examples among the first twenty retrieved set, in each round of feedback iteration and the remaining are negative examples for updating the weight parameters and revising the features. In most of the cases, CSD has been able to produce better results than EHD in terms of color distribution. EHD produced better results where spatial distribution of edges, is more important. However, with CSD descriptors the retrieval accuracy is limited when semantic significance is more important than color distribution only. One such example may be the case of an elephant where the improvement in precision is from 50 % - 75% at third iteration as shown in Figs. 1 -2. The weight updating approach is tested in terms of Average precision (with n=20 images) which is the average value obtained using eqn. 8 considering all randomly selected queries of the databases. The average precision obtained from the set of same query after different iterations are shown in Fig. 3(a). As seen from Fig. 3(a) that EHD and moment based methods generated almost similar results whereas CSD features performed better in terms of average retrieval precision. The experiments have been implemented on a Dell (T7400, 4GB, RAM) machine using MATLAB R2008a package. The cputime taken for each iteration is approximately 500ms. for database (A) and 3 secs for database (B).



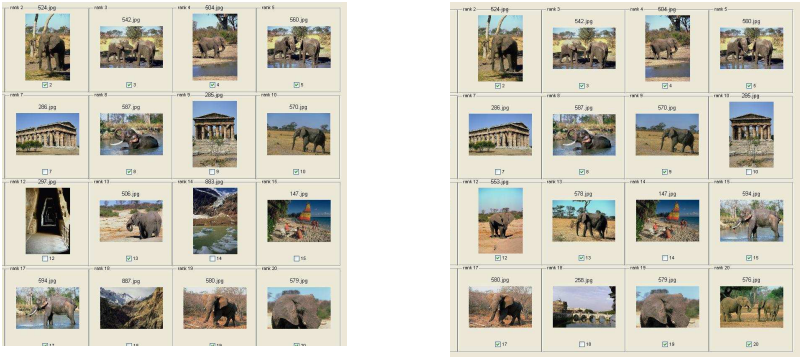
(a) (b)

Fig. 1. The proposed relevance feedback scheme on MPEG-7, CSD visual descriptors, (a)First pass (b) iteration1

The results for different queries are defined in terms of Precision as,

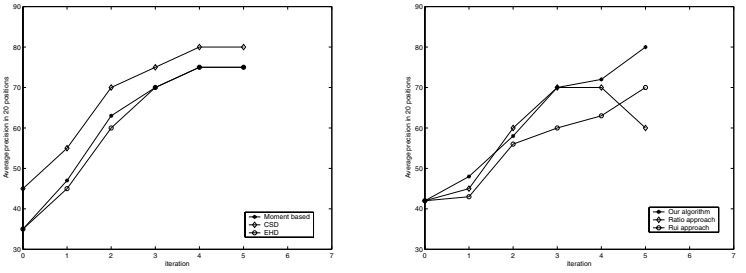
$$P(n) = \frac{target\ \ images\ \ within\ \ n\ \ positions}{n} \times 100 \tag{8}$$

The average precision as obtained from other two methods namely, (a) Ratio approach [16] in which the relative ratios of standard deviations of the component features of positive and negative examples are considered (b) Rui's method [6] in



(a) (b)

Fig. 2. The proposed relevance feedback scheme on MPEG-7, CSD visual descriptors, (a)iteration2 (b) iteration3



(a) (b)

Fig. 3. Comparative studies (a) using different feature extraction schemes (b) using different weight updation schemes

which a feature's weight is assigned in inverse proportion to the features variance across the images marked relevant is shown in Fig. 3(b). Experimental results show that, the proposed relevance feedback mechanism enhances the results using the MPEG-7 feature descriptors better than Ratio and Rui's method in most of the cases. As the computed FEI is the measure of heterogeneity in the relevant and irrelevant set of the component feature space it is able to improve the results consistently, with the S-type function in the interval $(0, +1)$.

3 Conclusion

Experimental results show that the proposed image retrieval system based on MPEG-7 features is able to improve the retrieval performance more than 20% in most of the cases, within two iterations of relevance feedback. Since, MPEG-7 Visual descriptors, describe features for effective image or video retrieval, effectiveness of the proposed feedback mechanism could be tested for video retrieval as future scope of research.

References

1. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
2. Cheng, K.O., Law, N.F., Siu, W.C.: Multiscale directional filter bank with applications to structured and random texture retrieval. Pattern Recognition 40(4), 610–621 (2007)
3. Martínez, J.C., Medina, J.M., Barranco, C.D., Perales, G., Hidalgo, J.M.S.: Retrieving images in fuzzy object-relational databases using dominant color descriptors. Fuzzy Sets and Systems 158(3), 312–324 (2007)
4. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content description Interface. John Wiley and Sons, Inc., Chichester (2002)
5. Manjunath, B.S., Ohm, J.R., Vasudevan, V.V., Yamada, A.: Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Technology 11(6), 703–715 (2001)
6. Rui, Y., Huang, T.S., Mehrotra, S.: Relevance feedback: a power tool for interactive content-based image retrieval. IEEE transactions on Circuits and Systems for Video technology 8(5), 644–655 (1998)
7. Yin, P.Y., Bhanu, B., Chang, K.C., Dong, A.: Integrating relevance feedback techniques for image retrieval using reinforcement learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1536–1551 (2005)
8. Lim, J.H., Jin, J.S.: Combining intra-image and inter-class semantics for consumer image retrieval. Pattern Recognition 38(6), 847–864 (2005)
9. Jin, Z., King, I., Li, X.Q.: Content-based retrieval by relevance feedback. In: Laurini, R. (ed.) VISUAL 2000. LNCS, vol. 1929, pp. 521–529. Springer, Heidelberg (2000)
10. Ves, E.D., Domingo, J., Ayala, G., Zuccarello, P.: A novel bayesian framework for relevance feedback in image content-based retrieval systems. Pattern Recognition 39(9), 1622–1632 (2006)
11. Qian, F., Zhang, B., Lin, F.: Constructive learning algorithm-based rbf network for relevance feedback in image retrieval. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 352–361. Springer, Heidelberg (2003)
12. He, X., King, O., Ma, W., Li, M., Zhang, H.J.: Learning a semantic space from user's relevance feedback for image retrieval. IEEE transactions on Circuits and Systems for Video technology 2003 13(1) (2003)
13. Banerjee, M., Kundu, M.K., Das, P.K.: Image Retrieval with Visually Prominent Features using Fuzzy Set Theoretic Evaluation. In: IEE International Conference on Visual Information Engineering VIE 2006, India, pp. 298–303 (2006)
14. Pal, S.K., Chakraborty, B.: Intraclass and interclass ambiguities (fuzziness) in feature evaluation. Pattern Recognition Letters 2, 275–279 (1984)
15. Pal, S.K., Majumder, D.D.: Fuzzy mathematical Approach to Pattern Recognition. Willey Eastern Limited, New York (1985)
16. King, I., Jin, Z.: Integrated probability function and its application to content-based image retrieval by relevance feedback. Pattern Recognition 36, 2177–2186 (2003)

Modelling the Effects of Internal Textures on Symmetry Detection Using Fuzzy Operators

Maurizio Cardaci[1,2], Filippo Millonzi[3], and Marco E. Tabacchi[3,4]

[1] Università degli Studi di Palermo, Dipartimento di Psicologia, Italy
[2] C.I.T.C. Centro Interdipartimentale di Tecnologie della Conoscenza, Palermo, Italy
[3] Università degli Studi di Palermo, Dip. di Matematica ed Applicazioni, Italy
[4] Istituto Nazionale di Ricerche Demopolis, Italy
{cardaci,millonzi,tabacchi}@unipa.it

Abstract. Symmetry is a crucial dimension which aids the visual system, human as well as artificial, to organize its environment and to recognize forms and objects. In humans, detection of symmetry, especially bilateral and rotational, is considered to be a primary factor for discovering and interacting with the surrounding environment. Rotational symmetry detecting can be affected by less-known factors, such as the stimulus internal texture. This paper explores how fuzzy operators can be usefully employed in modeling the effects of the internal texture on symmetry detection. To this aim, we selected two symmetry detection algorithms, based on different computational models, and compared their output with the outcome of an experiment in symmetry preferences on humans.

1 Introduction

Symmetry is a fundamental principle, which aids the visual system, human as well as artificial, to organize its environments and to recognize natural and artificial forms and objects. From a Gestaltist point of view, the law of symmetry stresses that we tend to perceive objects as symmetrical structures around a center, and it assigns a relevant role in attentive mechanism both in visual and auditory systems. In particular, by facilitating perceptual grouping [1], as well as figure/ground organization, symmetry is one of the most important factors allowing perceptual structures to emerge. Indeed, when we perceive disconnected but alike elements, that are symmetrical to each other, we tend to integrate them in a coherent percept. Moreover, in figure/ground segregation process, symmetrical images generally emerge as "figure", rather than as "ground". Symmetry detection is also highly relevant in shape recognition. Indeed, the description of a shape may be different when it is embedded in a context with horizontal or vertical symmetry [2]. Besides, in tasks requiring the completion of partially occluded visual stimuli, subjects tend to produce systematically symmetrical figures [3]. The concept of symmetry is not univocal: various kinds of properties of an image are defined as symmetry [4,5]; in this paper we will specifically focus on the *rotational symmetry*. A figure has rotational symmetry when it can be rotated less than 360° around its central point, or axis, and still match the original figure. In order to explore how fuzzy operators can be successfully employed in modeling the effects of such external cues as internal texture on symmetry detection, we compared the results obtained from an experiment

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 254–262, 2009.

on humans with the output of two algorithms based on different computational models, a field metaphor and a memetic algorithm, respectively. In this paper Sect. 2 briefly reports the results from the experiment on humans; Sect. 3 and 4 show the used algorithms and their results; finally, in Sect. 5 we compares the results and present the conclusions.

2 Human Judgement of Symmetry

From a human point of view, the vertical axis is predominant in rotational symmetry. Since early Mach's intuitions on the predominance of vertical symmetry axis in humans, the axial symmetry effects have been fully confirmed by the literature: symmetry detection is easier and faster when the symmetry axis is vertical [6]. More recent researches showed how detecting symmetry in human observers is influenced also by scanning strategies and attentional cueing effects [5]. Concerning this, several studies stressed the facilitating effects produced by external frames or cue lines aligned at the symmetry axis [7,8]. Other studies showed that in symmetry detection scanning strategies are also affected by qualitative features of the figure, i.e. presence of curvature minima along boundaries, concavity of the figure and number of vertices mismatches [9]. Moreover, symmetry judgment is faster and more accurate if the figure contains multiple symmetries. Other less studied variables could be relevant. In particular, the internal structure (texture) of stimulus and its interaction with the vertical bias has often been neglected.

2.1 An Experiment on Human Subjects

Here, are presented results obtained from an evaluative experiment with human subjects on equilateral triangles. Human observers had to choose the main symmetry axis on a equilateral triangle having global symmetries along three axes (30°, 90° and 150°). By manipulating internal texture, four triangles were generated (see Fig. 1): TR1 and TR2 globally maintain the three main symmetries, and are called tri-symmetric, whereas TR3 and TR4 maintain only one symmetry and are called mono-symmetric. All the test triangles have the same number of black areas (4) and the same ratio between gray and black areas (7:9). When the triangles are set as shown in Fig. 1, we call A, B and C their vertices starting from top in clockwise order (see Fig. 1(*a*)). It notice that we will use the same set of images and the same nomenclature as input for our algorithms.

2.2 Results

Statistical analysis revealed significant differences between responses to three-symmetry triangles vs one-symmetry ones. When three-symmetry triangles (as well

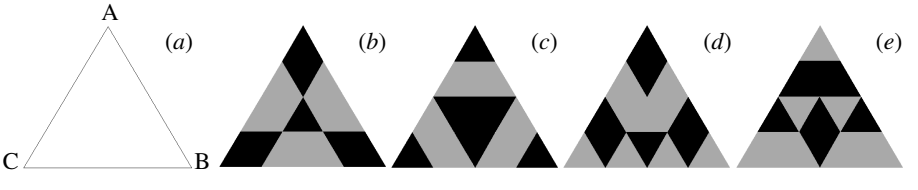


Fig. 1. Triangles used in the experiments. PRO (*a*); TR1 (*b*); TR2 (*c*); TR3 (*d*); TR4 (*e*).

as the blank one) were rotated 0° or 180°, the subjects selected the vertex A, following the vertical axis, while when they were rotated 40°, 320°, 140° or 220° no specific axis was preferred. Instead in one-symmetry triangles subjects tended to select the vertex A regardless of the rotation, hence the texture could have led the choice of the symmetry axis. These results indicate the possibility that human symmetry detection is due not only to vertical axis or to external frames, but also to internal structure of the stimulus.

3 Field Symmetry Transform

The Field Symmetry Transform (FiST) is a novel algorithmic approach to detect the symmetry in digital objects. The FiST algorithm takes as input an image I, and outputs an histogram of the symmetry distribution in the image itself; analysis of maxima and minima in the histogram reveals the main symmetric axes. The process employed by the algorithm is illustrated in Fig. 2: FiST treats the input image as a bidimensional plane (see Fig. 2(*a*)), in which each pixel p is a virtual charged particle at continue plane coordinates (x, y), with positive intensity proportional to the fuzzy grey intensity of *p*: e.g., in a black-object-on-white-background 8 bit image context, this means that a black pixel represents a positive unitary charged particle, while a pixel with intensity 128 (mid-grey) is considered as a positive charged particle of intensity $\frac{1}{2}$. Once all of the pixel have been represented as virtual charged particles, an equal-spaced orthogonal grid is superimposed on the plane (see Fig. 2(*b*)), and the vector field resulting from the contribution of all the virtual charges is computed at each crossing point of the grid (see Fig. 2(*c*)). The contribution given by the charged particle p lying at coordinates (x, y) to the field in a grid point g of coordinates (i, j), according to Coulomb's law, is:

$$E_p(g) = \frac{1}{4\pi}\frac{Q}{r^2}\hat{r}$$

where Q is the charge of p, and $\hat{r}$ is the unit vector pointing from the particle p to the evaluation point g, or from (x, y) to (i, j). Due to the superimposition principle of the charges, the total field vector E in p given by from all the points g of G is:

$$E(g) = \sum_{g \in G} E_p(g).$$

The vectors obtained through this process are then clustered according to their direction, and a histogram (see Fig. 2(*d*)) is obtained by counting the number of vectors in each cluster; the resulting histogram is the FiST of the original image, on which further analysis can be done in order to extract information pertaining the symmetry axes of the image. The choice of parameters for FiST (resolution of the grid, width of the clusters, linearity of the charge with respect to the pixel intensity) depends strictly on the kind of image in input, its resolution and bit depth. FiST has some interesting peculiarities: the algorithm is highly parallelizable, the only serial step is the construction of the final histogram; while not real-time for any reasonable sized input, the algorithm is quite fast. The fuzzy nature of the algorithm (pixel values are treated as fuzzy charges) means that the input images is not limited in bit depth, allowing the same implementation to manipulate high-bit depth images such as the ones in the medical imaging field; the

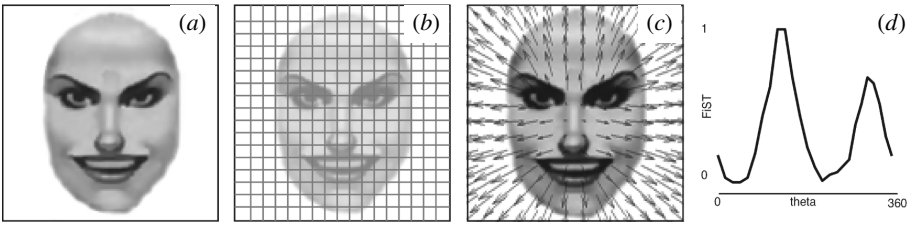


Fig. 2. The FiST process

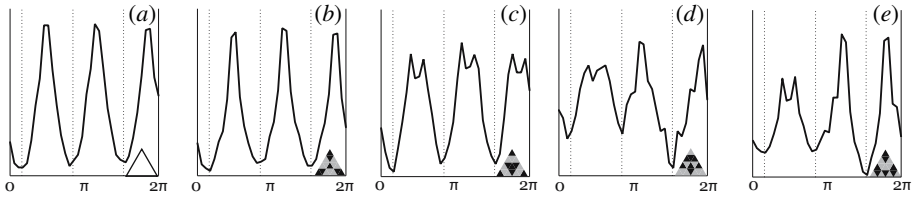


Fig. 3. Results using FiST. PRO (*a*); TR1(*b*); TR2 (*c*); TR3 (*d*); TR4 (*e*).

algorithm is not restricted to greyscale: since the charges' values are conventional, the method can be used as well with color images, providing some sensible interpretation of the chrominance data; the algorithm can easily be extended to include a pyramidal iterative version by varying the resolution of the superimposed grid.

3.1 Experiment

The triangles were rendered at an height of 400 pixels, in order to allow enough white space at the borders, and then placed in a 512×512 white pixels image box; rendering of the originally vectorial triangles used anti-aliasing techniques. Procedure FiST has been applied to our stimuli as follows: the FiST grid was set to 128×128 points, starting at the top left of the image, and in order to balance the contribution from each part of the image, the vectors considered were the ones lying in a circle centred on the image, with a 256 pixels radius. The interval considered for the histogram building phase was of $6°$, in order to obtain a smoothing of the interferences due to the digital nature of the image. The whole procedure was carried on using custom developed MATLAB code.

3.2 Results

As evident from the results shown in Fig. 3, in both the reference triangle PRO and the two tri-symmetric triangles TR1 and TR2, the histogram has almost-equal minima corresponding to the axes di $30°$, $90°$ and $150°$), and almost-equal maxima (or, in the case of TR2, double-peaked maxima) corresponding to the vertices of the triangle (at $90°$, $210°$ and $330°$). Meanwhile, the two mono-symmetric triangles TR3 and TR4 present a well-defined minimum only in correspondence to the axis passing through to

the vertex determining the symmetry. FiST results are invariant to rotation, although a gravity can be included by constant alteration of the field.

4 Memetic Symmetry Transform

The Memetic Symmetry Transform (MST) is an application of *memetic algorithms* [10,11] which are evolutionary algorithms inspired by the culture evolution in the human civilization and by models of adaptation in natural systems. They combine evolutionary adaptation of populations of individuals with individual learning within a lifetime.

4.1 Preliminaries

Agents. Let n be the number of grey or black equilateral triangles (*texture triangles*) adjacent to one side of the input triangle, the texture is a composition of n^2 texture triangles belonging to one of two disjointed sets: $\Delta - up$ or $\Delta - down$, triangles with one vertex directed upward or downward, respectively (see Fig. 4 *a*,*b*). We assign to each texture triangle one direction among those indicated by its vertices and we call *agent* an array $V = (v_1, v_2, \ldots v_{n^2})$ of integer, where $v_i = 1$, 2 or 3 is the direction of the $i - th$ texture triangle (see Fig. 4 c).

Fitness function. Let *V* be an agent and a_i the line with direction i, where $i = 1, 2, 3$, passing through the center of mass of the triangle, the value of the fitness function is:

$$f(V) = \max_{i=1,2,3} f_i(V) \qquad \text{where} \qquad f_i(V) = \sum_{j=1}^{n^2} \alpha_j^i + \sum_{j=1}^{\frac{n^2}{2}} \beta_j^i \quad i = 1, 2, 3$$

and the coefficients α_j^i and β_j^i are the following:

$$\alpha_j^i = \begin{cases} 3/2 & \text{if the } j - th \text{ texture triangle is black and its direction is parallel to } a_i \\ 1 & \text{if the } j - th \text{ texture triangle is gray and its direction is parallel to } a_i \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\beta_j^i = \begin{cases} +1/5 & \text{if the } j - th \text{ texture triangle has a simmetrical with respect to } a_i \\ -1/5 & \text{otherwise.} \end{cases}$$

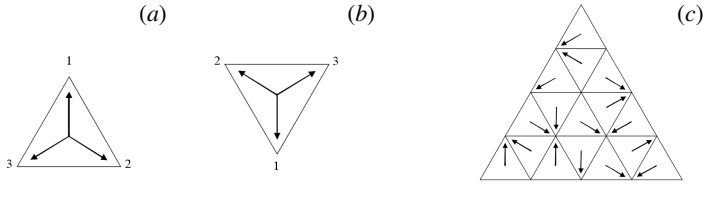


Fig. 4. $\Delta - up$ (*a*) and $\Delta - down$ (*b*) texture triangles with their three possible directions 1, 2 and 3. An example (*c*) of agent $V = (1, 2, 1, 1, 2, 3, 3, 2, 1, 2, 3, 3, 3, 2, 2, 3)$ with $n = 4$.

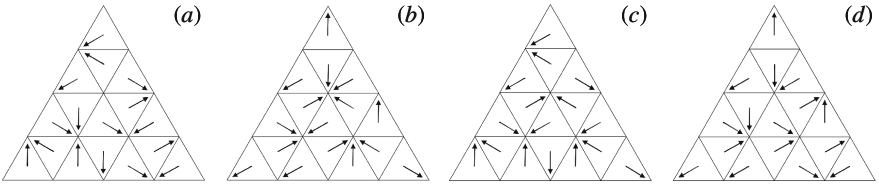


Fig. 5. Two parents (*a* and *b*) and their respective offsprings (*c* and *d*), when cut-points are $h = 4$ and $k = 11$

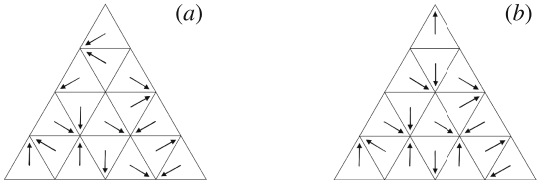


Fig. 6. Agent V (*a*) and mutated agent $\overline{V}$ (*b*) when T={5,13,14,16}

The values of α_j^i is equal to the ratio between the values of the colors of the texture (gray=169 and black=255), while the value of β_j^i has been chosen after several experimental sessions.

Operators. The crossover operator produces two new agents by combining two parents agents. It chooses two cut-points h and k, such that $1 \leq h \leq k \leq n^2$, in a random fashion and swaps the subsequences of the agents between the two cut-points. Formally, the offsprings obtained from $V = (v_1, v_2, \ldots, v_{n^2})$ and $W = (w_1, w_2, \ldots, w_{n^2})$ are $\tilde{V} = (v_1, \ldots, v_h, \ldots, w_k, \ldots, v_{n^2})$ and $\tilde{W} = (w_1, \ldots, w_h, \ldots, v_k, \ldots, w_{n^2})$ (Fig. 5). Mutation changes the orientations of some texture triangles. This operator randomly chooses a subset of indices T=$\{t_1, t_2, \ldots, t_m : 1 \leq t_i \leq n^2 \; \forall \, 1 \leq i \leq m\}$ and reassigns the directions of the corresponding texture triangles in V. Hence, the mutated agent will be $\overline{V} = (\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_{n^2})$, with $\overline{v}_i = v_i$ if $i \neq t_j$ (see Fig. 6).

Selection Function. The selection function selects the agents with the best fitness values. It has be noted that this function maintains constant the population size and it preserves the best agent produced during the evolution.

4.2 Algorithm

We designed the MST taking inspiration from the human subjects' behavior, who carry out their choices by combining a limited number of particular symmetries and colors arrangements. The algorithm generates a starting population P of m agents in a random fashion, then it evaluates each agent to find an optimal solution, that is an assignation of the same direction to all texture triangles; in the case of n^2 texture triangles the probability of finding it is $\frac{m}{3^{n^2-1}}$. If the solution is not in P then MST evolves P, by adding new agents obtained by applying both crossover and mutation to all individuals

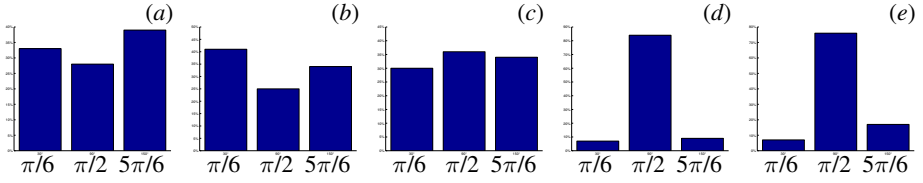


Fig. 7. Results using MST: PRO (*a*); TR1(*b*); TR2 (*c*); TR3 (*d*); TR4 (*e*)

of P. The fitness function evaluates the agents and the selection function takes the best individuals to create the next generation. If a solution is found then algorithm stops, else it continues the evolution until a solution is found or maximum number of generations ($gmax$) is reached. The MST algorithm can be sketched as follows:

MST
generating a population P of m agents;
evaluating the fitness for each agent in P;
$g \leftarrow 1$;
while no solution is in P and $g \leq gmax$
  $P' \leftarrow$ applying the crossover to each agent in P;
  $P'' \leftarrow$ applying the mutation to each agent in P;
  evaluating the fitness for each agent in $P' \bigcup P''$;
  $P \leftarrow$ selecting m best agents in $P \bigcup P' \bigcup P''$;
  $g \leftarrow g + 1$;
end;
return direction of the best agent in P;

According to fuzzy logic's principles [12], the MST does not determine exact solutions of the problem, but it often chooses as solution the direction of the agent having the best fitness value. Moreover, it has to be note that MST is triangle's rotation independent.

4.3 Experiment and Results

The experiments have been carried out by running the MST 100 times for each triangle in Fig. 1 and results are reported in Fig. 7. It is evident the strong correlation between human decision, when no external cues are given, and the behavior of MST. The human preference for vertex A in mono-symmetric triangles also with rotations different from $0°$ or $180°$ is matched by the preference for $90°$ orientation of MST; instead, in tri-symmetric triangles, rotated as above, there are no evident preferences among vertices in humans such as orientations in MST. We can consider each run as an human subject and each agent as a possible decision. Since the human subject has to choose among a limited number of changes by taking into account both personal bias and remarks, then both population size and maximum number of generations was set to 50.

5 Conclusion and Further Works

We analyzed a less-known aspect concerning the effects of the stimulus internal texture on the rotational symmetry detection. Results of an experiment carried out with human

subjects have been compared with those of symmetry detection algorithms, showing both interesting analogies and differences between human/artificial symmetry detection. In particular, it was possible to recognize a generalized behavioral pattern in human observers: exposed to tri-symmetric triangles rotated $0°$ or $180°$ human subjects were led by the vertical orientation, while, when these triangles were rotated $40°$, $320°$, $140°$ and $220°$, no axis was preferred. Also observing mono-symmetric triangles rotated $0°$ or $180°$, human subjects selected the vertical axis, but when these triangles were rotated $40°$, $320°$, $140°$ and $220°$, internal texture forced their choice toward the specific axis emphasized by the symmetry. These results suggest that when the texture appears as visually pregnant, it plays a role of reference frame, facilitating human observers' detection of the symmetry primary axis. About computational results (invariant to roto-translation), in this preliminary approach we noted that the algorithm based on geometrical or physical models, such as FiST, tends to provide a stable output when the input is constant, which means no disturbance from contingent factors usually present in human judgment (subjective preferences, external cues, etc). Contrarily, the evolutive based approach of MST is less precise in spotting physical aspects of symmetry devoid of contest, but it takes into account the whole visual experience. Since the human assessment of visual properties is subjective, in the framework we are dealing with, they can be measured only as mean of the different individuals' responses and hence each run of the MST can be considered as an individual. This leads us to believe that the symmetry detection algorithms cannot abstract from the usually cognitive dual approach (bottom-up or top-down). Hence, algorithms based on low-level features (such as FiST) are more sensible to physical and geometrical measures such in the case of texture, and they have properties of precision, repeatability and focus. Instead, high-level algorithms (such as MST), take into account bottom-up information as well as knowledge from external cues, both needed to the specific visual task. Such properties derive from the evolutive nature of the MST that searches for near optimal solutions in a quite similar way to the adaptive human behavior. These results have important applications ranging from the low level computer vision tasks or integration of stimuli, to building or improving tools able to measure and compare human perceptive abilities. From a cognitive point of view, further development will explore the role of other perceptual factors on the human symmetry judgment, while from a computational point of view, new studies will be aimed to the design of even more refined computational models of human perception.

References

1. Wertheimer, M.: Untersuchungen zur Lehre von der Gestalt II (Investigations into Gestalt theory II). Psychologische Forschung 4, 301–350 (1923)
2. Palmer, S.E.: The Role of Symmetry in Shape Perception. Acta Psychologica 59, 67–90 (1985)
3. van Lier, R., Wagemans, J.: From Images to Objects: Global and Local Completion of Self-Occluded Parts. Journal of Experimental Psychology: Human Perception and Performance 25, 1721–1741 (1999)
4. Zabrodsky, H.: Symmetry - A Review. Tech. Rep. 90–16, CS Dept., The Hebrew University of Jerusalem (1990)

5. Wenderoth, P.: The Salience of Vertical Symmetry. Perception 23, 221–236 (1994)
6. Palmer, S.E., Hemenway, K.: Orientation and Symmetry: Effects of Multiple, Rotational, and Near Symmetries. J. E. P.: H. P. and P. 4, 691–702 (1978)
7. Herbert, A.M., Humphrey, G.K., Jolicoeur, P.: The Detection of Bilateral Symmetry: Effects of Surrounding Frames. Canadian Journal of Experimental Psychology 48, 140–148 (1994)
8. Sekuler, A.B.: Axis of Elongation Can Determine Reference Frames for Object Perception. Ca. J. E. P. 50, 270–278 (1996)
9. Baylis, G.C., Driver, J.: Parallel Computation of Symmetry but not Repetition within Single Visual Shapes. Visual Cognition 1, 377–400 (1994)
10. Moscato, P.: On evolution, Search, Optimization, Genetic Algorithms and Martial Arts: towards Memetic Algorithms. Caltech Concurrent Computation Program, C3P Report 826 (1989)
11. Corne, D., Dorigo, M., Glover, F.: New Ideas in Optimization. McGraw-Hill, New York (1999)
12. Zadeh, L.: Fuzzy Sets. Information Control 8, 338–353 (1965)

Multivalued Background/Foreground Separation for Moving Object Detection

Lucia Maddalena[1] and Alfredo Petrosino[2]

[1] ICAR - National Research Council,
Via P. Castellino, 111, 80131 Naples, Italy
[2] DSA - University of Naples Parthenope,
Centro Direzionale, Isola C/4, 80143 Naples, Italy
lucia.maddalena@na.icar.cnr.it, alfredo.petrosino@uniparthenope.it

Abstract. The detection of moving objects is usually approached by background subtraction, i.e. by constructing and maintaining an up-to-date model of the background and detecting moving objects as those that deviate from such a model. We adopt a previously proposed approach to background subtraction based on self organization through artificial neural networks, that has been shown to well cope with several of the well known issues for background maintenance, featuring high detection accuracy for different types of videos taken with stationary cameras. Here we formulate a fuzzy approach to the background model update procedure to deal with decision problems typically arising when crisp settings are involved. We show through experimental results that higher accuracy values can be reached for color video sequences that represent typical situations critical for moving object detection.

Keywords: moving object detection, background subtraction, multivalued background modeling, self organization, neural network.

1 Introduction

Many computer vision applications, such as video surveillance or video compression, rely on the task of detecting moving objects in video streams, that provides the segmentation of the scene into background and foreground components.

The usual approach to moving object detection is through background subtraction, that consists in maintaining an up-to-date model of the background and detecting moving objects as those that deviate from such a model. Compared to other approaches, such as optical flow (e.g. [3]), this approach is computationally affordable for real-time applications. The main problem is its sensitivity to dynamic scene changes, and the consequent need for the background model adaptation via background maintenance. Such problem is known to be significant and difficult [14], and it has to take into account several well known issues in background maintenance, such as *light changes*, *moving background*, *cast shadows*, *bootstrapping*, and *camouflage*. Due to its pervasiveness in various contexts, background subtraction has been afforded by several researchers, and plenty of literature has been published (see surveys in [4,10,11], and more recently in [6]).

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 263–270, 2009.

In [9] we proposed the Self-Organizing Background Subtraction (SOBS) algorithm, which implements an approach to moving object detection based on the background model automatically generated by a self-organizing method without prior knowledge about the involved patterns. Such adaptive model can handle scenes containing moving backgrounds, gradual illumination variations and camouflage, has no bootstrapping limitations, can include into the background model shadows cast by moving objects, and achieves robust detection for different types of videos taken with stationary cameras.

One of the main issues to be pursued in background subtraction is the uncertainty in the detection caused by the cited background maintenance issues. Usually, crisp settings are needed to define the method parameters, and this does not allow to properly deal with uncertainty in the background model. Recently several authors have explored the adoption of fuzzy approaches to tackle different aspects of detecting moving objects. In [16] an approach using fuzzy Sugeno integral is proposed to fuse texture and color features for background subtraction, while in [2] the authors adopt the Choquet integral to aggregate the same features. In [12] a fuzzy approach to selective running average background modeling is proposed, and in [1] the authors model the background by the Type-2 Fuzzy Mixture of Gaussian Model proposed in [15].

Here we propose a fuzzy approach to the background model update procedure of SOBS algorithm, where a fuzzy function is computed, pixel-by-pixel, on the basis of the background subtraction phase. The idea is to introduce into the update phase an automatic and data dependent mechanism for further reinforcing into the background model the contribution of pixels that belong to it. It will be shown that the fuzzy approach, implemented in what will be called MSOBS (Multivalued SOBS) algorithm, further improves the accuracy of the corresponding crisp moving object detection procedure.

The paper is organized as follows. In Section 2 we detail the proposed fuzzy approach to moving object detection, describing the basic model adopted from [9] and the proposed modifications. In Section 3 we give a qualitative and quantitative evaluation of the proposed approach accuracy, comparing obtained results with those obtained by the crisp analogous approach. Conclusions are drawn in Section 4.

2 MSOBS Algorithm

The background model constructed and maintained in SOBS algorithm [9], here adopted, is based on a self organizing neural network, inspired by Kohonen [7], organized as a 2-D flat grid of neurons. Each neuron computes a function of the weighted linear combination of incoming inputs, with weights resembling the neural network learning, and can be therefore represented by a weight vector obtained collecting the weights related to incoming links. An incoming pattern is mapped to the neuron whose set of weight vectors is most similar to the pattern, and weight vectors in a neighborhood of such node are updated.

For each pixel we build a neuronal map consisting of $n \times n$ weight vectors, all represented in the HSV color space, that allows to specify colours in a way that

is close to human experience of colours. Each weight vector $c_i, i = 1, \ldots, n^2$, is therefore a 3D vector, initialized to the HSV components of the corresponding pixel of the first sequence frame I_0. The complete set of weight vectors for all pixels of an image I with N rows and M columns is represented as a neuronal map $\tilde{B}$ with $n \times N$ rows and $n \times M$ columns, where adjacent blocks of $n \times n$ weight vectors correspond to adjacent pixels in image I.

By subtracting the current image from the background model, each pixel p_t of the t-th sequence frame I_t is compared to the current pixel weight vectors to determine if there exists a weight vector that matches it. The best matching weight vector is used as the pixel's encoding approximation, and therefore p_t is detected as foreground if no acceptable matching weight vector exists; otherwise it is classified as background.

Matching for the incoming pixel p_t is performed by looking for a weight vector c_m in the set $C = (c_1, \ldots, c_{n^2})$ of the current pixel weight vectors satisfying:

$$d(c_m, p_t) = \min_{i=1,\ldots,n^2} d(c_i, p_t) \leq \epsilon \ , \tag{1}$$

where the metric $d(\cdot)$ and the threshold ϵ are suitably chosen as in [9].

The best matching weight vector $c_m = \tilde{B}_t(\overline{x}, \overline{y})$ and all other weight vectors in a $n \times n$ neighborhood of the background model $\tilde{B}$ are updated according to selective weighted running average:

$$\tilde{B}_t(i,j) = (1-\alpha_{i,j}(t))\tilde{B}_{t-1}(i,j)+\alpha_{i,j}(t)p_t(x,y), \quad \begin{array}{l} i = \overline{x} - \lfloor \frac{n}{2} \rfloor, \ldots, \overline{x} + \lfloor \frac{n}{2} \rfloor \\ j = \overline{y} - \lfloor \frac{n}{2} \rfloor, \ldots, \overline{y} + \lfloor \frac{n}{2} \rfloor \end{array} \tag{2}$$

Values for $\alpha_{i,j}(t)$ chosen in [9] can be expressed as

$$\alpha_{i,j}(t) = M(p_t)\ \alpha(t)\ w_{i,j} \ , \tag{3}$$

where $w_{i,j}$ are Gaussian weights in the $n \times n$ neighborhood, $\alpha(t)$ represents the learning factor, that is the same for each pixel of the t-th sequence frame and depends on scene variability, and $M(p_t)$ is the crisp hard-limited function

$$M(p_t) = \begin{cases} 1 & \text{if} \quad d(c_m, p_t) \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

that gives the background/foreground segmentation for pixel p_t.

It should be observed that if the best match c_m is not found, the background model $\tilde{B}$ remains unchanged. Such selectivity allows to adapt the background model to scene modifications without introducing the contribution of pixels not belonging to the background scene.

In this paper we propose to substitute the crisp function $M(\cdot)$ in eq. (3) with a fuzzy function that allows to take into account uncertainty in the background model. Specifically, we modify eq. (3) as follows:

$$\alpha_{i,j}(t) = (1 - F(p_t))\ \alpha(t)\ w_{i,j} \ , \tag{5}$$

where $F(p_t)$ is a saturating linear function given by

$$F(p_t) = \begin{cases} \dfrac{d(c_m, p_t)}{\varepsilon} & \text{if} \quad d(c_m, p_t) \leq \varepsilon \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

Values of the function $F(\cdot)$ are normalized in $[0, 1]$; the closer is the incoming sample p_t to the background model $C = (c_1, c_2, \ldots, c_{n^2})$, the smaller is the corresponding value $F(p_t)$. Therefore, choosing $\alpha_{i,j}(t)$ as in eq. (5) ensures that the closer is the incoming sample p_t to the background model, the more it contributes to the background model update, thus further reinforcing the corresponding weight vectors.

Other choices for learning factors $\alpha_{i,j}(t)$ as a function of $F(p_t)$ could have been considered according to the above considerations; for example, in [12] the authors propose a fuzzy running average approach where learning factors are chosen as an exponential function that, adapted to our case and notation, is given by

$$\alpha_{i,j}(t) = \exp(-5 * F(p_t))\ \alpha(t)\ w_{i,j}\ . \tag{7}$$

Summarizing, the background subtraction and update procedure considered in [9], as well as the modified versions proposed in the present paper, can all be stated in a similar way. Given an incoming pixel value p_t in sequence frame I_t, the estimated background model $\tilde{B}_t$ is obtained through the following algorithm:

Background subtraction and update algorithm

`Initialize weight vectors` C `for pixel` p_0 `and store it into` $\tilde{B}_0$
for `t=1, LastFrame`
  `Find best match` c_m `in` C `to current sample` p_t `as in eq. (1)`
  `Compute learning factors` $\alpha_{i,j}(t)$
  `Update` $\tilde{B}_{t-1}$ `in the neighborhood of` c_m `as in eq. (2)`
endfor

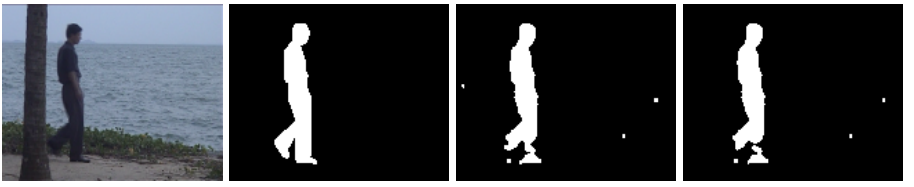
The original crisp SOBS algorithm is obtained if learning factors $\alpha_{i,j}(t)$ for the update step are chosen as in eq. (3), while the proposed multivalued algorithm, in the following denoted as MSOBS, is obtained if learning factors are chosen as in eq. (5). An alternative version of the multivalued algorithm, in the following denoted as MSOBS2, can be obtained if learning factors are chosen as in eq. (7). Results for all such algorithms will be compared in the following Section 3.

3 Experimental Results

Experimental results for moving object detection using the proposed approach have been produced for several image sequences. Here we describe two different publicly available sequences[1], that represent typical situations critical for moving object detection, and present qualitative and quantitative results obtained with the proposed method.

Sequence WS (*Water Surface*), where a person walks at a waterfront, has been chosen in order to test our method ability to cope with moving background (the

[1] http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html



(a) (b) (c) (d)

Fig. 1. Segmentation of sequence WS: (a) test image; (b) ground truth; (c) MSOBS detection mask; (d) SOBS detection mask

water surface). The sequence contains 633 frames of 160 × 128 spatial resolution. One representative frame is reported in Fig. 1-(a) and the corresponding hand-segmented background in Fig. 1-(b).

The indoor scene of sequence MR (*Curtain*) consists in an initially empty meeting room, with a curtain slightly blowing in the wind, where a man comes in and starts making his presentation. The sequence consists of 2964 frames of 160 × 128 spatial resolution, and we consider the hand-segmented background mask available for frame 1773. The considered test image and the related binary ground truth are reported in Figs. 2-(a) and 2-(b), respectively.

(a) (b) (c) (d)

Fig. 2. Segmentation of sequence MR: (a) test image; (b) ground truth; (c) MSOBS detection mask; (d) SOBS detection mask

The foreground masks computed by the proposed MSOBS algorithm are reported in Fig. 1-(c) for sequence WS and in Fig. 2-(c) for sequence MR, and those computed by SOBS algorithm are reported in Figs. 1-(d) and 2-(d), respectively. From such results it can be observed that both MSOBS and SOBS algorithms were quite successful in modeling the moving background (water surface and blowing curtain) and in detecting the moving person, both in the outdoor and in the indoor scene.

Results obtained by the proposed MSOBS algorithm have been compared with those obtained by the corresponding crisp SOBS algorithm in terms of different pixel-based metrics, namely *Precision, Recall*, and F_1.

Recall, also known as *detection rate*, gives the percentage of detected true positive pixels as compared to the total number of true positive pixels in the ground truth:

$$Recall = \frac{tp}{tp + fn},$$

where tp is the total number of *true positive* pixels, fn is the total number of *false negative* pixels, and $(tp + fn)$ indicates the total number of pixels present in the ground truth.

Recall alone is not enough to compare different methods, and is generally used in conjunction with *Precision*, also known as *positive prediction*, that gives the percentage of detected true positive pixels as compared to the total number of pixels detected by the method:

$$Precision = \frac{tp}{tp + fp},$$

where fp is the total number of *false positive* pixels and $(tp + fp)$ indicates the total number of detected pixels.

Using the above mentioned metrics, generally a method is considered *good* if it reaches high *Recall* values, without sacrificing *Precision*.

Moreover, we considered the F_1 metric, also known as *Figure of Merit* or *F-measure*, that is the weighted harmonic mean of *Precision* and *Recall*:

$$F_1 = \frac{2 * Recall * Precision}{Recall + Precision}.$$

Such measure allows to obtain a single measure that can be used to "rank" different methods.

All the above considered measures attain values in [0, 1], and the higher is the value, the better is the accuracy.

Accuracy values obtained by MSOBS and SOBS algorithms for sequences WS and MR are reported in Table 1. Here we can observe that both algorithms perform quite well, and that MSOBS performs slightly better than SOBS for both the sequences. More specifically, we can observe that the fuzzy approach achieves higher Recall values, but correspondingly lower Precision values. This is due to the fact that the fuzzy approach indeed allows to reinforce the contribution to the updating of the background model of pixels close to the model, thus leading to higher Recall values. At the same time, however, the fuzzy approach reinforces also the contribution of false positive pixels, thus reducing the Precision values. Nonetheless, results obtained by MSOBS algorithm are to be preferable to those obtained with the crisp approach, as shown by higher F_1 values.

Moreover, concerning different possible choices of learning factors for background updating, in Table 1 we also compare results obtained with MSOBS and with MSOBS2. Accuracy results attained are quite similar, and therefore we prefer to adopt learning factors defined as in eq. (5), since their computation is less computationally demanding.

Since we have already shown [9] that SOBS results are generally more accurate than those obtained by several state-of-the-art background subtraction algorithms (such as *BNN* [5], *Mixture of Gaussian* [13], and the method of Li et al. [8]), we can conclude that the proposed MSOBS algorithm compares favorably with them, too.

Table 1. Accuracy values for sequences WS and MR

	WS			MR		
	Recall	Precision	F_1	Recall	Precision	F_1
SOBS	0.8606	0.9684	0.9113	0.8751	0.9496	0.9108
MSOBS	0.8788	0.9571	0.9163	0.8653	0.9679	0.9138
MSOBS2	0.8844	0.9402	0.9114	0.8968	0.9288	0.9125

4 Conclusions

In this paper we propose to extend a previously proposed method for moving object detection [9] by introducing a fuzzy learning factor into the background model update procedure. The adopted method is based on self organization through artificial neural networks, and implements the idea of using visual attention mechanisms to help detecting objects that keep the user attention in accordance with a set of predefined scene features. Here we introduce a fuzzy function, computed pixel-by-pixel on the basis of the background subtraction phase. Such function is used in the background model update phase, providing an automatic and data dependent mechanism for further reinforcing into the background model the contribution of pixels that belong to it. It has been shown that the proposed fuzzy approach further improves the accuracy of the corresponding crisp moving object detection procedure, providing experimental results on real color video sequences that represent typical situations critical for moving object detection.

References

1. Baf, F.E., Bouwmans, T., Vachon, B.: Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling. In: 4th International Symposium on Visual Computing, Las Vegas, USA, pp. 772–781 (2008)
2. Baf, F.E., Bouwmans, T., Vachon, B.: Fuzzy Integral for Moving Object Detection. In: IEEE International Conference on Fuzzy Systems, Hong-Kong, China, pp. 1729–1736 (2008)
3. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of Optical Flow Techniques. International Journal of Computer Vision 12, 42–77 (1994)
4. Cheung, S.-C., Kamath, C.: Robust Techniques for Background Subtraction in Urban Traffic Video. In: SPIE Electronic Imaging - Video Communications and Image Processing, pp. 881–892 (2004)
5. Culibrk, D., Marques, O., Socek, D., Kalva, H., Furht, B.: Neural Network Approach to Background Modeling for Video Object Segmentation. IEEE Transactions on Neural Networks 18, 1614–1627 (2007)
6. Elhabian, S.Y., El-Sayed, K.M., Ahmed, S.H.: Moving Object Detection in Spatial Domain using Background Removal Techniques - State-of-Art. Recent Patents on Computer Science 1, 32–54 (2008)

7. Kohonen, T.: Self-Organization and Associative Memory, 2nd edn. Springer, Berlin (1988)
8. Li, L., Huang, W., Gu, I.Y.-H., Tian, Q.: Statistical Modeling of Complex Backgrounds for Foreground Object Detection. IEEE Transactions on Image Processing 13, 1459–1472 (2004)
9. Maddalena, L., Petrosino, A.: A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. IEEE Transactions on Image Processing 17, 1168–1177 (2008)
10. Piccardi, M.: Background Subtraction Techniques: a Review. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 3099–3104 (2004)
11. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image Change Detection Algorithms: A Systematic Survey. IEEE Transactions on Image Processing 14, 294–307 (2005)
12. Sigari, M.H., Mozayani, N., Pourreza, H.R.: Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application. International Journal of Computer Science and Network Security 8, 138–143 (2008)
13. Stauffer, C., Grimson, W.E.L.: Adaptive Background Mixture Models for Real-Time Tracking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 246–252 (1999)
14. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: Principles and Practice of Background Maintenance. In: Seventh IEEE Conference on Computer Vision, pp. 255–261 (1999)
15. Zeng, J., Xie, L., Liu, Z.: Type-2 Fuzzy Gaussian Mixture Models. Pattern Recognition 41, 3636–3643 (2008)
16. Zhang, H., Xu, D.: Fusing Color and Texture Features for Background Model. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 887–893. Springer, Heidelberg (2006)

Periodic Pattern Detection for Real-Time Application

Giovanni Puglisi and Sebastiano Battiato

Dipartimento di Matematica e Informatica
University of Catania, Italy
{puglisi,battiato}@dmi.unict.it

Abstract. Digital video stabilization approaches typically degrade their performances in presence of periodic patterns. Any kind of matching between consecutive frames is not usually able to work in presence of these kind of signals: the motion estimation engine is deceived and its performances degrade abruptly. In this paper we propose a fast fuzzy classifier able to recognize periodic and aperiodic pattern in the images that takes into account the peculiarities of digital video stabilization. Finally, the proposed classifier can be used as a filtering module in a block based video stabilization approach.

Keywords: Video Stabilization, periodic pattern, fuzzy classifier.

1 Introduction

In the last years video stabilization techniques have gained consensus, for they permit to obtain high quality and stable video footages even in non-optimal conditions. The best techniques, by using some mechanical tools, measure camera shake and then control the jitter acting on lens or on the CCD/CMOS sensor [1]. On the other hand, digital video stabilization techniques [2,3,4,5,6] make use only of information drawn from footage images and do not need any additional knowledge about camera physical motion.

Digital video stabilization systems have been widely investigated and several techniques have been proposed, with different issues and weak points. However in presence of regular or near regular texture, digital video stabilization approaches typically fail. These patterns, due to their periodicity, create multiple matching that degrade motion estimator performances. Even if some interesting approaches able to reliable find near regular texture have been recently developed [7,8], they are pretty complex and cannot be applied in real-time digital video stabilization. In this paper we propose a fast fuzzy classifier able to find regular and low distorted near regular texture tacking into account video stabilization peculiarities.

The rest of the paper is organized as follows. In Section 2 the analysis of the regular texture is performed. In Section 3 the classifier effectiveness applied to the video stabilization problem is shown whereas conclusions are summarized in Section 4.

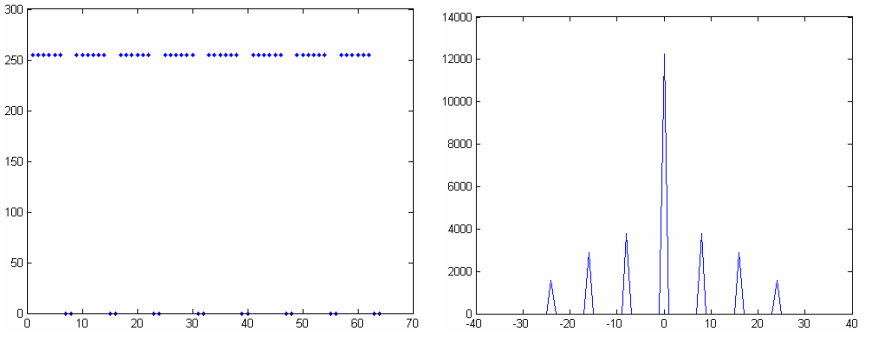
V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 271–278, 2009.

2 Regular Texture Analysis

In real images we can find many regular and near regular texture such us: buildings, wallpapers, windows, floors, etc. In particular regular texture and low distorted near regular texture, due to the multiple matching candidates, typically create a lot of problems to motion estimation algorithms. On the contrary, in presence of high distorted near regular texture (very often created by perspectively skewed patterns) video stabilization algorithms typically work in the correct way. Due to the limited number of samples in each selected patch and the spectral leakage that disperses frequencies over the entire spectrum a simple analysis (with proper thresholding) on Fourier peaks cannot be done. Also methods based on a simple analysis of bank filters (e.g., Gabor, etc.) are not able to properly detect the presence of such regions.

In this paper we propose a fuzzy classifier able to detect this kind of pattern in presence of some predefined constraints. It is based on Fourier domain analysis taking into account the following considerations (Fig. 1, 2):

1. The highest Fourier spectrum values of a periodic signal have a greater distance from the axes origin than aperiodic signal values;
2. Fourier components of periodic signals typically have a lower density than aperiodic signal values.



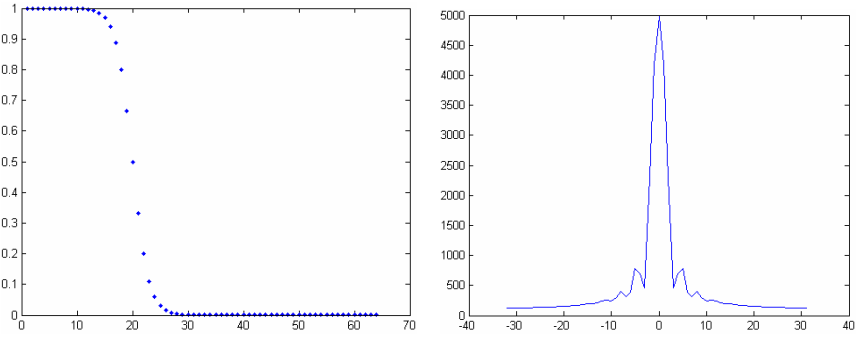
(a) (b)

Fig. 1. Example of 1-D periodic signal (a) with its spectrum (b)

The classifier makes use of the following two formulas:

$$distance = \frac{\sum_{i=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{j=-\frac{M}{2}}^{\frac{M}{2}-1} f_a(i,j)d(i,j)}{\sum_{i=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{j=-\frac{M}{2}}^{\frac{M}{2}-1} f_a(i,j) - f_a(0,0)} \tag{1}$$

where $d(i,j)$ is the Euclidean distance from the axes origin and $f_a(i,j)$ is the Fourier spectrum component (defined in $[-\frac{N}{2}, \frac{N}{2}-1]\times[-\frac{M}{2}, \frac{M}{2}-1]$) of a sequence of size $N \times M$.



(a) (b)

Fig. 2. Example of 1-D aperiodic signal (a) with its spectrum (b)

$$density = \frac{\sum_{k=1}^{componentsNumber} neighbors(k)}{componentsNumber} \tag{2}$$

where $componentsNumber$ represents the number of non-zero values of the Fourier spectrum and $neighbors(k)$ the number of non-zero values close to the component k. The concept of closeness depends on the constraints of the particular application.

The noise contribution in the formulas described above has been reduced considering only the most important Fourier component values. We discard all the values less than 30% of the maximum without considering the DC (Direct Current) component.

3 Regular Texture Fuzzy Classifier

The formulas (1) and (2) can be effectively used as discriminant features in a simple fuzzy classifier with rules listed in Table 1.

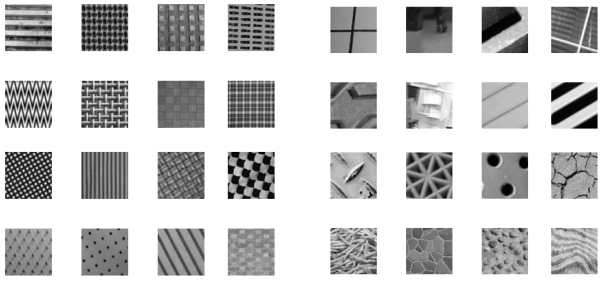
The membership values of the fuzzy system have been derived considering the peculiarities of the particular application. We consider a video stabilization technique using a BM (block matching) estimation module with block size 16×16 and search range ± 16 pixels. Block size defines the upper limit of periodic signal to be detected. The only periodic signals that must be taken into account, in this case, have a period less than 17 pixels.

A proper dataset containing both periodic and aperiodic images has been built by considering both synthetic and real texture downloaded from [1] [2] [3]. All the dataset (200 images) has been manually labeled in two classes: periodic and aperiodic. In the aperiodic group are also present corners, edges, regular texture

[1] http://www.ux.uis.no/ tranden/brodatz.html
[2] http://texturewarehouse.com/
[3] http://www.mayang.com/textures/

Table 1. Fuzzy rules of the system

	distance		*density*		*periodicity*
if	*Low*	and	*Low*	then	Low_1
if	*Low*	and	*Medium*	then	Low_2
if	*Low*	and	*High*	then	*VeryLow*
if	*Medium*	and	*Low*	then	$High_1$
if	*Medium*	and	*Medium*	then	*Medium*
if	*Medium*	and	*High*	then	Low_3
if	*High*	and	*Low*	then	*VeryHigh*
if	*High*	and	*Medium*	then	$High_2$
if	*High*	and	*High*	then	Low_4



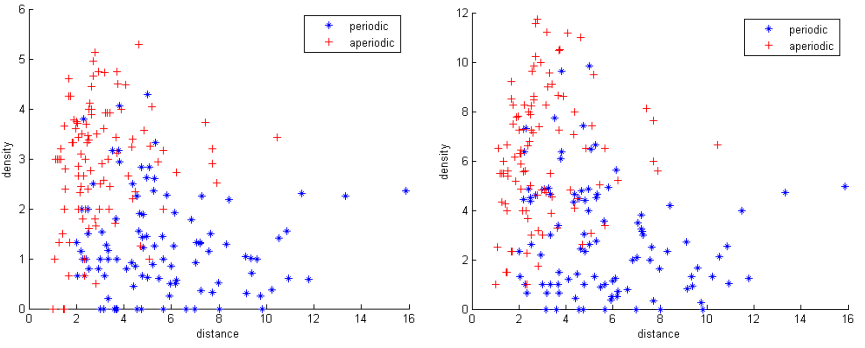
(a) (b)

Fig. 3. Some periodic (a) and aperiodic (b) images belonging to our dataset (comprising images with period greater than 16 pixels)

with period greater than 16 pixels (our motion estimation algorithm, due to its local view considers them aperiodic) and irregular texture (Fig. 3).

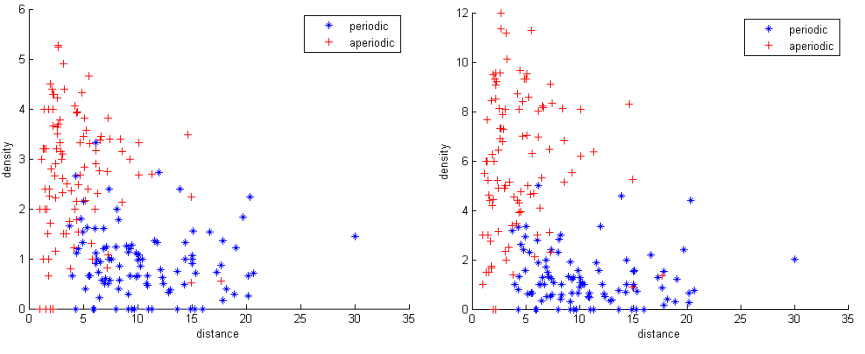
In order to obtain better classification performances we have analyzed the distribution of periodic and aperiodic images by considering different image dimension and neighborhood size (Equation 2). As can be seen from Fig. 4, 5 image dimension equal to (64×64 pixels) and neighborhood size equal to 2 pixels provide the best distribution: periodic and aperiodic images are divided pretty well.

The training process, devoted to find membership parameters, has been performed using a continuous GA (genetic algorithm), an optimization and search technique based on the principle of genetics and natural selection. An initial population, usually randomly selected, of possible solutions evolves toward a better solution. In each step some population elements are stochastically selected based on their fitness (the function to be optimized), and new elements are created through some techniques inspired by evolutionary biology (mutation, crossover). Genetic algorithms have found application in many fields [9]: computer science, engineering, economics, chemistry, physics, etc. Notice that for training simplicity we have considered a Sugeno fuzzy model.



(a) (b)

Fig. 4. Periodic and aperiodic images (32×32 pixels) in the features space just considering the neighborhood size equal to 1 (a) and 2 (b)



(a) (b)

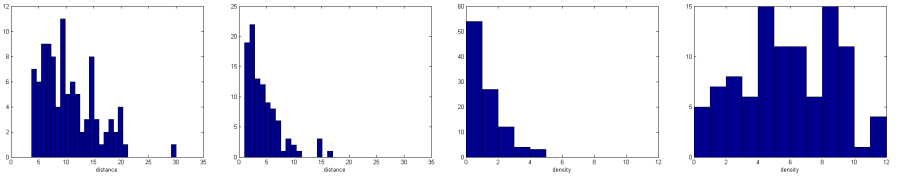
Fig. 5. Periodic and aperiodic images (64×64 pixels) in the features space just considering the neighborhood size equal to 1 (a) and 2 (b)

Genetic optimization is realized by using standard approaches in the field. In particular we have used default crossover and mutation algorithms provided by Genetic Toolbox functions of MATLAB 7. In order to speed-up the overall process, the initial population has been carefully defined. For our purposes the following considerations can be done (Fig. 6):

1. the minimum *distance* value of periodic patterns typically is greater than 4;
2. periodic signal typically have *density* values minor than 3.

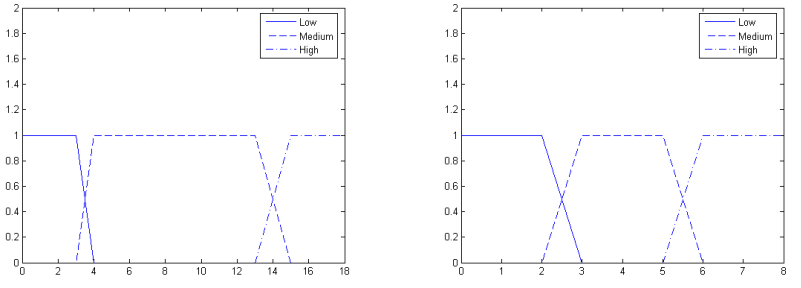
The initial population elements have been derived through a gaussian random perturbation of the parameters shown in Fig. 7 and Tab. 2.

To validate our classifier we have performed a leave-one-out cross-validation. A single data is considered the validation dataset, and the remaining data the



(a) (b) (c) (d)

Fig. 6. Histograms of *distance* (a, b) and *density* (c, d) features for periodic (a, c) and aperiodic signals (b, d)



(a) (b)

Fig. 7. Fuzzy membership functions for *distance* (a) and *density* (b) features

Table 2. Fuzzy membership output values

membership	$VeryLow$	Low_1	Low_2	Low_3	Low_4	$Medium$	$High_1$	$High_2$	$VeryHigh$
value	0	0.25	0.1	0.1	0.25	0.5	0.75	0.75	1

Table 3. Confusion matrix

	periodic	aperiodic
periodic	95	5
aperiodic	9	91

training dataset. Such procedure has been repeated until each data has been used as validation dataset.

For each input signal our fuzzy system produces a value belonging to [0-1] that is related to its degree of periodicity. In our case we choose as defuzzification strategy a simple thresholding process (threshold equal to 0.5). Table 3 reports the relative confusion matrix that confirms the robustness of the method for both classes, reaching an overall accuracy of 93%.

The proposed classifier can be used as a filtering module in the video stabilization algorithms. Each region classified as periodic should be removed before starting the actual video stabilization process. In order to confirm the effectiveness of the classifier for this purpose we have compared the performances of a block based video stabilization approach [2] with and without periodic patterns

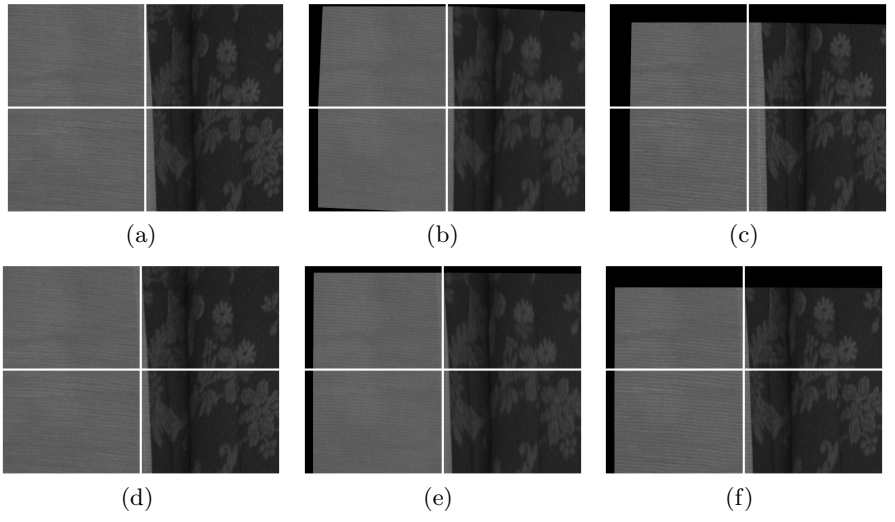


Fig. 8. Stabilized images obtained by [2] with (d, e, f) and without (a, b, c) the periodic pattern removal step. The grid is overlaid for better visualization.

removal step. As can be easily seen from Fig. 8, periodic patterns degrade [2] performances. On the contrary by simply using the filtering step proposed in this paper the video stabilization systems works pretty well.

4 Conclusions

In this paper we have proposed a novel fast fuzzy classifier for low distorted near regular texture detection (and removal). For each input signal our system produces a value belonging to [0-1] that is related to its degree of periodicity. The classifier has been validated with a leave-one-out cross-validation obtaining an accuracy of 93%.

Future works will be devoted to extend this approach to other motion estimation constraints. Also feature extraction on DCT domain will be analyzed.

References

1. Lukac, R.: Single-Sensor Imaging: Methods and Applications for Digital Cameras. CRC Press, Boca Raton (2008)
2. Battiato, S., Bruna, A.R., Puglisi, G.: A robust video stabilization system by adaptive motion vectors filtering. In: Proc. of ICME, Hannover, Germany (June 2008)
3. Battiato, S., Gallo, G., Puglisi, G., Scellato, S.: SIFT features tracking for video stabilization. In: Proc. of ICIAP, Modena (Italy), pp. 825–830 (September 2007)
4. Chang, J., Hu, W., Cheng, M., Chang, B.: Digital image translational and rotational motion stabilization using optical flow technique. IEEE Transactions on Consumer Electronics 48(1) (February 2002)

5. Mercenaro, L., Vernazza, G., Regazzoni, C.: Image stabilization algorithms for video-surveillance application. In: Proc. of ICIP, Thessaloniki (Greece) (October 2001)
6. Tico, M., Alenius, S., Vehvilainen, M.: Method of motion estimation for image stabilization. In: Proc. of IEEE ICASSP, Toulouse (France) (May 2006)
7. Hays, J.H., Leordeanu, M., Efros, A.A., Liu, Y.: Discovering texture regularity as a higher-order correspondence problem. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 522–535. Springer, Heidelberg (2006)
8. Liu, Y., Lin, W.C., Hays, J.H.: Near regular texture analysis and manipulation. ACM Transactions on Graphics 23(3), 368–376 (2004)
9. Haupt, R.L., Haupt, S.E.: Practical Genetic Algorithms. John Wiley & Sons, Hoboken (2004)

A System for Deriving a Neuro-Fuzzy Recommendation Model

Giovanna Castellano, Anna Maria Fanelli, and Maria Alessandra Torsello

CILab - Computational Intelligence Laboratory
Computer Science Department, University of Bari,
Via E. Orabona, 4 - 70126 Bari, Italy
{castellano,fanelli,torsello}@di.uniba.it
http://www.di.uniba.it/~cilab/

Abstract. In this paper, we present a system designed to discover recommendation fuzzy rules useful to provide personalized link suggestions to the users of a Web site. The system is mainly based on two processes. A fuzzy clustering process is applied to identify user categories by grouping users with similar interests. Then, a neuro-fuzzy strategy is applied to derive a set of recommendation fuzzy rules. A tool for the proposed system provides a wizard-based interface made of a sequence of panels that support users in the overall rule extraction process. An illustrative example is provided to show the performance of the system through the use of the developed tool.

Keywords: Web mining, Fuzzy clustering, access log, Web personalization, user profiling.

1 Introduction

The proliferation of information available on the Web has prompted the need for Web personalization. Recommendation systems represent the most notable application of Web personalization. Such systems attempt to meet the interests of their users by suggesting them information/services that they need without explicitly asking for them [7]. Hence, in the development of a recommendation system, the process of knowledge discovery from Web data covers a fundamental role. In literature, the Web Usage Mining (WUM) methodology was widely adopted to discover information about the user preferences that can be used to derive a knowledge base (i.e. recommendation model) useful to determine recommendations. In general, WUM involves the application of Data Mining techniques on usage data (characterizing the interactions of users with the Web site) to extract interesting patterns in the user navigational behavior [5], [9]. One of the most important sources of usage data is represented by log files storing all the information about the accesses made by users to a Web site. Starting from these, it is possible to understand the user interests and to identify user categories by grouping together users exhibiting similar behavior. This knowledge can be conveniently employed to derive a model that could be exploited to online recommend interesting items.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 279–286, 2009.

In this paper, we present a system for deriving a neuro-fuzzy recommendation model from user behavior data identified by preprocessing log files. The system, called REXWERE (Rule EXtraction for WEb REcommendation), employs a hybrid approach based on the combination of fuzzy reasoning and neural learning to extract knowledge in two successive phases: user categorization and recommendation model discovery. In user categorization, users with similar interests are grouped into clusters (user categories) through a fuzzy clustering approach. Then, a neuro-fuzzy strategy is applied to learn fuzzy rules capturing the association between user behaviors and Web pages to recommend. In this paper, we present an extended version of the system that has been presented in a previous work which includes additional fuzzy clustering algorithms and provides several indexes to evaluate the results of clustering.

The rest of the paper is organized as follows. Section 2 describes the working scheme of the proposed system. In sections 3, 4 and 5, we detail the functions of the three modules involved in REXWERE, as well as the user category extraction, dataset creation and recommendation model extraction. Finally, section 6 concludes the paper by providing an application example and giving the obtained results.

2 Working Scheme of the System

The starting point for the rule extraction process in REXWERE is represented by the behavior data derived by LODAP [2], a tool which we implemented to preprocess log files. Behavior data express models of the navigational behavior which the users exhibit during their visits to the Web site. To create such models, for each user, LODAP evaluates the interest degree for each visited page by considering two factors: the access frequency and the visit time of the user to each page. Finally, LODAP maps these values into a $n \times m$ behavior matrix $\mathbf{B} = [b_{ij}]$ where n is the number of users, m is the number of pages and each component b_{ij} represents the interest degree of the i-th user for the j-th page. In this way, the navigational behavior of the i-th user is modeled by the behavior vector $\mathbf{b}_i$, i.e. the i-th row of the behavior matrix.

Starting from the behavior data, REXWERE executes two main activities: identification of user categories and extraction of the recommendation model. The tool organizes these activities into three modules:

- The User Category Extraction (UCE) module that extracts categories of users sharing common interests by clustering the available behavior data;
- The Dataset Creation (DC) module that creates the training set and the test set needed for the learning process of the fuzzy recommendation rules;
- The Recommendation Model Extraction (RME) module that derives a fuzzy rule base representing the recommendation model by means of the learning of a neuro-fuzzy network.

The recommendation rules extracted by means of REXWERE can be used for the online suggestion of interesting links to the users of a Web site. REXWERE has been implemented in the Matlab environment (ver. 7) using GUIDE (Graphical

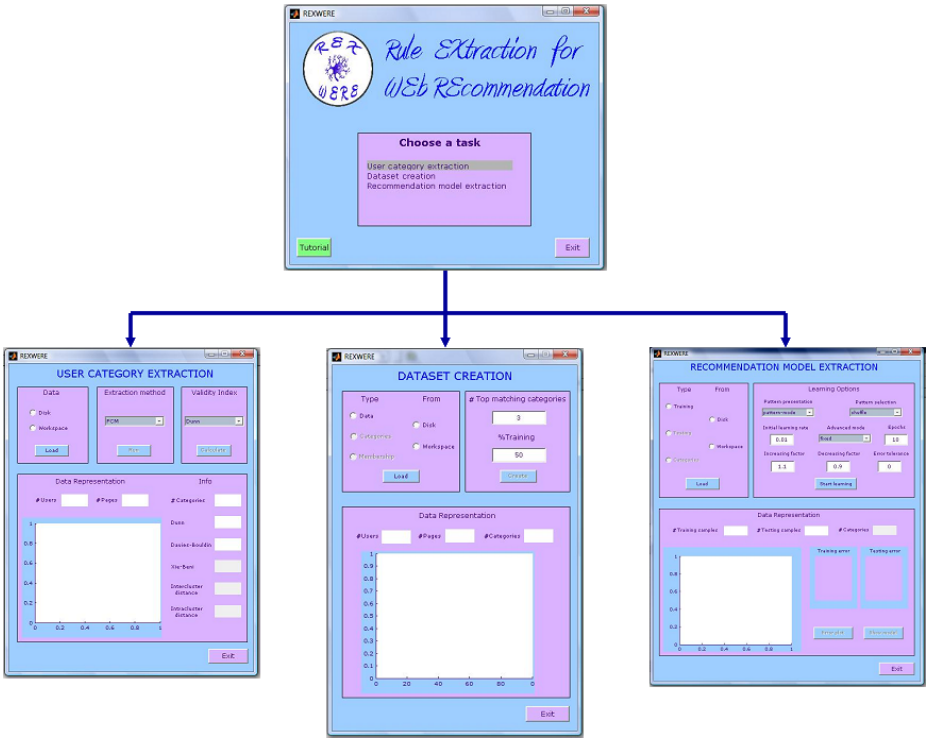


Fig. 1. The panels of REXWERE

User Interface Development Environment) for the development of the graphical interfaces. It provides a wizard-based user interface that supports the overall knowledge discovery process through a sequence of panels. Each panel is associated to a procedural module (as can be seen in fig. 1) and it consists in a window that offers, through a graphical user interface, the basic functionalities concerning the related module of the tool (such as data loading, data saving, parameter configuration, etc.). Moreover, the tool has been endowed with a tutorial that provides information about the use of the tool. The graphical interface of the tool includes a start-up window (see fig. 1) that lists all the modules involved in its working scheme. On the architectural level, all the information coming from each panel are collected and properly organized into a shared workspace. This enables the exchange of the information between the different tool panels and their use in separate experimental sessions involving each single phase of the entire knowledge discovery process. In the following, we detail the functions performed by each module of REXWERE.

3 The User Category Extraction Module

The extraction of user categories is performed by clustering the available behavior data with the aim of grouping together users with similar navigational

behavior. Due to their capacity to examine large quantity of data in a fairly reasonable amount of time, clustering algorithms are widely used to identify user categories. In particular, fuzzy clustering techniques reveal to be suited to this aim due to their ability to derive overlapping clusters (user categories). Hence, a user may belong to more than one category with a certain membership degree. To extract user categories, three different fuzzy clustering algorithms have been implemented:

- The well-known Fuzzy C-Means (FCM) algorithm [1];
- The Competitive Agglomeration Relational Data (CARD) algorithm [8];
- The CARD+ algorithm that is a modified version of CARD [3].

These algorithms differ in some features. While FCM belongs to the category of clustering algorithms working on object data expressed in the form of feature vectors, CARD and CARD+ fall in the category of fuzzy clustering algorithms based on the relational approach. This means that both these algorithms work on a relation matrix, typically containing the dissimilarity values among all pairs of data (user behavior vectors in our case). A variety of measures may be used to evaluate the (dis)similarity among data. In REXWERE, two different measures are included to estimate the similarity among behavior vectors: the cosine measure that defines the similarity degree by considering only the common pages accessed by the users and a fuzzy similarity measure which we proposed in [3]. In addition, unlike the other two algorithms, CARD+ allows to automatically determine the final number of clusters starting from an initial random number. Since the actual number of user categories visiting a Web site is not known in advance, this feature is especially required in the task of user categorization.

At the end of the clustering process, all algorithms provide the following results:

- C cluster prototypes represented as vectors $\mathbf{v}_c = (v_{c1}, v_{c2}, ..., v_{cm})$ for $c = 1, ..., C$ corresponding to the extracted user categories.
- A fuzzy partition matrix $\mathbf{M} = [m_{ic}]_{i=1...n}^{c=1...C}$ where each component m_{ic} represents the membership degree of the i-th user to the c-th category.

To evaluate the quality of the obtained partitions, different validity indexes have been implemented: the Xie-Beni index, the Dunn index, the Davies-Bouldin's index, the average intercluster and intracluster distances [6]. For good partitions, we expect small values for the Xie-Beni index, the Dunn index and the intercluster distances and high values for the Davies-Bouldin's index and the intracluster distances. On the basis of these index values, the user can establish the better partition to be used for providing user categories.

The panel of the UCE module enables the user to choose between the three clustering algorithms and the validity indexes. The panel also provides a graphical representation of data and a summary of the most important information about data, as well as the results of clustering and the validity index values.

4 The Dataset Creation Module

The DC module is mainly devoted to the creation of the dataset needed for the learning of the network employed to extract recommendation rules. Specifically, the dataset contains examples of associations between user behaviors and relevance of Web site pages to be suggested. These associations are derived by combining information about the available behavior data and the extracted user categories. More precisely, the dataset is composed of a set of n input-output samples expressed in the following form:

$$\mathbf{T} = \langle (\mathbf{b}_i, \mathbf{rd}_i) \rangle_{i=1,\ldots,n} \tag{1}$$

where the input vector $\mathbf{b}_i$ represents the i-th user behavior vector and the output vector $\mathbf{rd}_i$ expresses the amount of page recommendation for the i-th user. To compute the values in $\mathbf{rd}_i$, information embedded in the user categories are exploited. Precisely, for each vector $\mathbf{b}_i$, its membership to the user categories expressed by membership values $\{m_{ic}\}_{c=1,\ldots,C}$ in the partition matrix $\mathbf{M}$ are considered. Then, the l top matching user categories $c_1, \ldots, c_l \in \{1, \ldots, C\}$ are identified as those with the highest membership values. The number l may be set by the user. The values in the output vector $\mathbf{rd}_i = (rd_{i1}, rd_{i2}, \ldots, rd_{im})$, $(i = 1, \ldots, n)$ are hence calculated as:

$$rd_{ij} = m_{ic_1}\mathbf{v}_{jc_1} + \ldots + m_{ic_l}\mathbf{v}_{jc_l} \quad j = 1, \ldots, m \tag{2}$$

Starting from the constructed dataset, the user can create a training set and a test set by specifying a percentage of the total number n of samples as size of the training set (as can be seen in fig. 1).

5 The Recommendation Model Extraction Module

The RME module extracts the recommendation model by the learning of a specific neuro-fuzzy network. Such model represents the knowledge base expressed as a set of fuzzy rules which may be used for the online suggestion of interesting links. Each recommendation rule expresses a fuzzy relation between a behavior vector $\mathbf{b}_i = (b_{i1}, b_{i2}, ..., b_{im})$ and relevance of pages in the following form:

IF b_1 is A_{1k} AND ... AND b_m is A_{mk}
THEN relevance of $Page_1$ is r_{1k} AND ... AND relevance of $Page_m$ is r_{mk}

for $k = 1, \ldots, K$ where K is the number of fuzzy rules, A_{jk}, $j = 1, \ldots, m$ are fuzzy sets with Gaussian membership functions defined over the input variables b_j and r_{jk}, $j = 1, \ldots, m$ are fuzzy singletons expressing the amount of recommendation (relevance degree) of the j-th page.

The main advantage of using a fuzzy knowledge base for recommendation is the readability of the extracted knowledge. Actually, a fuzzy rule for recommendation can assume the following linguistic form:

IF (the interest degree for $Page_1$ is LOW) AND ...
(the interest degree for $Page_m$ is HIGH)
THEN (recommend $Page_1$ with relevance 0.3) AND ...
(recommend $Page_m$ with relevance 0.8)

Starting from the training set, the network can enter the learning phase to extract the knowledge embedded into the available data and represent it as a collection of fuzzy rules.

Firstly, the structure and the parameters of the network are initialized deriving an initial fuzzy rule base. In particular, the number of fuzzy rules (and the number of fuzzy sets used to partition data) together with the parameters that define the premise and the consequence of each rule are established. More precisely, a fuzzy rule is derived from each user category. The premise parameters of each rule depend on the center (cluster prototype) and the spread of the corresponding cluster. Hence, the centers of Gaussian membership functions coincide with the centers of clusters and their widths are simply determined by using a first-nearest-neighbor heuristic. The consequent values of each rule are calculated by weighting each of the data in the output domain by the degree of activation of the premise part of such a rule. Successively, the neural network enters in a learning phase by a back-propagation algorithm to optimally adjust the premise and the consequent parameters of the derived initial fuzzy rule base. Once the learning process has been completed, a set of fuzzy rules is derived representing the recommendation model that can be used in the online module to suggest interesting links.

6 Results and Conclusions

As an application example, REXWERE was applied to extract recommendation rules starting from log files of the Italian Web site of the Japanese movie Dragon Ball. In particular, we considered a 200×42 behavior matrix derived by LODAP from the log files of the considered site.

Firstly, the UCE module was run on the behavior matrix to derive user categories. All the implemented clustering algorithms were applied. Different runs of the FCM algorithm were carried out by setting, in each trial, a different value for the initial number of clusters (C=5,6,7,8,9,10). CARD and CARD+ were executed by using both the cosine measure and the fuzzy similarity measure. Several trials were performed by setting a different initial number of clusters $C_{max} = (5, 10, 15)$. At the end of each run, we calculated the values of the validity indexes. Among the validity indexes included in REXWERE, to establish the best partition, we considered the values obtained for the Dunn index and the Davies-Bouldin's index shown in fig. 2. In this figure, for CARD and CARD+, in correspondence of each trial, the final number of obtained clusters is indicated. By analyzing the obtained values, we observed that a good partition categorizes the available data into 5 final clusters. CARD with the use of both the similarity measures was not very stable by providing different final number of clusters in each trial. Moreover, by comparing the values of the validity indexes obtained

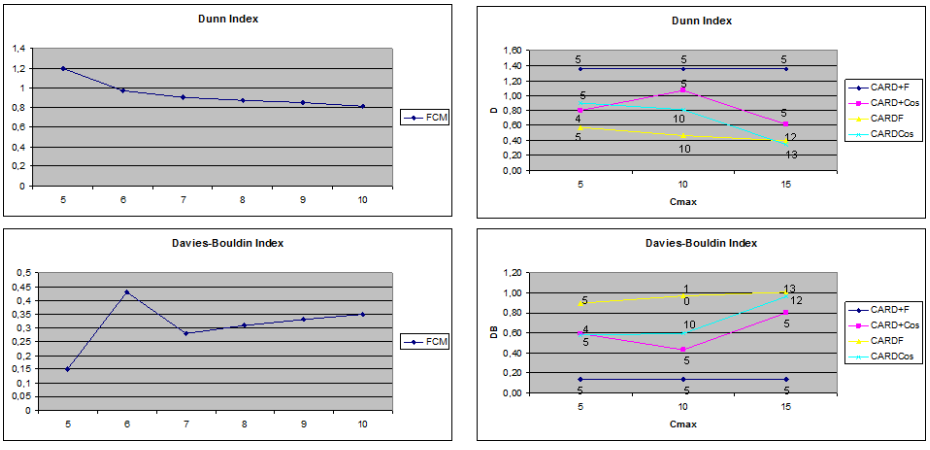


Fig. 2. The values obtained for the Dunn index and the Davies-Bouldin's index

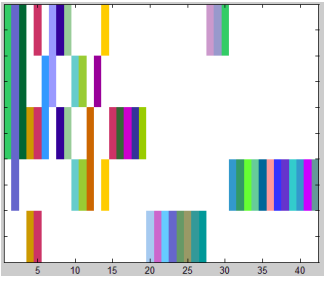


Fig. 3. Graphical representation of the identified user categories

by FCM (in correspondence of C=5) and CARD+ with both the similarity measures the best values correspond to the partition derived by CARD+ equipped with the fuzzy similarity measure. Fig. 3 shows a graphical representation of clusters identified by CARD+ with the fuzzy similarity measure. Here, each row represents a user category and each column represents a page. Different gray levels are used for different pages; for pages which are not visited the white color is used.

Next, the DC module was run. A dataset of 200 input-output samples was created, where each sample included 84 components (42 corresponding to the pages of each behavior vector and the remaining 42 to the relevance degrees). We set the size of the training set to the 70% of the dataset and the number of top matching categories to 3.

Finally, the RME module was applied to learn the fuzzy rule base. A neuro-fuzzy network with 42 inputs (corresponding to the components of the behavior vector), 42 outputs (corresponding to the relevance values of the Web pages) and 5 rule nodes was trained for 1000 epochs. In order to evaluate the quality of the recommendation model, a 10-fold cross-validation procedure was performed.

Among the 10 created models, we choose as final recommendation model the model having the lowest error on the test set (the mean error on the test set was equal to 0.18).

The derived recommendation model is represented by a fuzzy rule base composed of 5 rules. For each input variable of each rule, only two fuzzy sets are identified. In this way, rules can be expressed in a linguistic fashion making the recommendation model more comprehensible.

This model can be exploited to determine pages to be recommended to users visiting the considered site. How to provide recommendations to the users is addressed in our ongoing works.

References

1. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York (1981)
2. Castellano, G., Fanelli, A.M., Mencar, C., Torsello, M.A.: Log data preprocessing for mining Web browsing patterns. In: Proc. of the 8th Asian Pacific Industrial Engineering And Management Systems Conference (APIEMS 2007), Kaohsiung, Taiwan (2007)
3. Castellano, G., Fanelli, A.M., Mencar, C., Torsello, M.A.: Similarity-based fuzzy clustering for user profiling. In: Proc. of the IEEE/WIC/ACM International Conferences 2007, Silicon Valley, CA, USA, pp. 75–78 (2007)
4. Castellano, G., Fanelli, A.M., Torsello, M.A.: REXWERE: A tool for Rule EXtraction in WEb Recommendation. In: Proc. of the IEEE annual meeting of North American Fuzzy Information Processing society (NAFIPS 2007), San Diego, CA, pp. 129–134 (2007)
5. Facca, F.M., Lanzi, P.L.: Mining interesting knowledge from weblogs: a survey. Data and Knowledge Engineering 53, 225–241 (2005)
6. Maulik, U., Bandyopahyay, S.: Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(12), 1650–1654 (2002)
7. Nasraoui, O.: World Wide Web Personalization. In: Wang, J. (ed), Encyclopedia of Data Mining and Data Warehousing, Idea Group (2005)
8. Nasraoui, O., Frigui, H., Joshi, A., Krishnapuram, R.: Mining Web access log using relational competitive fuzzy clustering. In: Proc. of the Eight International Fuzzy System Association World Congress (1999)
9. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.D.: Web usage mining as a tool for personalization: a survey. User Modeling and User-Adapted Interaction 13(4), 311–372 (2003)

A Type-1 Approximation of Interval Type-2 FLS*

Janusz T. Starczewski

Department of Computer Engineering, Czestochowa University of Technology, Czestochowa, Poland
janusz.starczewski@kik.pcz.pl

Abstract. The paper presents a method for the approximation of the interval type-2 fuzzy logic system (FLS) by the type-1 FLS, when the interval type-2 FLS is assumed to perform the extended minimum Cartesian product and to have singleton consequents. The approximation error is discussed in details.

1 Introduction

Many authors [1,2,3,4,5,6,7,8,9] employ interval type-2 fuzzy logic systems (IT2FLSs) for variety of application tasks. Quite often noisy training data are acknowledged as ones of the sources of uncertainty. Usually the system designers translate input uncertainties into interval antecedent type-2 membership functions. The common approach is to equip all antecedents with an equal interval of membership and to use algebraic product or minimum reasoning mechanism. To our knowledge, this approach discards the potential of type-2 fuzzy sets.

In this letter we demonstrate that under specific working conditions, which commonly occur, there exist type-1 fuzzy logic system (T1FLS) equivalent to IT2FLS. However, we examine this type-1 system and similar one in the context of an approximation of ITFLSs with no specific conditions. Accordingly, the letter presents two methods for approximation of IT2FLSs by T1FLSs.

1.1 Type-2 Fuzzy Logic System

Let us now recall some basic preliminaries about type-2 sets. Let the set of all fuzzy subsets of the unit interval $[0,1]$ be denoted by $\mathcal{F}([\,0,1])$. The type-2 fuzzy set in the real line $\mathbb{R}$, denoted by $\tilde{A}$, is a set of pairs $\{x, \mu_{\tilde{A}}(x)\}$, denoted in the fuzzy union notation $\tilde{A} = \int_{x\in\mathbb{R}} \mu_{\tilde{A}}(x)/x$, where x is an element of the fuzzy set associated with the fuzzy membership function $\mu_{\tilde{A}} : \mathbb{R} \to \mathcal{F}([0,1])$. The values of $\mu_{\tilde{A}}$ are fuzzy membership grades $\mu_{\tilde{A}}(x)$ being classical fuzzy subsets of the unit interval $[0,1]$, i.e.,

$$\mu_{\tilde{A}}(x) = \int_{u\in[0,1]} f_x(u)/u \ , \tag{1}$$

* This work was partly supported by Polish Ministry of Science and Higher Education (Habilitation Project N N516 372234 2008–2011, Special Research Project 2006–2009, Polish-Singapore Research Project 2008–2010).

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 287–294, 2009.

where $f_x\colon [0,1] \to [0,1]$. The membership function (MF) f_x is referred as a secondary MF. Although, the secondary MF may take various shapes, almost all applications employ normal and rectangular secondary MFs constituting so called interval type-2 fuzzy sets [1,4,6,9]. This study is devoted to the logical systems based on the interval type-2 fuzzy sets, called Interval Type-2 Fuzzy Logic Systems (IT2FLS).

We start from the typical type-2 FLS architecture [10] that consists of the type-2 fuzzy rule base, the type 2 fuzzifier, the inference engine modified to deal with type 2 fuzzy sets, and the defuzzifier split into the type reducer and type 1 defuzzifier.

The rule base is composed of K rules:

$$\tilde{R}^k : \text{IF } x_1 \text{ is } \tilde{A}_1 \text{ and } x_2 \text{ is } \tilde{A}_2 \text{ and } \cdots \text{ and } x_N \text{ is } \tilde{A}_N \text{ THEN } y \text{ is } \tilde{B}.$$

where $\tilde{A}_n$, is the n-th antecedent fuzzy set of type 2, $\tilde{B}$ is the consequent type 2 fuzzy set, x_n is the n-th input variable, $k = 1, \ldots, K$. Fuzzy rules are fuzzy relations usually expressed by extended t-norms. The inference engine produces the type-2 fuzzy consequence by means of the extended sup-star composition [10] of the premise and the type 2 fuzzy relation, i.e.,

$$\mu_{\tilde{B}'^k}(y) = \widetilde{\sup_{\mathbf{x}\in\mathbf{X}}} \tilde{T}\left(\mu_{\tilde{A}'}(\mathbf{x}), \mu_{\tilde{R}^k}(\mathbf{x}, y)\right) \quad . \tag{2}$$

If fuzzy premises $\tilde{A}'_n$ are singletons, i.e., $\tilde{A}'_n = (1/1)/x'_n$, the inference is based on the premises combined by an extended T-norm (Cartesian product) and on the consequents according to the following scheme:

$$\begin{aligned} h_k(y) = \mu_{\tilde{B}'^k}(y) &= \mu_{\tilde{A}' \circ \left(\tilde{A}^k \cap \tilde{B}^k\right)}(y) = \tilde{T}\left(\mu_{\tilde{A}^k}(\mathbf{x}'), \mu_{\tilde{B}^k}(y)\right) \\ &= \tilde{T}\left[\underset{n=1}{\overset{N}{\tilde{T}}} \mu_{\tilde{A}^k_n}(x'_n), \mu_{\tilde{B}^k}(y)\right] \quad . \end{aligned} \tag{3}$$

Commonly the consequents are also singletons at each point y_k. Therefore the k-th conclusion MF may be expressed as:

$$h_k(y_k) = \mu_{\tilde{B}'^k}(y_k) = \underset{n=1}{\overset{N}{\tilde{T}}} \mu_{\tilde{A}^k_n}(x'_n) \quad , \tag{4}$$

while in all other points $y \neq y_k$ the membership is certainly 0, i.e. 1/0.

In the case of interval type-2 fuzzy sets, the well known Karnik&Mendel type-reduction method is a standard transformation of type-2 fuzzy conclusions into a type-1 fuzzy set [11]. In the assumed case of singleton consequents, the discrete version of the Karnik&Mendel type-reduction, i.e. the height type-reduction, is employed. Let us denote the bounds of the type-reduced fuzzy set as $y_{\min}$ and $y_{\max}$. Throughout this study, only the overall output will be taken into account,

$$y_{T2} = \frac{y_{\min} + y_{\max}}{2} \quad . \tag{5}$$

Since throughout this study only interval memberships of type-2 FLS are considered, any type-2 fuzzy set $\tilde{A}$ is described by a fuzzy membership grade of

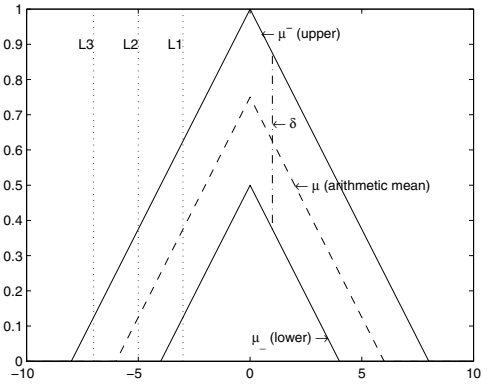


Fig. 1. Example of Type-2 Antecedent with Uniform Interval of Uncertainty

the form $\mu_{\tilde{A}}(x) = 1/[\underline{\mu}_A(x), \overline{\mu}_A(x)]$, where the lower and upper boundaries are called respectively an upper membership grade and a lower membership grade. That reduces calculations of extended t-norms to the following:

$$\tilde{T}(\mu_{\tilde{A}}, \mu_{\tilde{B}}) = 1/\left[T\left(\underline{\mu}_A, \underline{\mu}_B\right), T(\overline{\mu}_A, \overline{\mu}_B)\right] \quad . \tag{6}$$

2 Mean Type-1 Approximation of IT2FLS for Equal Uncertainties and Minimum Cartesian Product

Consider the IT2FLS with all antecedents having equal intervals of uncertainty as one shown in Fig. 1. The upper and lower memberships are almost entirely equidistant except the parts clipped by x axis. We may denote the interval of uncertainty by δ, as the distance between the upper and the lower membership grades at the center of the interval type-2 fuzzy set.

Combining all degrees of compatibility between inputs and antecedents by the minimum Cartesian product, we obtain the function of the lower firing grade as a dependence of the upper firing grade.

$$\overline{h}_k = \min_{i=1}^{n} \overline{\mu}_i^k \quad , \tag{7}$$

$$\underline{h}_k = \min_{i=1}^{n} \left(\max\left(0, \overline{\mu}_i^k - \delta\right)\right) = \max\left(0, \min_{i=1}^{n} \overline{\mu}_i^k - \delta\right) = \max\left(0, \overline{h}_k - \delta\right) \quad . \tag{8}$$

The proposed method for the type-1 approximation relies on the substitution of type-2 antecedents by arithmetic means of upper and lower boundaries $\mu_i^k = (\overline{\mu}_i^k + \underline{\mu}_i^k)/2$, i.e.,

$$\mu_i^k = \begin{cases} \overline{\mu}_i^k - \frac{\delta}{2} & \text{if } \overline{\mu}_i^k \in (\delta, 1] \\ \frac{\overline{\mu}_i^k}{2} & \text{if } \overline{\mu}_i^k \in (0, \delta) \end{cases} = \max\left(\frac{\overline{\mu}_i^k}{2}, \overline{\mu}_i^k - \frac{\delta}{2}\right) \quad . \tag{9}$$

Therefore, the k-th rule firing grade of the approximation T1FLS is as follows:

$$h_k = \min_{i=1}^{n}\left(\max(\frac{\overline{\mu}_i^k}{2}, \overline{\mu}_i^k - \frac{\delta}{2})\right) = \max\left(\min_{i=1}^{n}\frac{\overline{\mu}_i^k}{2}, \min_{i=1}^{n}\left(\overline{\mu}_i^k - \frac{\delta}{2}\right)\right)$$
$$= \max\left(\frac{1}{2}\overline{h}_k, \overline{h}_k - \frac{\delta}{2}\right) \quad . \tag{10}$$

Recall that we assumed the consequents of IT2FLS to be singletons. In the case of only one active rule, it would be trivial to demonstrate that both IT2FLS and T1FLS give the same output value. Hence, we shall analyze the cases of multiple active rules. In the most typical circumstances, the fuzzy partition of the input domain is such uniform that it guarantees the firing at most two of rules. Although the use of non-zero membership functions as Gaussian ones always brings about the firing of many rules, we may assume that only two of them are of the great importance. Consequently, the objective is to compare the overall output of the IT2FLS with the approximation T1FLS in the most frequent instances of two fired rules. Singleton inputs may be projected on type-2 antecedents in the ways depicted by lines L1, L2 and L3 in Fig. 1. In the sequel we distinguish several combinations of firing grades.

2.1 Case 1 — Equivalence between IT2FLS and T1FLS

Suppose that distance δ between the upper and the lower firing grades is constant, i.e. $\overline{h}_k - \underline{h}_k = \delta$ for the two fired rules. Evidently, the lower firing grade shall be positive, $\overline{h}_k - \delta \geq 0, k = 1, 2$. According to (5), the IT2FLS output is an average of the maximal and minimal bounds of the type-reduced set obtained by the discrete version of the Karnik&Mendel iterative type-reduction procedure. With the restriction of two fired rules and constant δ, the output of the type-2 system is calculated as

$$y_{T2} = \frac{1}{2}\left(\frac{\overline{h}_1 y_1 + \left(\overline{h}_2 - \delta\right) y_2}{\overline{h}_1 + \overline{h}_2 - \delta} + \frac{\left(\overline{h}_1 - \delta\right) y_1 + \overline{h}_2 y_2}{\overline{h}_1 - \delta + \overline{h}_2}\right)$$
$$= \frac{\left(\overline{h}_1 - \frac{\delta}{2}\right) y_1 + \left(\overline{h}_2 - \frac{\delta}{2}\right) y_2}{\left(\overline{h}_1 - \frac{\delta}{2}\right) + \left(\overline{h}_2 - \frac{\delta}{2}\right)} = y_{T1} \quad . \tag{11}$$

Surprisingly, the interval type-2 system has given the same output as the corresponding T1FLS such that we have no approximation error. We have noticed this property in [12,13]. Moreover, this case is the most common situation in fuzzy reasoning.

2.2 Case 2 — Clipping the Interval of Uncertainty for One Rule

Suppose the second rule has a clipped interval as projection L2 in Fig. 1, such that $0 < \overline{h}_2 \leq \delta, \overline{h}_1 > \delta$. It ensures that $\underline{h}_2 = 0$ and $h_2 = \overline{h}_2/2$. Therefore, the type-2 output is

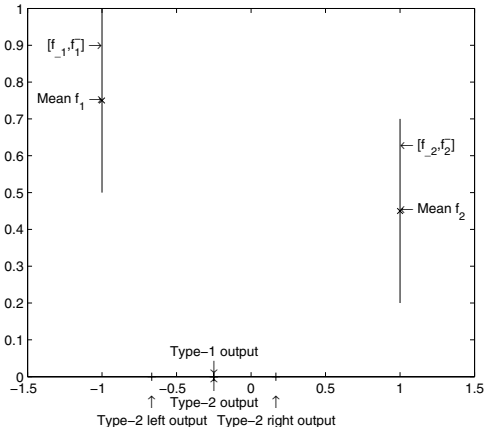


Fig. 2. Outputs of Type-2 and Arithmetic Mean Approximated Type-1 FLSs for Case of Equivalence of Systems

$$y_{T2} = \frac{1}{2}\left(\frac{\overline{h}_1 y_1 + 0 y_2}{\overline{h}_1 + 0} + \frac{\left(\overline{h}_1 - \delta\right) y_1 + \overline{h}_2 y_2}{\overline{h}_1 - \delta + \overline{h}_2}\right)$$
$$= \frac{1}{2}\left(\frac{2\overline{h}_1 + \overline{h}_2 - 2\delta}{\overline{h}_1 + \overline{h}_2 - \delta} y_1 + \frac{\overline{h}_2}{\overline{h}_1 + \overline{h}_2 - \delta} y_2\right) \tag{12}$$

and the corresponding type-1 system output is

$$y_{T1} = \frac{\left(\overline{h}_1 - \frac{\delta}{2}\right) y_1 + \frac{\overline{h}_2}{2} y_2}{\overline{h}_1 - \frac{\delta}{2} + \frac{\overline{h}_2}{2}} \quad . \tag{13}$$

The difference between the outputs may be converted as follows:

$$e = \frac{1}{2}\left(\frac{2\overline{h}_1 + \overline{h}_2 - 2\delta}{\overline{h}_1 + \overline{h}_2 - \delta} y_1 + \frac{\overline{h}_2}{\overline{h}_1 + \overline{h}_2 - \delta} y_2\right) - \frac{\left(\overline{h}_1 - \frac{\delta}{2}\right) y_1 + \frac{\overline{h}_2}{2} y_2}{\overline{h}_1 - \frac{\delta}{2} + \frac{\overline{h}_2}{2}}$$
$$= \frac{\left(\delta - \overline{h}_2\right)\overline{h}_2}{2\left(\overline{h}_1 + \overline{h}_2 - \delta\right)\left(2\overline{h}_1 + \overline{h}_2 - \delta\right)}\left(y_2 - y_1\right) \quad . \tag{14}$$

We define the approximation error that its formula does not depend on y dimension.

$$\varepsilon = \frac{e}{y_2 - y_1} = \frac{\left(\delta - \overline{h}_2\right)\overline{h}_2}{2\left(\overline{h}_1 + \overline{h}_2 - \delta\right)\left(2\overline{h}_1 + \overline{h}_2 - \delta\right)}$$
$$= \frac{\left(\delta - h_2\right) h_2}{2\left(h_1 + h_2 - \delta\right)\left(2h_1 + h_2 - \delta\right)} \quad . \tag{15}$$

Since $\frac{\partial \varepsilon}{\partial \overline{h}_2} = \frac{-\overline{h}_1 - \delta}{2\left(\overline{h}_1 + \overline{h}_2\right)^2} < 0$, the error achieves its maximum when δ is maximal and both $\overline{h}_1$ and $\overline{h}_2$ are minimal. It means that $\delta \to 1, \overline{h}_2 \to \frac{\delta}{2}, \overline{h}_1 \to \delta$ and the

error value is 0.5. Choosing lower value of interval δ diminishes the approximation error.

2.3 Case 3 — Clipping the Interval of Uncertainty for Two Rules

Suppose that the firing intervals of both rules are clipped, i.e., $0 < \overline{h}_2 \leq \delta, 0 < \overline{h}_1 \leq \delta$. Therefore, $\underline{h}_2 = 0, \underline{h}_1 = 0$ and $h_2 = \overline{h}_2/2, h_1 = \overline{h}_1/2$. The type-2 system output becomes

$$y_{T2} = \frac{1}{2}\left(\frac{\overline{h}_1 y_1}{\overline{h}_1} + \frac{\overline{h}_2 y_2}{\overline{h}_2}\right) = \frac{1}{2}\left(y_1 + y_2\right) \quad . \tag{16}$$

It is interesting that the firing grades do not have any effect on the output of the IT2FLS, while the corresponding T1FLS has the following output value:

$$y_{T1} = \frac{\frac{\overline{h}_1}{2} y_1 + \frac{\overline{h}_2}{2} y_2}{\frac{\overline{h}_1}{2} + \frac{\overline{h}_2}{2}} = \frac{\overline{h}_1 y_1 + \overline{h}_2 y_2}{\overline{h}_1 + \overline{h}_2} \quad . \tag{17}$$

The difference between outputs is

$$\begin{aligned} e &= \frac{1}{2}\left(y_1 + y_2\right) - \frac{\overline{h}_1 y_1 + \overline{h}_2 y_2}{\overline{h}_1 + \overline{h}_2} = \frac{\left(\overline{h}_2 - \overline{h}_1\right) y_1 + \left(\overline{h}_1 - \overline{h}_2\right) y_2}{2\left(\overline{h}_1 + \overline{h}_2\right)} \\ &= \frac{1}{2}\frac{\overline{h}_1 - \overline{h}_2}{\overline{h}_1 + \overline{h}_2}\left(y_2 - y_1\right) \quad , \end{aligned} \tag{18}$$

and the approximation error is

$$\varepsilon = \frac{1}{2}\frac{\overline{h}_1 - \overline{h}_2}{\overline{h}_1 + \overline{h}_2} \quad . \tag{19}$$

Obviously, the theoretical maximal approximation error occurs for the maximal difference between $\overline{h}_1$ and $\overline{h}_2$. Consequently, it should be that $\overline{h}_1 = \delta, \overline{h}_2 \to 0$. As a result, the error $\varepsilon \to 0.5$.

2.4 Approximation Error

In Fig. 3, the approximation error surfaces are demonstrated. For small values δ, there is no difference between type-2 and type-1 fuzzy systems in most configurations of the firing degrees. Hence, for thin and uniform intervals of uncertainty, the profits of the type-2 fuzzy logic are not great.

3 Simulation

In order to demonstrate the accuracy of our type-1 approximation, we used IT2FLSs to the "Glass" classification problem, from the UCI Repository of machine learning databases. The dataset consisting 214 patterns belonging to 6

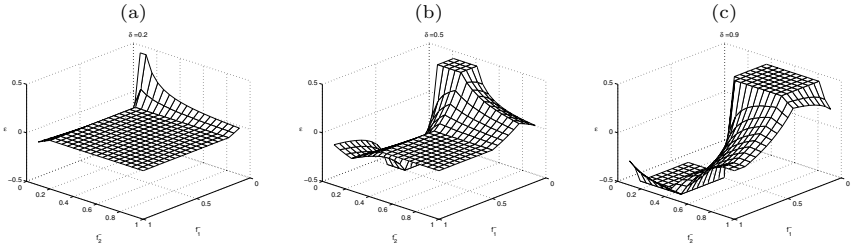


Fig. 3. Error of Arithmetic Mean Approximation for Equal Uncertainties; (a) $\delta = 0.2$, (b) $\delta = 0.5$, (c) $\delta = 0.9$

classes was divided into two equal parts, the one used for training and the second used for testing. We implemented our approximation method with respect to the One-against-All fuzzy classification system, i.e., each class was trained against all other classes in independent IT2FLS. Hence, the number of these independent subsystems was equal to the number of classes. The maximal output of the subsystems indicated the implied class.

Each interval type-2 fuzzy logic subsystem had 3 rules with the triangular upper and lower antecedent MFs (Fig. 1). The rules were set up by the standard FCM algorithm ($m = 2$) and the least squares fitting. We assumed the interval of uncertainty to be 0.2. Due to the uniform uncertainty and the minimum Cartesian product, we could apply the type-1 approximation of the system. In both systems, IT2FLS and the approximation T1FLS, the number of correct classifications on the testing set was equal to 57. Although the classification accuracy could be improved with the use of gradient learning techniques, it is only important that in the case of uniform uncertainty the performance of the typical T1FLS is identical as the performance of the expensive IT2FLS.

4 Conclusion

In many practical situations developers do not know whether the interval type-2 fuzzy logic approach is more appropriate than the type-1 fuzzy logic. Quite often their simulation results are not so successful as they wish to be. For this reason, we proposed the method of approximation of the minimum-based IT2FLS by the classical T1FLS. The method is accurate for uniform membership uncertainties, if they are not too wide. This restricted the type of problems the interval type-2 fuzzy logic can be addressed to the problems requiring not uniform or wide membership uncertainty. The approximation method can be easily extended to the systems based on triangular type-2 fuzzy sets [14] (for its mathematical background see [15]). We hope the proposed method will be very useful for validations of the type-2 FL applications by the comparison with the approximation T1FLSs.

References

1. Castillo, O., Aguilar, L.T., Cazarez-Castro, N.R., Boucherit, M.S.: Application of type-2 fuzzy logic controller to an induction motor drive with seven-level diode-clamped inverter and controlled infeed. Electrical Engineering 90(5), 347–359 (2008)
2. Castillo, O., Melin, P.: Intelligent systems with interval type-2 fuzzy logic. International Journal of Innovative Computing, Information and Control 4(4), 771–783 (2008)
3. Castillo, O., Melin, P.: Adaptive noise cancellation using type-2 fuzzy logic and neural networks. In: Proc. IEEE-FUZZ 2004, Budapest (July 2004)
4. Hagras, H.A.: A hierarchical type-2 fuzzy logic control architecture for autonomous robots. IEEE Transactions on Fuzzy Systems 12(4), 524–539 (2004)
5. Hagras, H.: Type-2 fuzzy logic controllers: A way forward for fuzzy systems in real world environments. In: Zurada, J.M., Yen, G.G., Wang, J. (eds.) WCCI 2008. LNCS, vol. 5050, pp. 181–200. Springer, Heidelberg (2008)
6. Liang, Q., Mendel, J.M.: Interval type-2 fuzzy logic systems: Theory and design. IEEE Transactions on Fuzzy Systems 8, 535–550 (2000)
7. Mendez, G.M.: Interval type-2 anfis. In: Advances in Soft Computing – Innovations in Hybrid Intelligent Systems, pp. 64–71 (2007)
8. Torres, P., Sáez, D.: Type-2 Fuzzy Logic Identification Applied to the Modeling of a Robot Hand. In: Proc. FUZZ-IEEE 2008, Hong Kong (June 2008)
9. Uncu, O., Turksen, I.B.: Discrete interval type 2 fuzzy system models using uncertainty in learning parameters. IEEE Transactions on Fuzzy Systems 15(1), 90–106 (2007)
10. Karnik, N.N., Mendel, J.M., Liang, Q.: Type-2 fuzzy logic systems. IEEE Transactions on Fuzzy Systems 7(6), 643–658 (1999)
11. Karnik, N.N., Mendel, J.M.: Centroid of a type-2 fuzzy set. Information Sciences 132, 195–220 (2001)
12. Starczewski, J., Rutkowski, L.: Neuro-fuzzy systems of type 2. In: Proc. 1st Int'l Conf. on Fuzzy Systems and Knowledge Discovery, Singapore, vol. 2, pp. 458–462 (2002)
13. Starczewski, J.: What differs type-2 FLS from type-1 FLS? In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) ICAISC 2004. LNCS, vol. 3070, pp. 381–387. Springer, Heidelberg (2004)
14. Starczewski, J.T.: Efficient triangular type-2 fuzzy logic systems. International Journal of Approximate Reasoning (accepted for publication)
15. Starczewski, J.T.: Extended triangular norms. Information Sciences 179, 742–757 (2009)

Control of a Non-isothermal CSTR by Type-2 Fuzzy Logic Controllers

Mosé Galluzzo and Bartolomeo Cosenza

Dipartimento di Ingegneria Chimica dei Processi e dei Materiali,
Università di Palermo,
Viale delle Scienze, 90128, Palermo, Italy
{galluzzo,b.cosenza}@dicpm.unipa.it

Abstract. The paper describes the application of a type-2 fuzzy logic controller (FLC) to a non-isothermal continuous stirred tank reactor (CSTR) characterized by the presence of saddle node and Hopf bifurcations. Its performance is compared with a type-1 fuzzy logic controller performance. A full analysis of the uncontrolled CSTR dynamic was carried out and used for the feedback-feedforward fuzzy controllers development. Simulation results confirm the effectiveness and the robustness of the type-2 FLCs which outperform their type-1 counterparts, particularly when uncertainties are present in the system.

Keywords: Type-2 fuzzy logic controller; Non-isothermal CSTR; Bifurcation; Non-linear system.

1 Introduction

The dynamics of many non linear systems can be strongly dependent on one or more parameters since their operative condition remains stable only if the values of these parameters remain in a limited range [1]. If the system parameters go out of this range then the equilibrium point becomes unstable. For this reason, nonlinear controllers like fuzzy logic controllers are used to control such systems because they are more robust then traditional controllers and can handle the changes in the system parameters. In the last years fuzzy logic controllers (FLCs) that use type-2 fuzzy sets [2,3] have been developed. It has been shown that often their performance is higher than type-1 FLCs [4,5,6] and also of traditional Proportional-Integrative-Derivative (PID) controllers [7]. With type-2 fuzzy logic it is in fact possible to handle the uncertainties [8,9] present in the system and in the input data to the controller. A few type-2 FLCs have been developed for applications in the field of process control [4,6,7]. In this paper the use of type-2 FLC for the control of a continuous stirred tank reactor (CSTR) that is characterized by the presence of saddle node and Hopf bifurcations is analyzed. A feedback and a feedback-feedforward control strategy are adopted and simulated in the presence of uncertain parameters and measurement noise. The simulation results are compared with those obtained using type-1 FLCs in the same contest.

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 295–302, 2009.

2 Non-isothermal CSTR

The case considered in this paper is the non-isothermal CSTR reported in [10], where an irreversible reaction A → B occurs.

2.1 Model Equations

The equations of the mathematical model are obtained by a component mass balance, an energy balance in the reactor and a energy balance in the jacket. The dynamics related to the jacket temperature can be considered much faster than that related to the reactor temperature thus the jacket time constant is negligible. The water cooling jacket is assumed to be perfectly mixed, and the mass of the metal walls are considered negligible, so that the thermal inertia of the metal is not considered. Therefore a simplified model with two equations can be derived and to generalize the mathematical model of the reactor, the equations can be expressed in a dimensionless way (1), (2), that represent the state-space model in the dimensionless reactor concentration x_2 and temperature x_3:

$$\frac{dx_2}{dt} = \frac{x_{60}}{x_1}(x_{20} - x_2) - c_0 x_2 e^{-\frac{1}{x_3}} \quad . \tag{1}$$

$$\frac{dx_3}{dt} = \frac{x_{60}}{x_1}(x_{30} - x_3) + c_1 x_2 e^{-\frac{1}{x_3}} - \frac{c_2 c_3 x_5 (x_3 - x_{40})}{x_1 (c_3 x_5 + c_4)} \quad . \tag{2}$$

A more detailed description of the model and the dimensionless variables can be found in [10].

2.2 Control Strategy

The considered exothermic CSTR without control can have multiple steady states and bifurcation points. The bifurcation parameter of this case is the coolant flow rate (dimensionless parameter x_5). Figure 1 explains the effect on the system of a disturbance in the dimensionless inlet temperature x_{30}. The line with pronounced stroke is the equilibrium curve corresponding to the initial stable condition of the reactor obtained with constant reactant flow rate $x_{60} = 1.5$ and constant input temperature $x_{30} = 0.0373$. The horizontal line indicates instead the desired dimensionless temperature of the reactor (set-point value = 0.039). If x_{30} assumes the value 0.038, the initial curve moves to the right and the CSTR remains in the stable region. A decrease of x_{30} to 0.0353 moves instead the curve to the left and the CSTR into the unstable region (the horizontal line is now included in the dash line of the new curve obtained at $x_{60} = 1.5$ and $x_{30} = 0.0353$). The behavior of the system in this zone is characterized by oscillations with a large amplitude. A possible solution to control the system is to increase the proportional gain of a feedback controller, until the oscillations disappear; but experience has however shown that this solution is not acceptable in real systems because it would result in noise amplification and instability. An alternative solution could be to manipulate a second process

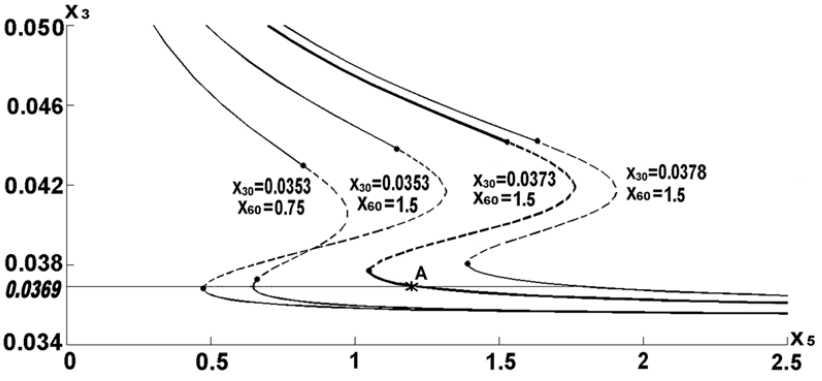


Fig. 1. Bifurcation plots x_3 *vs* x_5 for different values of parameters x_{30} and x_{60}

variable such as the inlet flow rate of the CSTR (x_{60}) in a feedforward control loop, measuring the disturbance variable x_{30}. By manipulating x_{60} it is in fact possible to shift the unstable region of the equilibrium curve over the set-point value, making sure to keep the system in a stable region, without oscillations.

3 Type-2 Fuzzy Logic

3.1 Type-2 Fuzzy Sets

A type-2 fuzzy set $\tilde{A}$ is characterized by a type-2 membership $\mu_{\tilde{A}}(x,u)$, where $x \in X$ and $u \in J_x \subseteq [0,1]$, and defined as:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \frac{\mu_{\tilde{A}}(x,u)}{x,u} \quad . \tag{3}$$

$$0 \leq \mu_{\tilde{A}}(x,u) \leq 1 \quad . \tag{4}$$

In (3) and (4) $\mu_{\tilde{A}}(x,u)$ is the *secondary grade* while the *primary membership* of x is the domain of the secondary membership function. For computational reasons in this paper only a particular case of type-2 fuzzy sets is used: the interval type-2 fuzzy sets (IT2FS) [3,11]. An interval type-2 fuzzy set is defined as:

$$\tilde{A}_I = \int_{x \in X} \frac{\int_{u \in J_x \subseteq [0,1]} (\frac{1}{u})}{x} \quad . \tag{5}$$

In a fuzzy system there can be different causes of uncertainty: from the meaning of words used in defining rules to the uncertainty (noise) present in measurements. The main characteristic of type-2 fuzzy sets is that they take into account the uncertainty of a system through a bounded region (in the primary membership) that is called the Footprint of Uncertainty (FOU) [9]. For a detailed general discussion about uncertainty see also [8]. The FOU is used to

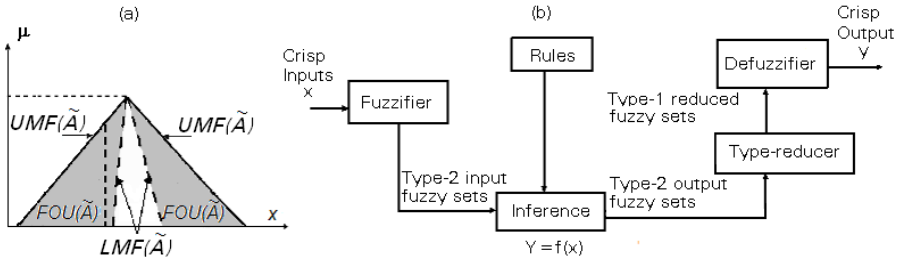


Fig. 2. (a) *FOU* (*shaded*), LMF (*dashed*), UMF (*solid*) for IT2FS ; (b) Type-2 FLS

model the shape and the position of all the uncertainties present in a system. For an IT2FS the shaded region in Fig. 2(a) denotes interval sets for the secondary membership functions. The *FOU* can be described in terms of upper (UMF) and lower (LMF) membership functions [11]. For a specific value of the definition interval there is no longer a single value for the membership function but the membership function takes on values wherever the vertical line intersects the blur between the two type-1 membership functions. Type-2 fuzzy logic systems are very useful in all circumstances in which measurements are characterized by uncertainty and when it is difficult to determine an exact membership function.

3.2 Type-2 Fuzzy Logic Systems

As for a type-1 fuzzy logic system (FLS), also a type-2 FLS contains four components: rules, fuzzifier, inference-engine and output-processor. The main difference between type-2 and type-1 FLS is the output-processor, in fact for a type-1 FLS it is just a defuzzifier, while, for a type-2 FLS it contains two components: the first component, the type-reducer, maps a type-2 fuzzy set into a type-1 fuzzy set, while the second component is just a normal defuzzifier that transforms a fuzzy output in a crisp output. In Fig. 2(b) a general type-2 FLS is depicted. The rules for a type-2 FLS represent a type-2 relation between the input space and the output space. Their structure is the same of type-1 rules, the only difference consists in the membership functions nature i.e.

$$R^1 : IF\ x_1\ is\ \tilde{F}_1^l\ and\ ...\ and\ x_p\ is\ \tilde{F}_p^l,$$
$$THEN \quad y\ is\ \tilde{G}^l \qquad\qquad l = 1, ..., M\ . \tag{6}$$

Equation (6) is the *lth* rule for a type-2 FLS with p inputs and *1* output. As we have already outlined the output of the inference engine in a type-2 FLS is a type-2 fuzzy set and it must be type-reduced before it can be defuzzified by the defuzzifier. One of the most used type-reduction methods is the *center of sets* type reducer, which can be expressed as:

$$Y_{cos}(x) = [y_l, y_r] = \int_{y^1 \in [y_l^1, y_r^1]} \cdots \int_{y^M \in [y_l^M, y_r^M]} \int_{f^1 \in [\underline{f}^1, \overline{f}^1]} \cdots \int_{f^M \in [\underline{f}^M, \overline{f}^M]} \frac{\sum_{i=1}^{M} f^i}{\sum_{i=1}^{M} f^i y^i} \quad . \tag{7}$$

In (7) $Y_{cos}(x)$ is an interval set while y_l and y_r are its end-points, $[\underline{f}^i, \overline{f}^i]$ and $[y_l^i, y_r^i]$ are respectively the interval firing level of the ith rule and the centroid of the type-2 interval consequent set. With the Karnik-Mendel iterative method [2,12] it is possible to compute the equation (7). $Y_{cos}(x)$ as interval set must be defuzzified; this operation can be carried out very simply because of the use of interval type-2 fuzzy sets: in this case the defuzzified output is the average of y_i and y_r.

The control of the CSTR makes use of two FLCs with Sugeno inference, the first in a feedback control loop and the second in a feedforward control loop. All feedback FLCs use two input variables, *error* (e) and *integral error* ($inte$), and one dimensionless output variable (x_5) with a TISO (two inputs - single output) structure. The structure of the feedforward FLC is instead SISO (single input - single output) and it is very simple consisting of only two membership functions and two rules. Both the type-1 and the type-2 feedback FLCs have seven Gaussian membership functions. They were chosen minimizing the Integral of Square Error (ISE) for set point changes. For the variables of all fuzzy feedback controllers, seven Gaussian membership functions were chosen. The rule base used in the feedback FLCs was designed by simulation runs. The structure of type-1 FLC is the same of type-2 FLC and the only difference consists of Gaussian membership functions amplitude. Each type-1 fuzzy Gaussian membership function has in fact an amplitude value that is the average of type-2 fuzzy internal and external Gaussian membership functions amplitude values.

4 Results and Discussion

The objective of the CSTR control is to keep the reactor in the chosen initial equilibrium point (point A in Fig. 1) even in the presence of large disturbances and parameter changes. The use of a simple feedback control loop using the coolant flow rate as manipulative variable allows to control the reactor temperature at the set point value ($x_3 = 0.0369$) in some cases, for instance for a step change in the inlet temperature x_{30} from 0.0373 to 0.0378. But when a step change is introduced in x_{30} in the opposite direction, from 0.0373 to 0.0353, the feedback controller, both with type-1 and type-2 fuzzy logic, cannot control the reactor and the temperature oscillates, as it can be seen in Fig. 3. The response of the type-2 FLC slightly outperforms that of the type-1 FLC when feedforward action is added (see Fig. 4). With both controllers the response starts to oscillate around the set point value after the introduction of the step change at $\tau = 20$ with decreasing amplitude until the set point value is reached. The oscillations of the type-2 FLC are slightly smaller than those of the type-1 FLC and reach the set point in a reduced time. The superiority of type-2 FLCs over type-1 FLCs is more evident when uncertainty is present in the system. Figure 5 illustrates

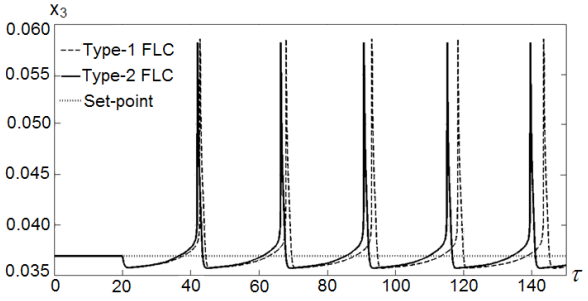


Fig. 3. Response of temperature x_3 to a step change in the disturbance (input temperature) from $x_{30} = 0.037$ to $x_{30} = 0.0353$ at $\tau = 20$ with a constant set-point (0.0369) with type-1 and type-2 fuzzy feedback control

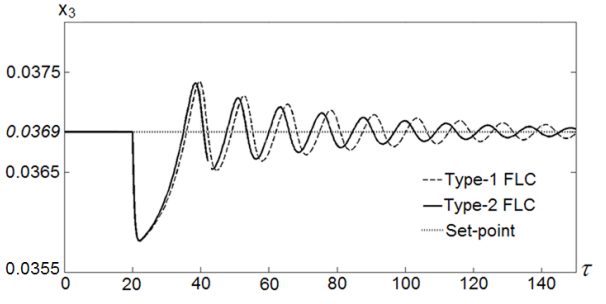


Fig. 4. Response of temperature x_3 to a step change in the disturbance (input temperature) from x_{30}=0.037 to $x_{30} = 0.0353$ at $\tau = 20$ with a constant set-point (0.0369) with type-1 and type-2 fuzzy feedback-feedforward control

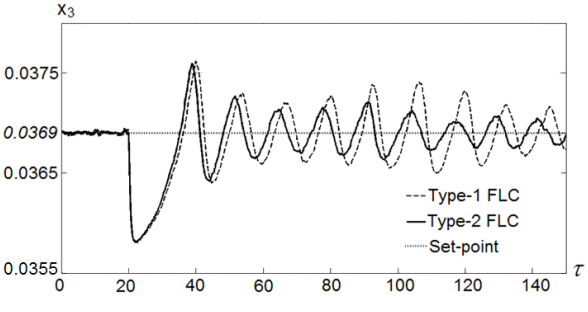


Fig. 5. Response of temperature x_3 to a step change in the disturbance (input temperature) from $x_{30} = 0.037$ to $x_{30} = 0.0353$ at $\tau = 20$, with a constant set-point (0.0369) and random variation of some system parameters, with type-1 and type-2 fuzzy feedback-feedforward control

the behaviour of both controller types when uncertainties are introduced as random variations of some system parameters (c_0 and c_2), making more difficult the control of the system. It can be seen that the control configuration using

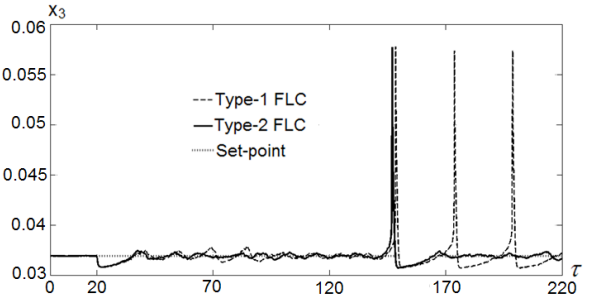


Fig. 6. Response of temperature x_3 to a step change in the disturbance (input temperature) from $x_{30} = 0.037$ to $x_{30} = 0.0353$ at $\tau = 20$, with a constant set-point (0.0369) and noise in the measurement of the disturbance x_{30} in the feedforward control loop, with type-1 and type-2 fuzzy feedback-feedforward control

type-2 FLCs is able to keep the reactor temperature under control reducing the amplitude of oscillations with time better than using type-1 FLCs, showing a greater robustness and minimizing the effects of the uncertainties present in the system. The superiority is even more evident when the uncertainty is introduced as a noise in the measurement of the disturbance x_{30}, in the feedforward control loop. In this case, see Fig. 6, the behaviour of the two fuzzy controllers is very similar until $\tau = 170$ (with a slightly better performance of the type-2) but then at $\tau = 170$ the type-2 FLC outperforms its type-1 counterpart, reducing the amplitude of the oscillations and reaching the set-point value. Type-1 FLC instead is not able to handle the uncertainties, showing a behaviour characterized by oscillations with constant amplitude.

5 Conclusions

It has been shown by simulation that the control of the considered non-isothermal CSTR with bifurcation points cannot be achieved through a simple feedback control loop in the presence of disturbances in the reactant temperature. A feedback-feedforward control scheme is able to maintain the control of the reactor temperature even in the presence of changes in some reactor parameters. If the disturbance measurement is affected by noise only using type-2 FLCs in the feedback-feedforward control scheme can avoid the oscillations of the reactor temperature. In all cases type-2 FLCs have a higher performance than type-1 FLCs.

References

1. Hale, J., Hocak, H.: Dynamics and Bifurcations. Springer, New York (1991)
2. Karnik, N.N., Mendel, J.M., Liang, Q.: Type-2 Fuzzy Logic Systems. IEEE Trans. on Fuzzy Systems 7, 643–658 (1999)

3. Mendel, J.M.: Uncertainty, Type-2 fuzzy Sets, and Footprints of Uncertainty. In: Proceeding of 9th Int'l. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems, Annecy, France, pp. 325–331 (2002)
4. Wu, D., Tan, W.W.: A type-2 Fuzzy Logic Controller for the liquid-level process. In: Proceedings of IEEE International Conf. on Fuzzy Systems, Budapest, Hungary, pp. 953–958 (2004)
5. Cazarez-Castro, N.R., Castillo, O., Aguilar, L.T., Cardenas, S.L.: From Type-1 to Type-2 Fuzzy Logic Control: A Stability and Robustness Study. In: Melin, P., Castillo, O. (eds.) International Congress on Fuzzy logic, Neural Networks and Genetic Algorithms, Tijuana, Mexico, pp. 308–321 (2005)
6. Lynch, C., Hagras, H., Callaghan, V.: Using Uncertainty Bounds in the Design of an Embedded Real-Time Type-2 Neuro-Fuzzy Speed Controller for Marine Diesel Engines. In: Proc. FUZZ-IEEE, Vancouver, Canada, pp. 7217–7224 (2006)
7. Galluzzo, M., Cosenza, B., Matharu, A.: Control of a nonlinear continuous bioreactor with bifurcation, by a type-2 fuzzy logic. Computers and Chemical Engineering 32, 2986–2993 (2008)
8. Klir, G.J., Wierman, M.J.: Uncertainty-based information. Physica-Verlag, Heidelberg (1998)
9. Mendel, J.M., Liang, Q.: Pictorial Comparisons of Type-1 and Type-2 Fuzzy Logic Systems. Intelligent Systems and Control, 280–285 (1999)
10. Perez, M., Albertos, P.: Self-oscillating and chaotic behaviour of a PI-controlled CSTR with control valve saturation. Journal of Process Control 14, 51–59 (2004)
11. Cao, J., Liu, H., Li, P., Brown, D.: Adaptive Fuzzy Logic Controller for Vehicle Active Suspension with Interval Type-2 Fuzzy Membership Functions. In: Proc. FUZZ-IEEE, Hong Kong (2008)
12. Karnik, N.N., Mendel, J.M.: Type-2 Fuzzy Logic Systems. In: Proceedings IEEE Conference of Systems, Man and Cybernetics, San Diego, pp. 2046–2051 (1998)

Evaluating Fuzzy Controller Robustness Using Model Checking

Giuseppe Della Penna[1], Benedetto Intrigila[2], and Daniele Magazzeni[1]

[1] Department of Computer Science, University of L'Aquila, Italy
[2] Department of Mathematics, University of Roma "Tor Vergata", Italy
{giuseppe.dellapenna,daniele.magazzeni}@univaq.it
intrigil@mat.uniroma2.it

Abstract. Fuzzy control is well known as a powerful technique for designing and realizing controllers. However, statistical evidence for their correct behavior may be not enough, even when it is based on a large number of samplings. Therefore, much work is being done to provide a systematic verification of fuzzy controllers and to asses their *robustness*, that is the ability of a controller to maintain good performance even in the presence of significant disturbances or parameter variations. In the present paper, we introduce a model checking based methodology for the fuzzy controller robustness analysis, that can be applied on plant-controller pairs in a nearly automatic way, giving higher precision results than other approaches, such as cell mapping. We support our conclusions with a case study that compares two different fuzzy controllers for the inverted pendulum on a cart problem.

1 Introduction

A control system (or, shortly, *controller*) is a small hardware/software component that controls the behavior of a larger system, called *plant*. In a closed loop configuration, the controller reads the plant state (looking at its *state variables*) and adjusts its *control variables* in order to keep it in a particular state, called *setpoint*, which represents its normal or correct behavior. In the last years, the use of sophisticated controllers has become very common in robotics, critical systems and, in general, in the hardware/software *embedded systems* contained in a growing number of everyday products and appliances.

Fuzzy control is well known as a powerful technique for designing and realizing controllers [1], especially suitable when a mathematical model is lacking or is too complex to allow an analytical treatment.

1.1 Motivations

The *robustness* of a controller is the ability to maintain good performance even in the presence of disturbances and/or parameter variations outside the design ranges.

Namely, if the state read from the plant is *unexpected*, i.e., not considered in the controller design, the controller may not be able to determine the right control action to take. In real applications, many causes, including environmental conditions, physical errors, deformation of materials, etc. may actually affect the state variables detection

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 303–311, 2009.

(or the state variables themselves) as well as the plant execution, and produce unexpected states. Therefore, ensuring a suitable robustness degree in the controller is often required, especially when the plant is a critical system, i.e., when malfunctioning may cause damage to persons or relevant economic losses.

For special kinds of systems this problem can be handled by analytical methods, directly including the robustness requirements into the controller design [2,3]. Unfortunately, these methods cannot be applied to fuzzy controllers, since they are defined in linguistic (and thus non-analityc) terms. In other cases, robustness problems can be handled by interpolation techniques. However, when the plant has a complex nonlinear dynamics the interpolation approach does not work well.

Cell mapping technology has been used to assess the robustness of fuzzy controllers [4]. Cell mapping allows an approximated analysis of the whole dynamics of the system. In particular, the state space is partitioned into a finite number of disjoint cells and the system dynamics is always approximated to the behavior of the geometric centers of the cells [5].

Since this technique involves a *complete* analysis of the approximated system dynamics, the increase of the number of cells gives rise to a noticeable increase of the computational effort. Therefore only a limited precision in the system analysis is possible. Moreover, due to the coarse granularity, it is difficult to analyze the effects of small disturbances on the executions of the control actions.

1.2 Our Contribution

In a previous paper, we applied explicit model checking techniques to the correctness verification of fuzzy controllers [6]. In this paper we extend the approach of [6] to a general methodology for the *robustness analysis* of fuzzy controllers, that can be easily applied to nonlinear systems. Our technique is based on the CMurφ explicit model checker [7], whose special features, described in Section 3, allow to easily interface with plant simulators and fuzzy controllers.

The use of explicit model checking, together with the features of CMurφ, allows us to handle very large systems (say, millions of states) without incurring in the memory explosion problem. Moreover, we can easily manage continuous systems with very high approximation, generating more complete robustness certifications. Finally, our technique is nearly automatic, requiring small efforts to be applied to any system.

The paper is organized as follows. In Section 2 we describe the parameters and measures used to evaluate the fuzzy controller robustness. In Section 3 we briefly introduce model checking techniques and the CMurφ tool, and give details about how we modeled robustness analysis within it. Section 4 shows the experimental results of our technique applied to the analysis of two different fuzzy controllers for the *inverted pendulum on a cart* problem. Finally, Section 5 contains some concluding remarks.

2 Controller Robustness Measures

To evaluate the robustness of a fuzzy controller, we analyze how it reacts to stimuli that are not addressed by the plant specification.

More formally, let a plant with *discrete time* dynamics $f(x, u)$ be given, where x is the vector of state variables and u the vector of control variables. Then a *controller*

is a function $k(x)$ that, given a plant state x, returns the values of control variables $u = k(x)$ that are required to reach the plant *setpoint* x_G from x.

A *trajectory* is a sequence of states $(x_0, \ldots, x_n)$ where $x_n = x_G$ is the setpoint and $x_{i+1} = f(x_i, k(x_i))$ for all $i = 0 \ldots n-1$. In other words, a trajectory is the sequence of states of the plant driven by the controller from an initial state to the setpoint.

To evaluate the controller robustness we measure how much a *disturbed* trajectory $\tilde{t} = (\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_n)$ deviates from the corresponding *ideal* trajectory $t = (x_0, x_1, \ldots, x_n)$ [4]. Here, an ideal trajectory is generated in the (ideal) environment given by the plant specification, whereas a disturbed trajectory is generated in a context that applies random or systematic variations to the plant parameters and/or variables, as it usually happens when the plant is executed in a real environment.

In particular we consider the following disturbances applied to trajectories:

- **Parameters Variation.** In this case, we test if the controller can handle plants whose design parameters are different, to a given extent, from the specification. Indeed, these variations occur in practice when the plant is a complex system, whose implementation is subject to errors and approximations.
- **State/Control Disturbances.** In this case, we test if the controller is able to drive the plant to its setpoint when the control and/or state variables are subject to external disturbances. This is very common in plants due to physical effects (such as friction).

In general, there are many different parameters that can be used to measure the deviation of such trajectories. As in [4], we consider the following measures.

1. **Final Trajectory Error.** The final error, e_f, is the Euclidean distance between the final state of a disturbed trajectory $\tilde{t}$ and the final state of the ideal trajectory t with the same starting point $x_0 = \tilde{x}_0$. It is given by $e_f = \|x_G - \tilde{x}_G\|$ where the symbol $\|\cdot\|$ denotes the Euclidean norm. We consider the maximum and average values of this error, calculated in each controller trajectory.
2. **Actual Trajectory Error.** The actual trajectory error, e_t, is an averaged Euclidean distance between a disturbed trajectory and its ideal trajectory. It is computed by averaging the distances between corresponding points in the two trajectories. It is given by $e_t = \frac{\sum_{i=1}^{n} \|x_i - \tilde{x}_i\|}{n}$ where n is the number of states in the trajectory. When a trajectory terminates at its setpoint before the other, distances continue to be measured between points of the unterminated trajectory and the setpoint of the terminated trajectory. We consider the maximum and average values of this error, calculated in each controller trajectory.
3. **Trajectory Length Error.** The trajectory length error, e_d, is the difference between the lengths of a disturbed trajectory and its ideal counterpart measured as follows: $e_{tl} = \left\| \sum_{i=1}^{n-1} \|x_{i+1} - x_i\| - \sum_{i=1}^{n-1} \|\tilde{x}_{i+1} - \tilde{x}_i\| \right\|$. We consider the maximum, minimum and average values of this error, calculated in each controller trajectory.
4. **Number of Controllable States.** A controllable state is a plant state from which the controller is able to reach the setpoint. Therefore, every state in a trajectory that reaches the setpoint is controllable. This value measures the difference between the number of controllable states of ideal and disturbed trajectories, to see if disturbances make some state not controllable.

5. **Time Optimality.** The time optimality value is computed as the difference between the average number of steps required to reach the setpoint in each disturbed and ideal trajectory, respectively.

3 Model Checking

Generally speaking, model checking [8,9,10,11,12] can be defined as the formal process of verifying the validity of a set of assertions on the behavior of a system, modeled by a *Finite State System*. The model checking algorithms always involve a (symbolic or explicit) analysis of the state space, reachable from the initial states. This analysis is called *reachability analysis*.

In our case, the continuous plant dynamics $f(x, u)$ is transformed, by a direct quantization of the continuous variables [13], into a discrete *transition relation*. This relation, in turn, defines a graph, (*transition graph*). The set of *reachable* states is obtained by visiting the transition graph, starting from the *initial* states. We consider as *initial* states all the states whose variables values range over suitable intervals or sets (the so-called *computational space* or *region*).

During the visit, the verifier checks whether the given constraints are satisfied by each state reached. Therefore, since the system has a finite number of states, the verifier will eventually explore all the transition graph, thus certifying that the constraints hold in each possible system state. Otherwise, if a state is reached that does not satisfy one of the constraints (i.e., it is an *error* state), the verifier will dump the graph path that led to this error state (the *error trace*).

In our robustness application, as described in Section 3.1, we take advantage of the reachability analysis algorithms to simultaneously generate a disturbed trajectory and the corresponding ideal trajectory, and set the constraints to verify that the robustness measures never exceed a specified threshold. This is accomplished using the CMurφ tool [7], an extended version of the Murφ [11] model checker, originally developed to verify protocol-like systems.

CMurφ has two important functionalities required to deal with the complexity of the plant simulators and the fuzzy controller [14], which make many other well-known verifiers such as SPIN [12] unsuitable for our purposes. Indeed, the verifier can use external C/C++ functions, so it can embed the fuzzy controller and the plant simulator, without having to re-model them in terms of another language, which could introduce errors and approximations in the system behavior and require substantial rewriting efforts. Moreover, CMurφ is able to handle (finite precision) real numbers, and this is important since most plants work on continuous variables or combinations of discrete/continuous variables.

3.1 Measuring Robustness with CMurφ

To measure the robustness of a fuzzy controller FC applied to the plant P, we use a particular CMurφ model M_P that embeds two copies of the plant model, P_1 and P_2, both controlled by FC.

The system state is a pair (s_1, s_2) where s_1 is the state of P_1 and s_2 of P_2, respectively. The model allows both plants to evolve in parallel, under the guidance of the same controller, starting from the same initial state and with the same setpoint.

In particular, P_1 is defined with the exact parameters given by the plant specification, and generates ideal trajectories, whereas P_2 is initialized with different parameters and/or has random disturbances added to each state (or control actions), so it generates the corresponding disturbed trajectory.

Note that the controller, and possibly the plant model simulator, are external programs linked to the CMurφ model, which acts as a simulation "runner" and "observer". Therefore, the use of CMurφ does not introduce any external error or approximation in the plant and controller behavior, ensuring a correct robustness measure.

While the model evolves, we use the CMurφ invariants to obtain the needed measures. In particular, we collect the information needed to measure the robustness parameters described in Section 2.

Thanks to the exhaustive state space exploration performed by the model checker, we are able to take our measures on a very large number of trajectories, so obtaining the full coverage of *all the possible trajectories handled by the controller*, within the given approximations. Then we use the minimum, maximum and average value of such measures to generate the final robustness report.

4 Experimental Results

To show the potential of our model checking based approach to the robustness evaluation, in this Section we present the experimental results related to the same case study proposed by Papa et al. in [4]. Namely, we consider a Feedback Takagi-Sugeno (TS) fuzzy controller [15], and a Smith-Comer TS fuzzy controller [16] for an *inverted pendulum on a cart*.

The robustness analysis and comparison have been performed considering both types of disturbances and parameter variations described in Section 2, namely variations in the half length of the inverted pendulum and random disturbances applied to the control force value.

In the following, we first describe the inverted pendulum problem, then we give some details about the two controllers and finally we compare their performances by considering the measures discussed in Section 2.

4.1 Inverted Pendulum on a Cart

A controller for the inverted pendulum on a cart has to bring the pendulum to equilibrium by applying a horizontal force to the cart.

The pendulum state is described by two real variables: the pendulum angle w.r.t. the vertical axis θ and the angular velocity $\dot{\theta}$.

The normalized system is described by the differential equation

$$\begin{aligned} \dot{x}_{(1)} &= x_{(2)} \\ \dot{x}_{(2)} &= \frac{g_e \sin(x_1) - \left[\cos(x_1)/(m_p+m_c)\right]\left[m_p l x_2^2 \sin(x_1) + u\right]}{4l/3 - \left[m_p/(m_p+m_c)\right] l \cos^2(x_1)} \end{aligned}$$

where $x_{(1)} = \theta$, $x_{(2)} = \dot{\theta}$, g_e is the gravitational constant, m_p is the mass of the pole, m_c is the mass of the cart, l is the half-length of the pole and u is the force applied to the cart. The model parameters are as follows: $\theta \in [-1.5, 1.5]\, rad$, $\dot{\theta} \in [-8, 8]\, \frac{rad}{sec}$, $u \in [-50, 50]\, N$, $e_g = 9.8 m/s^2$, $l = 0.5m$, $m_p = 0.1kg$, $m_c = 0.9kg$. The sampling period T_s is 0.01s.

4.2 The Takagi-Sugeno Fuzzy Controllers

In this section we give a short description of the two controllers under analysis. A more detailed explanation of their characteristics can be found in the cited papers.

Feedback TS Fuzzy Controller. We consider a Feedback TS Fuzzy controller designed using the methodology presented in [15]. The authors use genetic algorithms with cell map information as feedback, so obtaining a controller with maximal stability and controllability. The controller has nine linguistic rules of the form: IF error(x_1) is *Region* AND error-change(x_2) is *Region* THEN control action is $c_{j1}x_1 + c_{j2}x_2 + c_{j3}$, where *Region*=$\{$*Pos*, *Zero*, *Neg*$\}$. The values of c_{j1}, c_{j2}, c_{j3} can be found in [4].

Smith-Comer TS Fuzzy Controller. The controller described in [16] is generated through an automated method for calibrating fuzzy controllers, based on the cell state space concept. In particular, the system state space is quantized into cells, creating a discrete model of the system behavior. Then, given a cost function and a plant simulation model, the technique generates a numerical controller (of minimum cost) that is finally approximated using a fuzzy logic controller.

The control algorithms have been implemented as described in corresponding papers and embedded as external C functions in CMurφ model to perform the verification.

4.3 The Model Checking Framework

To set up our model checking framework, we discretized the model as follows. We used for the state variables a discretization step of $0.001 rad$ for θ and $0.01 \frac{rad}{sec}$ for $\dot{\theta}$. In this way, we have a *computable region* of 4,804,601 states. We fixed a *time horizon* of 60 steps: that is, if the controller is not able to drive the pendulum to the setpoint in no more than 60 steps, we mark as not controllable the initial state of the current trajectory.

The experiments have been carried on a Intel Core2 Duo T5600 1,83GHz machine with 2 Gb of RAM. To define the initial states, we discretized the variables splitting both θ and $\dot{\theta}$ in 502 regions, and thus resulting in 252,004 initial states. However, note that, since each state in a trajectory is also an initial state of a (sub)trajectory, actually with these settings our experiments cover more than 3,5 millions of states.

4.4 Controller Robustness in the Presence of Parameter Variations

To test robustness with respect to parameter variations, we considered variations in the inverted pendulum half length (l) ranging from 0.5m to 0.75m, that is up to 50% of its design value. The mass (m) has been adjusted accordingly. Results are in Table 1, where column *Ctrl* indicates the controller used (i.e., *FB* for the Feedback TS Fuzzy Controller and *SC* for the Smith-Comer TS Fuzzy Controller) and column *Dist* indicates the l parameter variation. The other columns contain the robustness measures discussed in Section 2. In particular, column *Controllable States* indicates the percentage

Table 1. Experimental results for parameters variation

Ctrl	Dist	Controllable States	Time Optim.	Traj Length Err avg	min	max	Final Traj Err avg	max	Actual Traj Err avg	max
FB	0.5	79.67%	25.2049	—	—	—	—	—	—	—
FB	0.55	76.96%	28.6845	0.0107	-0.01	0.08	0.0631	0.2	0.4088	8.1544
FB	0.6	76.36%	30.6161	0.0215	-0.01	0.12	0.0744	0.2021	0.6760	8.4856
FB	0.65	75.48%	31.4019	0.0283	0	0.16	0.0777	0.203	0.8842	8.7161
FB	0.7	74.27%	32.3019	0.036	0	0.19	0.081	0.2057	1.0519	9.0669
FB	0.75	72.85%	33.0624	0.0446	0	0.23	0.0835	0.2112	1.1971	9.1865
SC	0.5	76.93%	27.6753	—	—	—	—	—	—	—
SC	0.55	74.26%	33.0622	0.0193	-0.03	0.08	0.0582	0.2002	0.4701	9.0952
SC	0.6	73.19%	33.7244	0.0294	-0.04	0.15	0.0603	0.2003	0.8118	9.033
SC	0.65	70.92%	34.3319	0.0409	-0.05	0.2	0.0619	0.2043	1.063	9.1693
SC	0.7	68.97%	35.2210	0.0533	-0.04	0.23	0.0651	0.2077	1.2562	9.4164
SC	0.75	66.71%	36.2353	0.0678	-0.04	0.3	0.0678	0.2227	1.3999	9.8515

of controllable states with respect to the whole computable region of 4,804,601 states. Figure 1 shows a graphical comparison of some relevant measures.

Comparing our results with those of [4] we find that the general assessment of the performances of the two controllers are qualitative similar. However, there are several differences: as an example, we find that the Smith-Comer Controller has a better performance w.r.t. the final trajectory error, contrary to the findings of [4]. We think that our results are more accurate since our evaluation is based on the analysis of 3,827,825 states, to be compared with the analysis of 10,201 cells in [4].

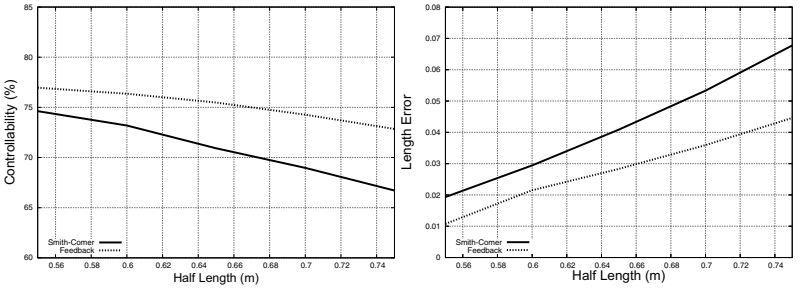


Fig. 1. Controllability and trajectory length error with parameters variation

4.5 Controller Robustness in the Presence of Random Disturbances

To test robustness with respect to state and control variables disturbances, we applied random disturbances of 5%,10%,15%,20%,25% to the force value (that is the output of the controller). Results are in Table 2, where columns have the same meaning of Table 1. Figure 2 shows a graphical comparison of some relevant measures. The results are comparable with the ones in Table 1.

Note that this kind of analysis cannot be performed with the cell mapping techniques since, generally speaking, the disturbances are small and do not essentially change the cell dynamics.

Table 2. Experimental results for random control disturbances

Ctrl	Dist	Controllable States	Time Optim.	Traj Length Err avg	min	max	Final Traj Err avg	max	Actual Traj Err avg	max
FB	0%	79.67%	25.2049	—	—	—	—	—	—	—
FB	5%	79.64%	25.3292	0.0003	-0.02	0.02	0.0127	0.07	0.0068	0.16
FB	10%	79.59%	25.5568	0.0004	-0.02	0.03	0.0239	0.11	0.0167	0.24
FB	15%	79.53%	25.9439	0.0005	-0.02	0.03	0.0336	0.16	0.0269	0.3
FB	20%	79.46%	26.4289	0.0007	-0.03	0.04	0.0419	0.2	0.0376	0.41
FB	25%	79.23%	26.9306	0.0008	-0.03	0.06	0.0495	0.2001	0.0485	0.55
SC	0%	76.93%	27.6753	—	—	—	—	—	—	—
SC	5%	76.88%	27.8387	0.0001	-0.02	0.02	0.017	0.16	0.009	5.1349
SC	10%	76.37%	28.2155	0.0003	-0.03	0.03	0.0301	0.18	0.0217	6.5223
SC	15%	76.53%	28.6886	0.0007	-0.03	0.03	0.0396	0.19	0.0356	6.873
SC	20%	76.25%	29.1636	0.0014	-0.03	0.05	0.0454	0.2	0.0492	8.4425
SC	25%	76.06%	29.6458	0.0019	-0.04	0.08	0.0491	0.2	0.0641	8.513

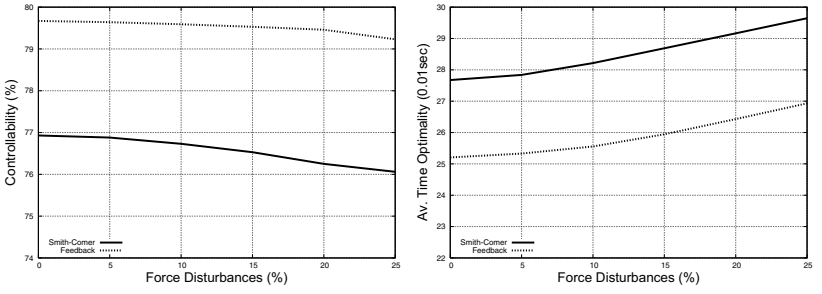


Fig. 2. Controllability and time optimality with random control disturbances

5 Conclusions

In this paper we showed a general methodology for the robustness analysis of fuzzy controllers. Our methodology exploits model checking techniques and, in particular, the CMurφ verifier, to deal with the complexity of the plant-controller pair, and can be therefore applied to fuzzy controllers even on nonlinear systems with minimal effort.

The results show that this technique can achieve a *better precision* than the one obtained with the cell mapping approach.

Indeed, the two techniques can be considered as *complementary*, since the cell mapping gives information about the *global dynamics* of the system.

Therefore the proposed technique, and in particular the CMurφ verifier, can be actually considered as an effective tool for robustness verification in real world applications.

References

1. Jin, J.: Advanced Fuzzy Systems Design and Applications. Physica-Verlag (2003)
2. Isidori, A.: Nonlinear Control Systems II. Springer, Heidelberg (1999)
3. Levine, W.: The Control Handbook. Jaico Publishing House (2005)
4. Papa, M., Wood, J., Shenoi, S.: Evaluating controller robustness using cell mapping. Fuzzy Sets and Systems 121(1), 3–12 (2001)

5. Hsu, C.: Cell-to-Cell Mapping. Springer, Heidelberg (1987)
6. Intrigila, B., Magazzeni, D., Melatti, I., Tofani, A., Tronci, E.: A model checking technique for the verification of fuzzy control systems. In: Proceedings of Int. Conf. on Computational Intelligence for Modelling Control and Automation (CIMCA 2005) (2005)
7. CMurphi Web Page, http://www.di.univaq.it/gdellape/murphi/cmurphi.php
8. Clarke, E.M., Grumberg, O., Peled, D.A.: Model Checking. MIT Press, Cambridge (1999)
9. Burch, J.R., Clarke, E.M., McMillan, K.L., Dill, D.L., Hwang, L.J.: Symbolic model checking: 10^{20} states and beyond. Inf. Comput. 98(2)
10. Holzmann, G.J.: Design and Validation of Computer Protocols. Prentice-Hall, Englewood Cliffs (1991)
11. Murphi Web Page, http://sprout.stanford.edu/dill/murphi.html
12. SPIN Web Page, http://spinroot.com
13. Della Penna, G., Intrigila, B., Magazzeni, D., Melatti, I., Tofani, A., Tronci, E.: Automatic generation of optimal controllers through model checking techniques. In: Informatics in Control, Automation and Robotics. LNEE, vol. 15, pp. 107–122. Springer, Heidelberg (2008)
14. Della Penna, G., Intrigila, B., Melatti, I., Tronci, E., Venturini Zilli, M.: Exploiting transition locality in automatic verification of finite state concurrent systems. STTT 6(4) (2004)
15. Hu, H., Tai, H., Shenoi, S.: Incorporating cell map information in fuzzy controller design. In: Proceedings of the 3rd IEEE Conf. on Fuzzy Systems (1994)
16. Smith, S., Comer, D.: An algorithm for automated fuzzy logic controller tuning. In: Proceedings of IEEE Int. Conf. on Fuzzy Systems 1992, pp. 615–622 (1992)

Learning Fuzzy Systems by a Co-Evolutionary Artificial-Immune-Based Algorithm

Luiz Lenarth G. Vermaas, Leonardo M. Honorio,
Muriel Freire, and Daniele Barbosa

Institute of Technology and Electrical Engineering, UNIFEI,
Av. BPS 1303, Centro, MG - BR
{lenarth,demello,muriel,danieleb}@unifei.edu.br

Abstract. To create a Fuzzy System from a numerical data, it is necessary to generate rules and memberships representing the analyzed set. This goal demands to break the problem into two parts: one responsible for learning the rules and another responsible for optimizing the memberships. This paper uses a Gradient-based Artificial Immune System with a different population for each of these parts. By simultaneously co-evolving these two populations, it is possible to exchange information between them enhancing the fitness of the final generated system. To demonstrate this approach, a fuzzy system for autonomous vehicle maneuvering was developed by observing a human driver.

Keywords: Autonomous Vehicle, Co-Evolutionary Artificial Immune Systems, Fuzzy System Learning.

1 Introduction

In some applications the expert knowledge can be imprecise or may not be enough to assemble fuzzy controllers. In scenarios like these, concepts about Machine Learning and Data Mining can be used to analyze an initial data set (IDS) and generate both fuzzy rules and membership functions.

The literature presents several methodologies to extract fuzzy rules from numerical data [1]- [4]. One of the most citted approaches is presented in [4], which shows an algorithm that uses standard memberships and a simple logic to extract fuzzy rules capable of representing a given data set. This logic is based on associating the input and the output over the standard memberships where each entry of the original data set may generate a rule. The output of this step is a set of rules, which will be treated to eliminate inconsistency. This method also presents good results and is able to deal with large data sets. However, this approach assumes standard and fixed memberships and the final rule set may be quite inexpressive, because it uses several considerations to avoid inconsistency and redundancy. To improve a similar approach presented in [5], this work developed a method using a Co-Evolutionary Gradient-Based Artificial-Immune-System with two different types of populations: one is responsible for optimizing

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 312–319, 2009.

the memberships and the other for learning the rules. The Co-Evolutionary process enables information change between the populations enabling to find a more representative fuzzy system at the end of simulation.

In order to evaluate this methodology, an application of automatic parallel parking will be used, where a fuzzy system for autonomous maneuvering will be learnt by a data set of actions taken by a human driver.

2 Generating a Fuzzy System from a Numerical Data Set

Suppose a given numerical input and output data set acquired from a given environment. The goal is to automatically generate a fuzzy system that represents the relation between the inputs and outputs of this application. Such system is composed of two types of elements: membership functions and rules. Membership functions may be present in different sets of numbers, shapes and positions, and finding the best configuration depends on the systems data and rules. The literature presents several works dedicated to generating fuzzy systems from data [6]-[8] using three different approaches. One approach gets the rules from an expert and uses population-based algorithms to optimize the memberships [9], the second approach provides the memberships and learns the rules [4], and, finally, some algorithms use two different populations to learn both rules and memberships [6]. As the optimization of membership functions and the rule inference process are dependent on each other, they should be modeled together to reach the best representation. Adopting this approach, we present CAISFLO, which stands for Co-Evolutionary Artificial-Immune-Based Algorithm Applied to Fuzzy Systems Learning and Optimization, based on two different sets of populations co-evolving together to find the best representation of a real application. One set of population is responsible for optimizing the memberships while the other for learning the rules. However, even though they have the same goal, i.e. to generate an accurate representation of a data set, changes that improve one population may destroy the other. To avoid this situation, a full Pareto optimization [5] is adopted, meaning that improvements in one population will be only allowed if it does not jeopardize the other. As already stated, fuzzy rules are strongly connected and depended on membership functions. It is impossible to find any rule in a data set if these functions have not been defined. Thus, the first step of assembling a fuzzy system is to generate a population of membership functions fmPop= $\{\text{fAb}_1,\ldots,\text{fAb}_n\}$. For each individual fAbi, a new population rPopi = $\{\text{riAb}_1,\ldots,\text{riAb}_m\}$ responsible for learning the inference rules is created and evolved. After the rules are learnt, each individual fAbi of the first population will have its memberships optimized to enhance the accuracy. This process is shown in Fig. 1 and continues until the end of the simulation.

As the main purpose of this approach is to correctly represent a real system given a data set, the fitness of antibody fAb_i is given by

$$fitness(fAb_i) = \sum_{a=1}^{ne} (fAb_i(ipDs_a) - opDs_a)^2 + penalty(ipDs_a). \quad (1)$$

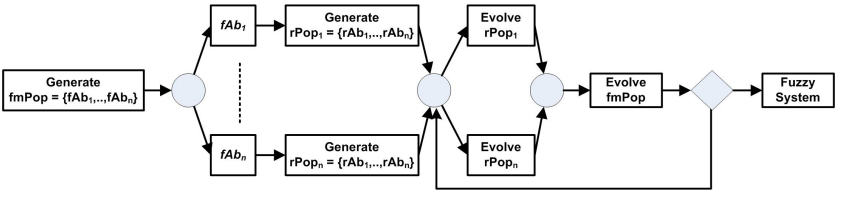


Fig. 1. Diagram of the Co-Evolutionary process

where, Ne is the number of entries in the data set (Ds), ipDS_a represents the input variables vector in entry a of Ds, opDs_a represents the output value of entry a, and $Penalty(.)$ is a function that returns a very large value (i.e. 10^{99}) if fAb_i does not have any rule to deal with ipDs_a. The main idea of how these populations individually work are shown next.

3 Learning Fuzzy Rules

In order to facilitate and emphasize the explanation of the proposed method, suppose a numerical data set with two input variables (x_1 and x_2) and an output variable (y). The given members of the data set are represented as:

$$(x_1^1, x_2^1; y^1), (x_1^2, x_2^2; y^2), ..., (x_1^n, x_2^n; y^n). \tag{2}$$

The first step of the proposed methodology consists of generating an available rule table (ART). This table contains all possible rules from a given data input taking into consideration a set of membership function configuration. The ART is generated from the division of the input and output range into fuzzy regions, similarly to step 1 of the proposal made by [4]. Each domain interval, i.e., the range comprehended between the minimum and the maximum values of a data set variable must be divided into 2N+1 regions where N can be different for each variable. Fig. 2 shows the division of the interval domains into fuzzy regions. In this case, the domain interval of the input variable x_1 is divided into three regions (N=1), the input variable x_2 into five regions (N=2), and the output variable y also into five regions (N=2). After that, according to [4], each data set entry may generate a single fuzzy rule based on the highest membership degree of every variable. For example, Figure 2 shows that x_1^1 has a membership degree of 0.6 at S1, 0.4 at M and zero at the other regions. Likewise, x_2^1 has a degree of 1 at M and zero at the other regions, whereas y^1 has a degree of 0.8 at L_1 and 0.2 at M. Thus, this entry generates the following rule:

It is well-known that the numerical output value of a fuzzy system depends on the activated membership functions with its respective degree. It is important to highlight that all of the activated membership functions with a higher or a lower degree contribute to the output value calculation. Therefore, generating fuzzy rules based on the highest membership degree, as previously shown, means a simplification of the problem. Instead of generating rules based on membership

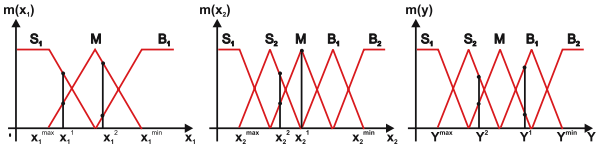


Fig. 2. Division of the input and output space into fuzzy regions

degrees, our proposal is to generate a set of rules combining all fuzzy regions that have activated membership functions, whatever the membership degree is. Thus, instead of generating a single fuzzy rule according to what was shown in Eq. 3, the given member of the data set (x_1^1, x_2^1, y^1) will generate:

$$\begin{array}{l}(x_1^1,x_2^1;y^1) \Rightarrow \text{IF } x_1 \text{ is } S_1 \text{ AND } x_2 \text{ is M, THEN y is } B_1 \Rightarrow \text{Rule 1}\\ (x_1^1,x_2^1;y^1) \Rightarrow \text{IF } x_1 \text{ is } S_1 \text{ AND } x_2 \text{ is M, THEN y is M } \Rightarrow \text{Rule 2}\\ (x_1^1,x_2^1;y^1) \Rightarrow \text{IF } x_1 \text{ is M AND } x_2 \text{ is M, THEN y is } B_1 \Rightarrow \text{Rule 3}\\ (x_1^1,x_2^1;y^1) \Rightarrow \text{IF } x_1 \text{ is M AND } x_2 \text{ is M, THEN y is M } \Rightarrow \text{Rule 4}\end{array} \quad (3)$$

At first, several conflicting rules, i.e., rules that have the same antecedent (IF part) but a different consequent (THEN part), are generated when this procedure is adopted. In order to solve this problem, the ART is built in a way that each line represents a unique antecedent and each column of the consequent part represents the output membership functions. Each antecedent will be associated with a certain number of activated output memberships, signed with a digit "1" in the ART. As an example, Tab. 1 shows the generated ART from the given members (x_1^1, x_2^1; y^1) and (x_1^2, x_2^2; y^2) taken from Fig. 2. In this case, the output variable y is represented by five columns, S_2, S_1, M, B_1 and B_2, that correspond to their membership functions.

Table 1. Available Rules Table - ART

Antecedent		Consequent – Output y				
		S2	S1	M	B1	B2
IF x_1 is S1 **AND** x_2 is M	**THEN**	0	0	1	1	0
IF x_1 is M **AND** x_2 is M	**THEN**	0	1	1	1	0
IF x_1 is M **AND** x_2 is S1	**THEN**	0	1	1	0	0
IF x_1 is B1 **AND** x_2 is S1	**THEN**	0	1	1	0	0
IF x_1 is B1 **AND** x_2 is M	**THEN**	0	1	1	0	0

4 The Gradient-Based Artificial Immune System

The main purpose of the Artificial Immune System (AIS) is to use the successful Nature Immune System process in optimization and learning [10]. As every intelligence-based method, AIS is a search methodology that uses heuristics to explore only interesting areas in the solution space. However, unlike other intelligent methods, it provides tools to simultaneously perform local and global searches. These tools are based on two concepts: hypermutation and receptor

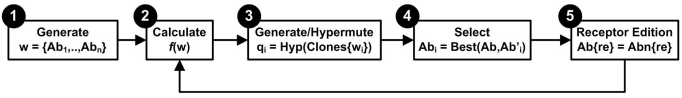


Fig. 3. GbCLONALG flowchart

editing. While hypermutation is the ability to execute small steps toward a higher affinity Ab leading to local optima, receptor editing provides large steps through the solution space, which may lead to a region where the search for a hAb is more promising.

The technical literature shows several AIS algorithms with some variants. One of the most interesting ones is the GbCLONALG algorithm presented in [11]. The main statement of GbCLONALG is that progressive adaptive changes can be achieved by using numerical information captured during the hypermutation process. There are several possible ways to capture this information. The one used in this article is the tangent vector technique, because it treats the objective function as a "black box", where small disturbances are individually applied over each dimension of the input vector yielding the vector:

$$\mathrm{TV}\left(\mathrm{f}\left(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n}\right)\right)=\begin{bmatrix}\dfrac{\mathrm{f}\left(\mathrm{x}_1+\Delta\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n}\right)-\mathrm{f}\left(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n}\right)}{\left|\Delta\mathrm{x}_1\right|}\\ \vdots \\ \dfrac{\mathrm{f}\left(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n}+\Delta\mathrm{x}_\mathrm{n}\right)-\mathrm{f}\left(\mathrm{x}_1,\ldots,\mathrm{x}_\mathrm{n}\right)}{\left|\Delta\mathrm{x}_\mathrm{n}\right|}\end{bmatrix} \tag{4}$$

where: n is the number of control or input variables, f(.) is the objective function to be optimized; $x_1, ..., x_n$ are input variables; and $\Delta \mathrm{x}_k$ is a random increment applied to x_k. This approach enhances the algorithm efficiency and enables a well defined stop criterion. The complete algorithm is shown in Fig.3.

Each step or block of the previous diagram is detailed as follows: *1.* Randomly choose a population $w = Ab_1, ..., Ab_n$, with each individual defined as $Abi = x_1, ..., x_j, ..., x_n c$, where nc represents the number of control variables or actions; *2.* Calculate the value of the objective function for each individual; this result provides the population affinity for the optimization process; *3.* Proceed with the Hypermutation process. *4.* The bests nb individuals among the original w population are selected to stay for the next generation. The remaining individuals are replaced by randomly generated new Ab's. This process simulates the receptor editing and helps in searching for better solutions in different areas.

Although this algorithm has presented very good results in continuous optimization scenarios [11] it is necessary to adapt its principles to carry out combinatorial search problems. To accomplish that goal, some considerations about the definition of antibodies, the tangent vector calculation and the receptor editing must be taken into account.

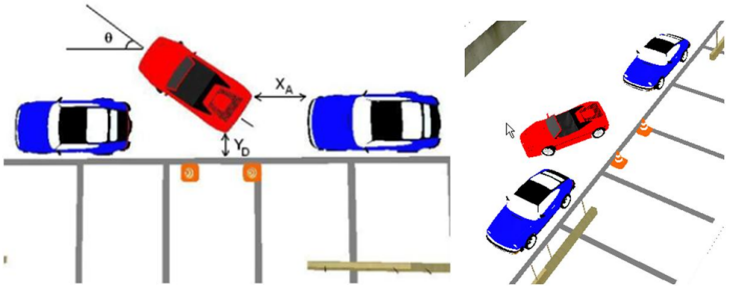


Fig. 4. Simulation software and the variables considered in the fuzzy control

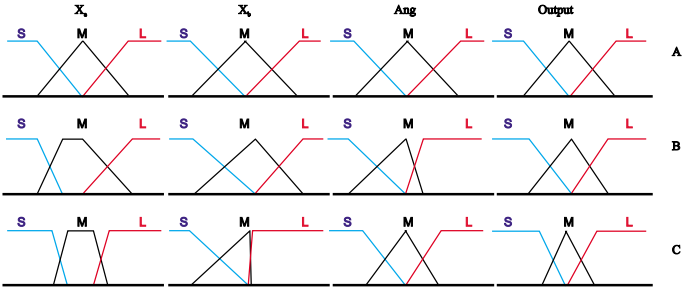


Fig. 5. (a) Membership Functions without optimization (b) Optimized Membership Functions, and (c) Optimized Membership Functions after co-evolution

5 Applications and Results

In order to validate the proposed method, this paper presents an application of automatic parallel parking using a 3D software [12] that allows the reproduction of vehicle dynamics. Fig. 4 illustrates the environment, as well as the considered input variables.

To build the data set, a parallel parking maneuver was manually carried out by using a joystick. During this process, the input variables XA, YD and Θ as well as the OUTPUT value obtained by reading a virtual encoder connected to the steering wheel, are stored in a data set, yielding 256 entries.

To extract a fuzzy system from this data set, capable of reproducing the human control over the vehicle the CAISFLO algorithm was used. The first step is to generate an initial population of memberships and, for each one, build the ART. To illustrate this process, Figure 5(a) shows the initial membership configuration of an antibody generated according to [4] for comparison reasons. For this antibody, the ART is shown in Table 2a, where 22 possible rules were considered. If the methodology proposed in Wang&Mendel [4] had been used, just 6 rules would have been generated, as shown in Table 2b. As the populations evolve, the memberships and rules start to assume new shapes with better results. Figure 5(b) shows the result of CAISFLO's first generation where it is

Table 2. (a) Available Rules Table (b) ART, Rules obtained from Wang and Mandel (c) Rules obtained from CAISFLO

a)

Antecedent												Consequent		
												S	M	B
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	S	**THEN**	0	1	1
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	M	**THEN**	0	1	1
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	0	0	1
IF	X_A	=	M	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	1	1	1
IF	X_A	=	M	**AND**	Y_D	=	M	**AND**	ANG	=	B	**THEN**	1	1	1
IF	X_A	=	B	**AND**	Y_D	=	M	**AND**	ANG	=	B	**THEN**	0	0	1
IF	X_A	=	S	**AND**	Y_D	=	M	**AND**	ANG	=	B	**THEN**	1	1	1
IF	X_A	=	S	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	1	1	1
IF	X_A	=	S	**AND**	Y_D	=	S	**AND**	ANG	=	B	**THEN**	1	1	0
IF	X_A	=	B	**AND**	Y_D	=	S	**AND**	ANG	=	B	**THEN**	1	1	0

b)

Antecedent												Consequent			
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	M	**THEN**	OUTPUT	=	B
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	OUTPUT	=	B
IF	X_A	=	M	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	OUTPUT	=	B
IF	X_A	=	M	**AND**	Y_D	=	M	**AND**	ANG	=	B	**THEN**	OUTPUT	=	B
IF	X_A	=	S	**AND**	Y_D	=	M	**AND**	ANG	=	B	**THEN**	OUTPUT	=	S
IF	X_A	=	S	**AND**	Y_D	=	S	**AND**	ANG	=	B	**THEN**	OUTPUT	=	S

c)

Antecedent												Consequent			
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	S	**THEN**	OUTPUT	=	M
IF	X_A	=	B	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	OUTPUT	=	B
IF	X_A	=	M	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	OUTPUT	=	B
IF	X_A	=	S	**AND**	Y_D	=	B	**AND**	ANG	=	B	**THEN**	OUTPUT	=	M
IF	X_A	=	S	**AND**	Y_D	=	S	**AND**	ANG	=	B	**THEN**	OUTPUT	=	S

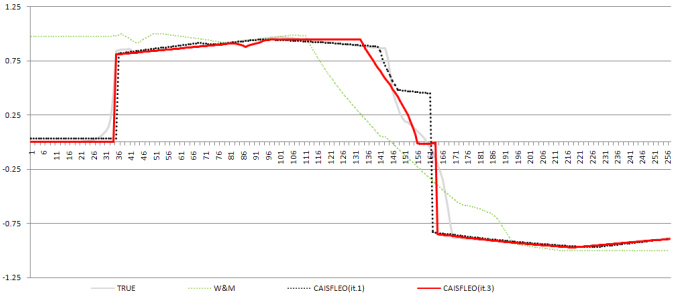


Fig. 6. The original (TRUE) output from the data set, the results of the W&M method, and the 1st and final generations of CAISFLO

possible to see that the membership function MXA assumed a trapezoidal shape. Figure 5(c) and Table 2c showed the final result of memberships and the rules, respectively, obtained after 3 generations. It is important to note that the final number of rules obtained from the present proposal have found only 5 rules. Although it was just one rule lower than the method of W&M, the rules are different and have more accuracy, as can be seen in Figure 6 where the y-axis represents the virtual encoder value and the x-axis the entries in the data set.

6 Conclusion

This paper presented a co-evolutionary artificial immune based algorithm applied to generate fuzzy systems from numerical data. For that purpose, two sets

of populations were used: one designed to learn rules and another to optimize membership functions. To reduce the solution space two strategies were adopted. The first one was to build a table containing just valid rules. The second was to adopt GbCLONALG in order to reduce the number of clones and to search just interesting areas. CAISFLO algorithm was tested in a maneuver learning scenario, where a user parked a virtual car and, using the stored data, the respective fuzzy system was achieved, tested and compared with another method. The results proved the efficiency of the algorithm. As future work, the algorithm will be tested in a real autonomous car maneuvering scenario.

References

1. Pal, N.R., Chakraborty, S.: Fuzzy Rule Extraction from ID3-Type Decision Trees for Real Data. IEEE Transactions on Systems, Man, and Cybernetics- Part B: Cybernetics 31(5), 745–754 (2001)
2. Quinlan, J.R.: C4.5.: Programs for Machine Learning. Morgan Kaufmann, San Mateo, Inc. (1993)
3. Alves, R.T., Delgado, R.M., Lopes, H.S., Freitas, A.A.: An Artificial Immune System for Fuzzy-Rule Induction in Data Mining. In: Yao, X., Burke, E.K., Lozano, J.A., Smith, J., Merelo-Guervós, J.J., Bullinaria, J.A., Rowe, J.E., Tiňo, P., Kabán, A., Schwefel, H.-P. (eds.) PPSN 2004. LNCS, vol. 3242, pp. 1011–1020. Springer, Heidelberg (2004)
4. Wang, L., Mendel, J.M.: Generating Fuzzy Rules by Learning from Examples. IEEE Transactions on Systems, Man, and Cybernetics 22(6), 1414–1427 (1992)
5. Abido, M.A.: A niched Pareto genetic algorithm for multiobjective environmental/economic dispatch. Int. Journal of Electrical Power and Energy Systems 25(2), 97–105 (2003)
6. Li, Y., Ha, M., Wang, X.: Principle and Design of Fuzzy Controller Based on Fuzzy Learning from Examples. In: Proc. of the 1st Int. Conf. on Machine Learning and Cybernetics, vol. (3), pp. 1441–1446 (2002)
7. Pei, Z.: A Formalism to Extract Fuzzy If-Then Rules from Numerical Data Using Genetic Algorithms. In: Int. Symposium on Evolving Fuzzy Systems, pp. 143–147 (2006)
8. Abe, S., Lan, M.: Fuzzy Rules Extraction Directly from Numerical Data for Function Approximation. IEEE Transactions on Systems, Man, and Cybernetics 25(1), 119–129 (1995)
9. Zhao, Y., Collins, E.G., Dunlap, D.: Design of genetic fuzzy parallel parking control systems. In: Proc. American Control Conference, vol. 5, pp. 4107–4112 (2003)
10. de Castro, L.N., Von Zuben, F.J.: Learning and Optimization Using the Clonal Selection Principle. IEEE Transactions on Evolutionary Computation 6(3), 239–251 (2002)
11. Honório, L.M., da Silva, A.M.L., Barbosa, D.A.: A Gradient-Based Artificial Immune System Applied to Optimal Power Flow Problems. In: de Castro, L.N., Von Zuben, F.J., Knidel, H. (eds.) ICARIS 2007. LNCS, vol. 4628, pp. 1–12. Springer, Heidelberg (2007)
12. Honorio, L.M.: Virtual Manufacture Software. Information Technology Institute/UNIFEI, http://www.virtualmanufacturing.unifei.edu.br

A Fuzzy Inference Expert System to Support the Decision of Deploying a Military Naval Unit to a Mission

Giuseppe Aiello, Antonella Certa, and Mario Enea

Dipartimento di Tecnologia Meccanica, Produzione ed Ingegneria Gestionale,
Università degli Studi di Palermo
90128, Palermo, Italy
{aiello,acerta,enea}@dtpm.unipa.it

Abstract. Naval military units are complex systems required to operate in fixed time frames in offshore tasks where maintenance operations are drastically limited. A failure during a mission is a critical event that can drastically influence the mission success. The decision of switching a unit to a mission hence requires complex judgments involving information about the health status of machineries and the environmental conditions. The present procedure aims to support the decision about switching a unit to a mission considering the vagueness and uncertainty of information by means of fuzzy theory and emulates the decision process of a human expert by means of a rule-based inference engine. A numerical application is presented to prove the effectiveness of the approach.

Keywords: fuzzy inference, decision support system, mission reliability.

1 Introduction

Naval military units are complex systems required to operate in fixed time frames in offshore tasks where maintenance operations are drastically limited. The sequence of operations to be performed in a fixed time frame constitutes the mission the ship is engaged for. A failure during a mission is a critical event that in some cases can be handled on board, but that in other cases, when it affects a critical component such as an engine or a steering mechanism, requires the ship to return to a port with suitable repair facilities. Critical failures may hence prevent the ship to return to the port and consequently require the ship to be towed by another vessel. In such situations, the impossibility to perform the mission tasks in the time frame assigned eventually results in a mission failure. The evaluation of the likelihood of a failures is hence an issue of primary interest for commercial and military navies, not only because it prevents the ship to perform its functions, but also because it can be a costly event and a treat for the safety of the crew. In the traditional military, in particular, units must be "operation ready", meaning they must be ready to accomplish the missions they are assigned to. The decision of switching a unit to a mission however involves

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 320–327, 2009.

complex judgments requiring information about the health status of machineries, the available resources and the environmental conditions. The health status of machineries here refers mainly to information about the availability and condition of equipment, resource information concerns the availability and condition of personnel, the types of training available and the training actually received, and environmental information, refers to information about the threat situations and alert conditions in which units must operate, and information about weather and ocean conditions etc. As it frequently happens in decision problems, some of these of information can be quantified, and some are merely descriptive: purely descriptive information may however be very valuable in the decision process.

In the present paper a support tool is presented to assist the decision maker in the decision of deploying a military unit to a mission. A problem that frequently arises when designing a decision support tool is to represent the vagueness and uncertainty that typically affects information which cannot be handled with traditional (crisp) mathematical models. The proposed approach takes into account such vagueness and uncertainty by means of fuzzy sets and emulates the decision process of a human expert by means of a rule-based inference engine. Experts' knowledge may in fact efficiently be represented in the form of rules when fuzzy logic is employed. Rule-based expert systems use human expert knowledge to solve real-world problems that normally would require human intelligence. Fuzzy Inference Systems (FIS) are popular computing frameworks based on the concepts of fuzzy set theory, which have been applied with success in many fields like control [1] [2], decision support [3], system identification, etc. Their success is mainly due to their closeness to human perception and reasoning, as well as to their intuitive handling and simplicity, which are important factors for acceptance and usability of the systems [4].

2 The Fuzzy Rule-Based Expert System for Decision Support

The procedure here presented, aims to be a support tool to make a decision about the switching a unit to a specific mission. As stated before several parameters that influence such decision should be taken into account, this paper however aims at presenting a methodology rather than formalizing the complete decision framework, therefore three representative parameters only have been taken into account and they have been identified by interviewing some captains of military ships. Such parameters are the reliability, the distance from the closest port (in marine miles) and the conditions of the sea.

The reliability of the system involved in a mission is a primary concern since, as stated before, maintenance operations are drastically limited in offshore conditions. In addition the operating conditions of systems and machineries must be considered according to the specific mission profile since only a limited number of machines are required in each mission. For this reason it is preliminarily needed to individuate the ship subsystems (propulsion, power generation, etc..) required to accomplish mission tasks. Moreover, for each subsystem, the critical components must be identified and their reliability must be linked to the

reliability of the entire ship according to the functional relations expressed by the reliability block diagrams (RBD). As mentioned before, the two more input parameters here proposed in addition to the reliability are the (maximum) distance from the closest port which is a key parameter in estimating the likelihood of returning to port in case of failure [5] and the sea conditions.

The FIS is applied to each subsystem by using IF-THEN rules and fuzzy operators to determine the impact of each subsystem on the operational readiness. The minimum of the values thus obtained represents the overall operational readiness of the ship related to a given mission. The minimum operator is chosen to assure a pessimistic assessment of the likelihood to successfully perform mission tasks.

Fig. 1 shows the whole procedure to evaluate the ship operational readiness with relation to a mission.

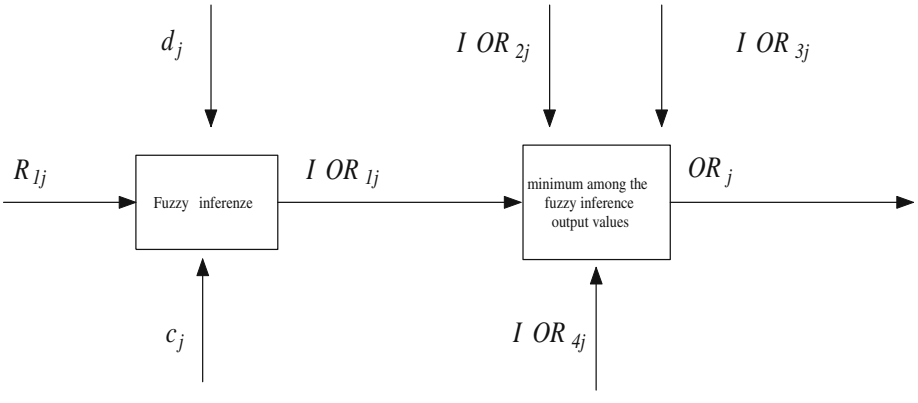


Fig. 1. Proposed procedure block diagram

where
$i = 1, \ldots, I$ is the generic subsystem;
$j = 1, \ldots, J$ represents the mission for which to make the decision;
R_{ij} = subsystem i reliability with relation to the mission j;
d_j = mission distance;
c_j = conditions of the sea;
$I\ OR_{ij}$ = impact of subsystem i on the global score of the likelihood to perform mission j;
OR_j = global score expressing the likelihood of mission j.

As it is well known, a basic fuzzy logic system is constituted by four components: a rules set, a fuzzifier, an inference engine and a defuzzifier. The core of a FIS is its knowledge base, which is expressed in terms of fuzzy rules and allows for approximate reasoning [6]. The fuzzy logic system here used is a Multi Input-Single Output System (MISO), using the Mamdani implication [7] and the center of area method (COA) as defuzzifier. At the first step of the inference process, it is needed to define the fuzzy sets to represent the crisp input values, that is the

fuzzification process, which consists in assigning fuzzy linguistic variables in the universe of discourse of each input value. In particular, in this paper each input parameter is described by triangular and trapezoidal fuzzy numbers. Triangular fuzzy numbers are widely used for their simplicity and solid theoretical basis [8]. The membership function of a triangular fuzzy number A is $\mu_A : R \to [0, 1]$ and it can be represented by the set of equations 1, where $l < m < u$. Consequently, a triangular fuzzy number is fully characterized by three real numbers (l, m, u). The parameter m corresponds to the maximum grade of $\mu_A(x)$ that is equal to 1, l and u are the lower and upper bounds of the definition interval.

$$\mu_A(x) = \begin{cases} \frac{x-1}{m-l} & \text{when } x \in [l, m] \\ \frac{u-x}{u-m} & \text{when } x \in [m, u] \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Analogously, the membership function of a trapezoidal fuzzy number B is $\mu_B : R \to [0, 1]$ and it can be represented by the set of equations 2:

$$\mu_B(x) = \begin{cases} \frac{x-1}{m-l} & \text{when } x \in [l, m] \\ l & \text{when } x \in [m, n] \\ \frac{u-x}{u-n} & \text{when } x \in [n, u] \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $l < m < n < u$.

Similarly, a trapezoidal fuzzy number is fully characterized by four real numbers (l, m, n, u). The parameters m and n give the maximum grade of $\mu_B(x)$. The next step in the fuzzy logic system is to define the possible rules arising from combining the fuzzy inputs. Rules are usually provided by a team of experts in the form of IF-THEN sentences and are introduced into the FIS. Later, since the values of the assessment parameters are crisp, the fuzzifier maps the input crisp numbers into the fuzzy sets to obtain their degrees of membership. The inference engine of the FIS maps the antecedent fuzzy (IF part) sets into consequent fuzzy sets (THEN part) taking into account the rules already stated. The inference process determines the fuzzy subset of the output variable for each rule by using the MIN (Mamdani operator) as implication operator. If more than one rule produces the same consequence, an operator must aggregate the results of these rules. In particular, the MAX operator is used. Finally, the defuzzifier maps the fuzzy output into a crisp number, which becomes the output of the FIS, that is, in the case here considered, the impact of generic subsystem on the ship operational readiness.

As mentioned before in this case the COA method is applied which is the most prevalent of all the defuzzification methods [9], [10]. The Fig. 2 represents the inference process.

3 Numerical Application

The proposed procedure is here applied to a simulated case referred to a military ship. The inference process is carried out by Informs software package Fuzzy Tech.

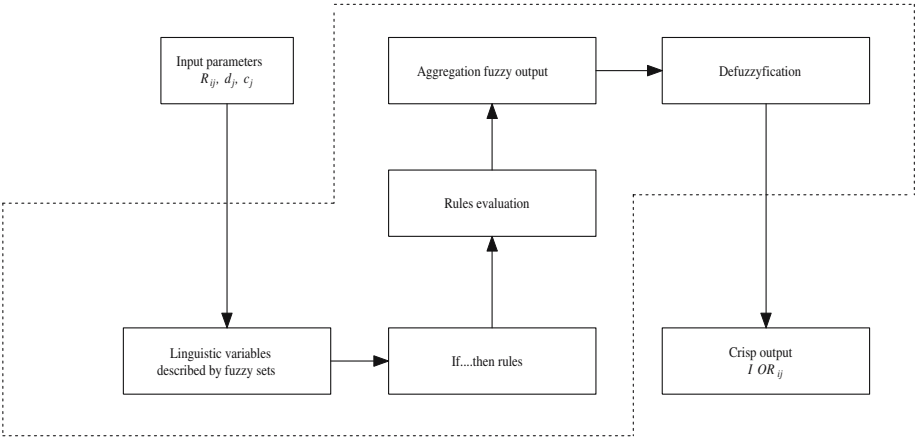


Fig. 2. Block diagram of fuzzy inference procedure

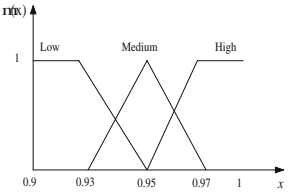


Fig. 3. Subsystem reliability

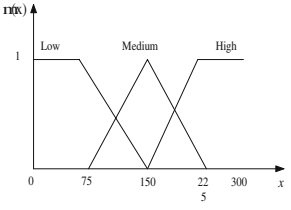


Fig. 4. Mission distance

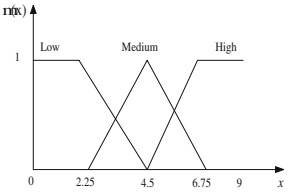


Fig. 5. Sea conditions

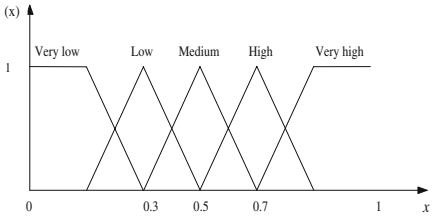


Fig. 6. Impact of each subsystem on the ship operational readiness

Table 1. Subsystem reliability

Subsystem	Reliability
Propulsion	0.98
Power generation	0.96
Command and control	0.94
Weapon	0.93

It is supposed that the ship is constituted by the following subsystems individuated as critical for the mission success: propulsion; power generation; command

Table 2. Other input parameters

Distance	Sea condition
220	5

Table 3. Set of rules

if			then
Reliability	Distance	Sea condition	Impact on ship O.R.
Low	Low	Low	Low
Low	Low	Medium	Very Low
Low	Low	High	Very Low
Low	Medium	Low	Low
Low	Medium	Medium	Very Low
Low	Medium	High	Very Low
Low	High	Low	Low
Low	High	Medium	Very Low
Low	High	High	Very Low
Medium	Low	Low	High
Medium	Low	Medium	High
Medium	Low	High	Medium
Medium	Medium	Low	Medium
Medium	Medium	Medium	Medium
Medium	Medium	High	Low
Medium	High	Low	Medium
Medium	High	Medium	Medium
Medium	High	High	Low
High	Low	Low	Very High
High	Low	Medium	High
High	Low	High	medium
High	Medium	Low	High
High	Medium	Medium	Medium
High	Medium	High	Low
High	High	Low	High
High	High	Medium	Medium
High	High	High	Low

and control and weapon. Such systems may undergo different loading and employment conditions in different mission profiles thus resulting in different reliability values. For example, the propulsion subsystem is constituted by redundant engines and, depending on the speed required to perform the mission, they can be employed in different configurations: the reliability relation is hence different according to the specific stand-by/parallel or series configuration. The conceptual framework here considered is hence constituted by the systems reliability, the distance from the closest port and the conditions of the sea (according to the mission profile). Each input parameter has three linguistic variables (low, medium and high) described by triangular and trapezoidal fuzzy numbers, as shown in the

Table 4. Subsystem impact on mission likelihood

Subsystem	Impact on mission success.
Propulsion	0.4630
Power generation	0.455
Command and control	0.2974
Weapon	0.2029

Fig. 3, 4 and 5. Instead, the output parameter has five linguistic variables (very low, low, medium, high and very high) as shown in the Fig. 6.

The contribution of the generic subsystem i to the likelihood of performing the mission j, OR_{ij} is expressed by values in the range $[0, 1]$ and it can be represented, for example, by a corresponding chromatic scale. The reliability of each subsystem i against the generic mission j, which constitutes an input to the decision system, is given in table 1. The other input data of the mission j are shown in table 2. The set of rules individuated by the experts is given in table 3. The related output values obtained by the inference process are reported in table 4. Thus, in this simulated case, by applying the proposed procedure, that is by taking the minimum value among the output values, the ship operational readiness with relation to a given mission is 0.2029 measured in the range $[0, 1]$.

4 Conclusions

In the present paper the decision of deploying a military naval unit to a mission has been considered. Such decision generally requires a human decision process involving information about the environmental conditions, the operational status of machines etc. Such information can be hardly formalized by means of traditional (crisp) mathematical models, due to its vagueness and uncertainty, whereas such characteristics can be efficiently taken into account using approximate reasoning. In the present paper an expert decision support system based upon a fuzzy inference engine is presented, which allows to take into account experts' experience in the judgments of the likelihood of a military naval unit performing a mission. The mission is described by a specific mission profile which defines the distance from the closest port, the conditions of the sea, and the reliability values of the subsystem involved. The numerical application shows that the methodology presented may efficiently be employed to support the decision maker in the decision process providing a global score expressing the likelihood of the ship to perform the mission tasks, thus confirming the effectiveness of fuzzy inference systems in decision analysis.

References

1. Carlsson, C., Fullér, R.: Fuzzy Reasoning in Decision making and Optimization. Physica-Verlag, New York (2002)
2. Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic-theory and applications. Prentice Hall PTR, New Jersey (1995)

3. Bojadziev, G., Bojadziev, M.: Fuzzy logic for business, finance and management. World Scientific Publishing Co. Pte. Ltd., Singapore (1999)
4. Ross, T.J.: Fuzzy logic with engineering applications. John Wiley & Sons, England (2004)
5. Christer, A.H., Lee, S.K.: Modelling Ship Operational Reliability over a Mission under Regular Inspections Author(s). The Journal of the Operational Research Society 48(7), 688–699 (1997)
6. Czogala, E., Leski, J.: Fuzzy and Neuro-Fuzzy Intelligent System. Physica-Verlag, New York (2000)
7. Mamdani, E.H., Assilian, S.: An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. International Journal Man-Machine Studies 7, 1–13 (1975)
8. Pedrycz, W.: An Why Triangular Membership Functions? Fuzzy Sets and Systems 64(1), 21–30 (1994)
9. Sugeno, M.: An introductory survey of fuzzy control. Inf. Sci. 36, 59–83 (1985)
10. Lee, C.: Fuzzy logic in control systems: fuzzy logic controller. Parts I and II. IEEE Trans. Syst., Man, Cybern. 20, 404–435 (1990)

A Reasoning Methodology for CW-Based Question Answering Systems

Elham S. Khorasani, Shahram Rahimi, and Bidyut Gupta

Computer Science Department,
Southern Illinois University, Carbondale, IL, USA
{elhams,rahimi,gupta}@cs.siu.edu

Abstract. Question Answering Systems or (QA systems for short) are regarded as the next generation of the current search engines. Instead of returning a list of relevant documents, QA systems find the direct answer to the query posed in natural language. The key difficulty in designing such systems is to perform reasoning on natural language knowledgebase. The theory of Computing with Words (CW), proposed by Zadeh, offers a mathematical tool to formally represent and reason with perceptive information. CW views a proposition in natural language as imposing a soft/hard constraint on an attribute and represents it in form of a *generalized constraint*. In this paper we develop a reasoning methodology for the restricted domain CW-based QA systems. This methodology takes, as input, the knowledgebase and the query in form of generalized constraints and organizes the knowledge related to the query in a new tree structure, referred to as a *constraint propagation tree*. The constraint propagation tree generates a plan to find the most relevant answer to the query and allows improving the answer by establishing an information-seeking dialog with user.

Keywords: Reasoning methodology, Question Answering, Computing with Words, Generalized Constraint.

1 Introduction

The current search engine technologies are much limited to pattern matching and are still relied on human effort for providing useful information. Instead of a direct answer to the query, users receive thousands of documents that contain the input keywords and they have to manually process these documents to extract the desired information. The QA systems are regarded as the next generation of the current search engines. They receive a query expressed in natural language, process their knowledge base (KB), which is also in natural language, and return the most relevant answer to the query. Therefore QA systems need more complex natural language processing than other type of information retrieval systems. The key difficulty in designing such systems lies in the imprecise nature of natural language expressions. The theory of Computing with Words [7], which is rooted in fuzzy set and fuzzy logic, provides a mathematical tool to

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 328–335, 2009.

model the imprecision of natural language propositions and perform reasoning among perceptions. CW views a proposition in natural language as imposing a soft/hard constraint on an attribute and represents it in form of a Generalized Constraint (GC).GC provides a basis for approximate reasoning. It serves as a generalized language for representing different kinds of uncertainty such as probability, possibility, truth qualification and so forth. A proposition in natural language may be expressed in GC with the form $GC : X\ isr\ R$, where X is the constraint variable, R is the constraint on the values that X can take and is called the constraining relation, and r is the semantic modality of the constraint that specifies how R is related to X. There are three primary modalities which represent the three primary aspects of uncertainty: probabilistic (r=p), possibilistic (r=*blank*), and veristic (r=v). Other types of constraints can be viewed as the mixture of the primary constraints [7].

For example the proposition: "gas is expensive" can be represented in GC as: "*price(gas) is expensive*". New GCs may be derived from sets of existing GCs by conjunction, projection and propagation operations (For more details on generalized constraint theory refer to [7]).

After representing knowledge in form of a GC, a set of deduction rules need to be defined to perform reasoning. To do so, a GC is then summarized and abstracted into a protoform (PF), abbreviation for prototypical form. Informally a protoform is an abstracted summary of an object and represents the semantic of such objects [6]. For example the GC expression "*price(gas) is expensive*", can be abstracted to protoform:"*A(B) is C*", where A is an abstraction of linguistic variable "*price*", B is an abstraction of "*gas*" and C is an abstraction of the granule value "*young*". The concept of protoform plays an important role in reasoning; it allows classifying the objects of the same semantic structure and defining inference rules for manipulating them. These rules are drawn from various domains such as probability, possibility, fuzzy arithmetic, fuzzy logic and so forth and they basically govern propagation of GCs. Some examples of these rules are listed in table 1. More rules can be found in [7]. Each rule has a symbolic part, which is in terms of protoforms, and a computational part which defines the computation that has to be carried out to arrive at a conclusion. The focus of this paper is to develop a methodology that uses GC propagation rules to make a sequence of inferences on a GC knowledgebase, in order to provide an answer to the input query. Although fuzzy set theory and fuzzy logic are well defined and have been extensively studied in literature, there are not yet many works that extended and utilized CW to develop a working QA system.

Sufyan Beg et.al. [4] designed a hybrid framework for a QA deduction engine that combines the phrase-based deduction with CW reasoning. This framework identifies and tags the query as well as the sentences in KB and extracts the facts that are relevant to the query. If the relevant facts are tagged as a protoform, then they will be processed according to protoform deduction rules. Otherwise the standard bivalent logic reasoning will be applied to find the appropriate answer to the question.Ahmad, et.al. [1] proposed a framework for developing a CW-based fuzzy expert system for automated question answering. The focus of

Table 1. Examples of protoform deduction rules

rule	**symbolic part**	**computational part**
rule (1) interpolation	$X \text{ is } A$ $\sum_i \text{if } x_i \text{ is } A_i \text{ then } Y \text{ is } B_i$ $Y \text{ is } B$	$\mu_B(v) = \sum_i m_i \wedge B_i$ $m_i = sup_u(\mu_A(u) \wedge \mu_{A_i}(u))$, $i = 1, \ldots, n$
rule (2) intersection syllogism	$Q_1 A's \text{ are } B's$ $Q_2 (A\&B)'s \text{ are } C's$ $Q_2 A's \text{ are } (B\&C)'s$	$Q_3 = Q_1 \times Q_2$
rule (3) basic extension principle	$Y = f(X_1, .., X_n)$ $X_i is A_i \quad i = 1, \ldots, n$ Y is B	$\mu_{f(A_1,..,A_n)}(v) =$ $sup_{u_1,\ldots,u_n}(\mu_{A_1}(u_1) \wedge \cdots \wedge \mu_{A_n}(u_n))$ $m_i = sup_u(\mu_A(u) \wedge \mu_{A_i}(u))$, $i = 1, \ldots, n$
rule (4) compositional rule of inference	$X \text{ is } A$ $(x, y) \text{ is } B$ $Y is A \circ B$	$\mu_{A \circ B}(v) = sup_u(\mu_A(u), \mu_B(u, v))$
rule (5) Basic Probability	$prob(X is A) \text{ is } B)$ $prob(X is C) \text{ is } D$	$\mu_D(v) = sup_r(\mu_B(\int_U (\mu_A(u) r(u) d(u)))$ $v = \int_U \mu_c(u) r(u) d(u), \quad \int_u r(u) d(u) = 1$ $r(u) = \text{probability denssity function of } u$

this framework is on using a probabilistic context-free grammar for translating the natural language sentences into GCs and protoforms.

None of the frameworks mentioned above presented a well-defined inference methodology that would be able to address the following issues:

- How to find the set of propositions in the knowledgebase that are relevant to the query?
- What is the inference chain for propagating constraints from a set of relevant propositions to the query?
- How to combine different answers obtained for the query?
- How to improve the quality of the answer obtained for the query?

This paper presents a reasoning methodology that addresses the above issues. The methodology that we propose here organizes the knowledge in a tree structure that we call a *constraint propagation tree* (CP). CPT extracts and organizes the set of relevant propositions in knowledgebase in response to a query. An evaluation algorithm then traverses the tree and propagates the constraints from this set to the query while aggregating different answers obtained for the query. CPT also allows one to identify the missing knowledge and establish a dialog with user when the information in knowledgebase is not enough for providing an answer.

2 The Reasoning Methodology

The reasoning methodology that is presented here takes the GC form of the query and the knowledgebase as input and makes a sequence of inferences to

obtain a direct answer to the query. We assume the availability of a tool that translates the knowledgebase and the query in to generalized constraints.

The query posed to the system may be of various types. Generally a query can be viewed as seeking a value for one or more variables. Given the GC expression "*X is R?*", the query may ask for instantiation of the constraint variable (X) or the constraining relation (R). This view of the query includes a wide range of question types such as factual questions, list questions, definitions, and so forth. Our reasoning methodology instantiates the query variables in two phases: first the information relevant to the query is extracted and organized in a constraint propagation tree. Next the tree is evaluated to find the value for the query variables and combine different values obtained for these variables.

There are two types of relevancy: direct and indirect. Direct relevancy can be assessed by pattern matching while indirect relevancy requires reasoning and deduction on knowledgebase. For example if the query is Q: *"price(gas) is ?"*, and the knowledge base contains the propositions: p1: *"relation(price(gas), production(oil)) is direct"*, and p2: *"production(oil) is low"*, then p1 is directly and p2 is indirectly relevant to the query. Formally a proposition p is directly relevant to the query if it satisfies one the following conditions:

1. p contains the constraint variable and the subject of the query. For example p: *"relation(price(gas),production(oil)) is direct"*, is directly related to the Q: *"price(gas) is ?"*, because it contains the constraint variable of the query *price* as well as its subject *gas*.
2. p contains the constraint variable of the query with a generic subject. For example p: *"if Age(x) is young then risk(BreastCancer(x)) is high*", is directly related to Q: *"risk(BreastCancer(Mary)) is ?"*.

CPT applies the protoform deduction rules in a hierarchical way to extract the propositions that are directly or indirectly relevant to the query and determine how they are related. The root node in CPT represents the input query and the intermediate nodes are sub goals. Each node is connected to its children through a protoform rule, where the parent node represents the consequent and the children represent the antecedents of the rule. A node in CPT is represented by a tuple: (N, GC, E), where:

- N: is an integer that represents the node number.
- GC: is a generalized constraint that has zero or more uninstantiated variables, e.g., *"Age(Mary) is ?R"* or *"if Age(x) is over 40 then risk(breastCancer(x)) is high"*.
- E: indicates the connection between the node and its children. $E = \{(r, \{C\})\}$, where r is the rule number and $\{C\}$ is a set of integers representing a group of immediate child nodes that form the antecedents of r. For example let us assume that a node i has children $\{j, k, m, n\}$ where nodes $\{j, k\}$ and $\{m, p\}$ are connected to their parent by rules a and b respectively. In this case $E = \{(a, \{j, k\}), (b, \{m, p\})\}$.

The following algorithm shows the procedure of generating a CPT.

```
Algoirthm CPTGeneration
Begin
  initialize the root node to the query
  repeat until no new nodes can be created
     let DRS be the set of propositions in KB that are directly
     related to the query.
     If DRS is not empty then
         for each proposition p in DRS
             if p matches with the query then
                instantiate the query variables
             if there is more than one instantiation for a variable then
                if the variable is veristic
                   instantiate it to the disjunction of individual values
              else instantiate it to the conjunction of individual values
             convert p and the query to protoforms: PF(p) and PF(Q)
             for each rule r in the protoform deduction rules:
                if PF(p) matches withthe antecedent of r &
                PF(Q) matches with the consequent of r then
                   create child nodes for the antecedents of r
         set the query to an uninstantiated leaf node
End
```

This algorithm first initializes the root node and extracts the set of propositions that are directly relevant to the query. This set is called directly related set (DRS). Then for each proposition in DRS, if it matches with the query, the query variables will be instantiated accordingly. This is the case where the answer to the query is explicitly stored in knowledgebase. For example, if we are interested to know the age of Mary: Q: *"Age(Mary) is ?R"*, and the knowledgebase contains the proposition *"Age(Mary) is middle-age"*, then we can instantiate R with the fuzzy subset that represents the granule value *"middle-age"*. If there is more than one proposition in DRS that matches with the query, the query variable will be instantiated to the conjunction of individual values. If for the above example KB also includes the proposition *"Age(Mary) is older than 30"*, then R will be instantiated to: *middle-aged* $\wedge$ *older than 30*. If the query variable is veristic [5], it will be instantiated to the disjunction of individual matches with DRS[1].

CPT allows seeking additional information from the user when the information in KB is not enough to answer a query. A leaf node in CPT can be tagged as a missing knowledge if it has at least one un-instantiated variable. Instantiating this variable may or may not be necessary for answering the query; however in the latter case it might improve the quality of answer by providing more constraints and thereby more robust estimates for the query variables.

[1] There are two classes of fuzzy variables verisitc and possibilistic. Possiblistic variables are disjunctive and can take only one value (e.g., *"Age(Mary)"*). In contrast the veristic variables are conjunctive and can take more than one value (e.g., *"big-Countries"*).

The second phase of reasoning is to propagate the constraints from the bottom of CPT to the top while combining different constraints obtained for each node. The propagation and aggregation algorithm is straightforward. It starts with the nodes in the level before the last level and applies the protoform inference rules (table 1) to the appropriate group of children to obtain a constraint for the parent node.If more than one value is obtained for a node variable then it will be instantiated to the conjunction of these constraints, however if the node variable is verisitic, it will be instantiated to the disjunction of the individual values.After instantiation, the value of a variable can be stored and reused for future queries, provided that the information about that variable will not change in the knowledgebase.It is worth noting that the fuzzy set obtained from applying the protoform rules should be normalized before being propagated to its upper level.

2.1 Evaluation of the Methodology

To evaluate our methodology, we applied it to a real world example taken from a web article about causes of breast cancer. Suppose that our knowledgebase consists of the following information:

The average chance that a woman being diagnosed by breast cancer is a function of age. From age 30 through age 39, it is about 0.4 %; from age 40 through age 49, it is about 1.5 %; from age 50 through age 59, it is about 2.5 % , and from age 60 through age 69 it is about 3.5 %. There are some factors that affect the average risk of breast cancer. Alcohol increases the average risk of breast cancer significantly; pregnancy in the age of 30 or before reduces the average risk of breast cancer by about 3 %, and in older women being overweight can increase the average risk of breast cancer slightly.

Suppose also that we have the following information about Mary:

Mary has a son who is about 20. She gave birth to her son when she was in her 20s. Mary is few years younger than Ann who is in her mid 50. Mary consumes about 1400 to 2000 calories a day. And she drinks moderately.

As rules of thumb, we also know that:

Overeating causes being overweight and the age of mother is equal to age of her son plus the age that she gave birth to her son.

Given the above information we are interested to know what Mary's chances of developing a breast cancer are. As mentioned before, the query and knowledgebase must be translated to GCs before the reasoning methodology can be applied. Generally this translation is not unique and depends on the question that is asked. Thus a proposition in knowledgebase can be translated according to all possible questions that may be asked about that proposition. Although this approach guarantees to find an answer for a question, if such answer exists, it can degrade the time performance considerably for large knowledgebases. A better approach is to find the questions that are most likely to be asked and translate the propositions in knowledgebase accordingly. For the purpose of this paper we assume that there exists a tool that performs such translation. By translating the above information to GCs we get:

1. *if age(x) is in 30s then average(risk(bc(x))) is about 0.4 % +If age(x) is in 40s then average(riskbc(x)) is about 1.5 % + If age(x) is in 50s then average(riskbc(x)) is about 2.5 % + If age(x) is in 60s then average(riskbc(x)) is about 3.5 %*

2. *if dirnkhabit(x) is regularly then alcoholFactor(riskbc(x)) is significant*
3. *if age(pregnancy(x)) is about 30 or before then pregnancyFactor(riskbc(x)) is about 3 %*
4. *if age(x) is old and weight(x) is overweight then weightFactor(riskbc(x)) is slightly*
5. *riskbc(x) is average(riskbc(x)) + alcoholFactor(riskbc(x)) + weightFactor(riskbc(x)) - pregnancyFactor(riskbc(x))*
6. *age(son(Mary)) is about 20*
7. *age(Mary) is Age(Ann) few years*
8. *age(Ann) is mid-50*
9. *age(pregnancy(Mary)) is in 20s*
10. *eatingHabit(Mary) is about 1400 to 2000 calories per day*
11. *if eatinghabit(x) is overeat then weight(x) is oveweight*
12. *age(x) = age(son(x)) +age(pregnancy(x))*
13. *drinkhabit(Mary) is moderate*

The CPT of this example is shown in figure 1. After defining appropriate fuzzy sets for the linguistic terms such as: *"about 3%", "old", "significant", "overweight", "mid-50", and so forth*, we calculated and deffuzified the answer as *"Risk(bc(Mary)) is 4 %"*.

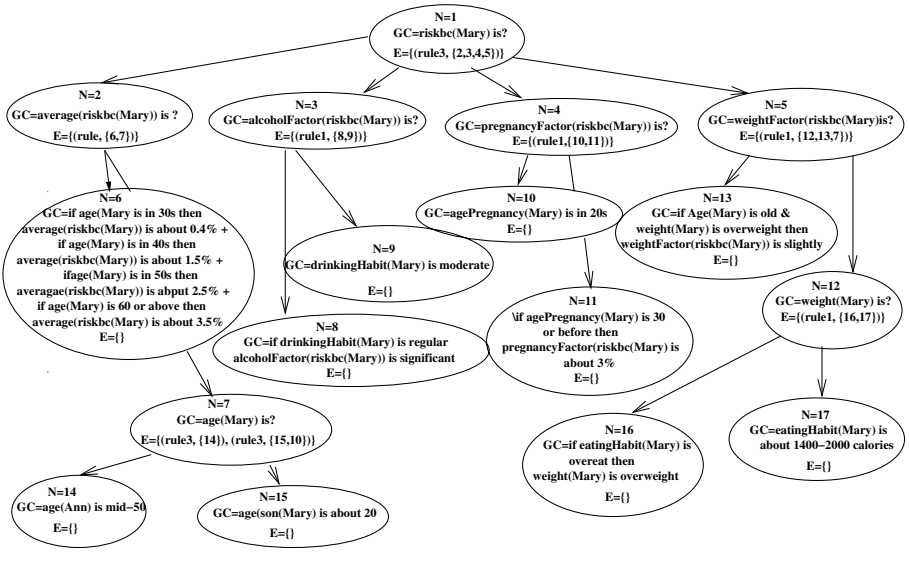


Fig. 1. The CPT of the example. Numbers of the rules are according to those listed in table 1.

3 Discussion and Summary

Current CW-based QA frameworks do not provide a systematic approach for extracting and combining the information in knowledgebase. In this work we developed a methodology that automates the process of inference in a CW-based QA system. The core of the methodology is the generation of a constraint

propagation tree which extracts and organizes the knowledge relevant to the query. CPT also helps to achieve a more robust answer by identifying the missing information in knowledge base in response to a query.

Two issues remain to be addressed as a future work to scale up this methodology to a larger domain knowledgebase such as World Wide Web.

1. *Time performance*. An open domain QA system contains a vast dynamic knowledge source with various types of questions posed to the system. In such systems CPT can be excessively large and it may not be effective to generate and evaluate CPT for each question posed to the system. Thus appropriate techniques should be developed to store data from a previously generated CPT in an indexed database for use in later queries. This data should also be kept updated due to the highly dynamic nature of the web. In order to reduce the size of CPT, the generation algorithm can be modified to stop searching after finding a reasonable answer according to the user expectations.
2. *Commonsense knowledge*. The commonsense knowledge is usually generic, context dependent and uncertain (for example the famous proposition birds can fly). Including commonsense knowledge to the knowledgebase introduces nonmontonicity and adds a great complexity to the reasoning process. Dealing with commonsense knowledge is an open research area and is studied under the name of default reasoning [2,3].

References

1. Ahmad, R., Rahimi, S.: A Perception Based, Domain Specific Expert System for Question- Answering Support. In: WI 2006: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society Press, Los Alamitos (2006)
2. Dubois, D., Prade, H.: Approximate and commonsense reasoning: From theory to practice. In: Michalewicz, M., Raś, Z.W. (eds.) ISMIS 1996. LNCS, vol. 1079, pp. 19–33. Springer, Heidelberg (1996)
3. Schwartz, D.G.: Default reasoning with qualified syllogisms. In: ISUMA 1995: Proceedings of the 3rd International Symposium on Uncertainty Modelling and Analysis, pp. 396–401. IEEE Computer Society, Los Alamitos (1995)
4. Sufyan Beg, M.M., Thint, M., Qin, Z.: PNL-Enhanced Restricted Domain Question Answering System. In: Fuzzy Systems Conference, FUZZ-IEEE 2007, pp. 1–7 (2007)
5. Yager, R.R.: Veristic variables. IEEE Transactions on Systems, Man, and Cybernetics, Part B 30, 71–84 (2000)
6. Zadeh, L.A.: Precisiated natural language (PNL). AI Mag. 25, 74–91 (2004)
7. Zadeh, L.A.: Toward a generalized theory of uncertainty (GTU): an outline. Inf. Sci. Inf. Comput. Sci. 172, 1–40 (2005)

An Intelligent Car Driver for Safe Navigation with Fuzzy Obstacle Avoidance

Francesco M. Raimondi and Ludovico S. Ciancimino

Dipartimento di Ingegneria dellAutomazione e dei Sistemi (D.I.A.S.),
Università degli Studi di Palermo, viale delle Scienze, 90128 Palermo, Italy
{fmraimondi,ludovico.ciancimino}@unipa.it

Abstract. In order to respond effectively to the environment uncertainties, autonomous vehicles are generally equipped with sensors. The proposed car guidance system is equipped with an intelligent controller, based on fuzzy logic, which calculates the speed and wheels orientation in order to follow a path while it is avoiding unknown obstacles. Better fluidity of driving are obtained using future-path, car dimension and car position prevision. Vehicle symmetries also speed-up and simplify the guidance system reducing the inputs and the rules numbers.

Keywords: Fuzzy Logic, Intelligent Guidance, Obstacle Avoidance.

1 Introduction

Autonomous vehicle moves in two main types of environment: simple and obstacle-free or complex and unknown (with moving obstacles). The second case requires Obstacle Avoidance (OA) strategies for safe navigation. The literature [3], [5] also considers the bounded steering wheel orientation angle speed since, if the cruise speed is high, a curvature discontinuity of the path cannot be followed correctly and the path will be lost. In [12], a dynamic window approach provides a local vs. global relationship and is used to store obstacles in memory for later analysis. In [9] a speed-dependent OA by dynamic active regions algorithm is presented. Heuristics along with additional feedback from sensors are used to provide motion and obstacle locations as seen in [6]. Other OA decisions are to use the cell-decomposition methods of VFH* [13] and Sentzs A* algorithm [7]. The probabilistic roadmap planners [8] output is a roadmap graph that connects, as nodes, points (called waypoint W) that are collision-free vehicle placements used to find out the best path that drives the vehicle to the target. In [1] a Continuous Curvature (CC) path planner algorithm for car-like vehicles is presented. The curvature and the sharpness of the path are considered but not the consequent path following problem. In [11] the line-of-sight (LOS) guidance scheme is applied in order to allow waypoint following diminishing the heading error in a finite time interval. In [2], a variable structure control based on fuzzy logic for car-like vehicle is presented. In [5], a fuzzy path-tracker for car-like vehicles is presented. The path to be followed is pre-encoded in order to obtain a matrix of point and angles used to calculate the fuzzy inputs as angle errors. In [10], hybrids control strategy for

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 336–343, 2009.

bounded-curvature vehicle that follow path (made of circular arc or line trajectories) with constant speed is presented.

In this paper, a car Guidance System (GS) with intelligent dynamic OA and a pseudo continuous Curvature Optimizer (CO) is presented. It is based on the Fuzzy Car Driver (FCD) OA controller and a CO that allows the following of paths (generated by an outside roadmap planner) with optimized curvature profile. The FCD takes place, gradually, when there are obstacles nearby otherwise the CO deforms the trajectory generated using the LOS in order to optimize the vehicle trajectory. In order to engage FCD gradually two adjustable wait are used ($wait_{fuzzy}$ and $wait_{LOS}$). These solutions allow efficient decomposed path following since the controller calculates the speed and the steering angle without vehicle parameters knowledge and model dependencies. The vehicle speed is calculated in real-time, instead of constant speed, in order to reject noise, wheels slipping and curvature discontinuity. Considering future positions and distances with respect to the front and the backs of the vehicle gives also better performance with long vehicles.

The developed GS is implemented using the VVM [4] that allows, real-time, realistic vehicle dynamics simulation and real vehicle control. The GS is tested, with some VVM experiments, in order to evaluate its performances in realistic real-time motion control applications.

2 Car-Like Vehicles Models and Sensor

The kinematics model [5], [3] of non-holonomic car-like vehicle (see Fig. 1 a), in the hypothesis of rolling without sliding, can be described as a bicycle-vehicle (with the limitations $|\dot{\phi}| < \dot{\phi}_{Max} = 0.5\text{r/s}$ and $|\phi| < \phi_{Max} = 1.2\text{r}$):

$$\dot{x} = v\cos(\phi)cos(\theta) \qquad \dot{y} = v\cos(\phi)sin(\theta) \qquad \dot{\theta} = v\sin(\phi)/L\ . \tag{1}$$

where (x, y) are the coordinates of the mid-point P of the rear axis, L is the distance between the front and rear axles, θ is its orientation angle, ϕ is the front wheels average orientation with respect to θ and v is the driving speed.

A reliable prediction of the vehicle future position P_{t2} (at time $t2 > t1$) with respect to the last state P_{t1} can be done considering constant ϕ and v. In these hypotheses, P_{t2} will be in position (see Fig. 1 b): A (if $\phi_{t1} = 0$), in B (if $\phi_{t1} > 0$) or in C (if $\phi_{t1} < 0$). The P_{t2} can be calculated as P_t variation ΔP(ΔX, ΔY)

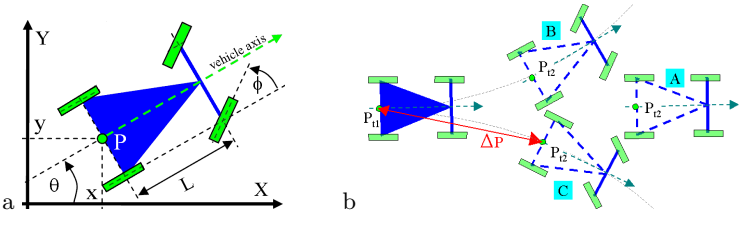


Fig. 1. Car-like vehicle schema and future position prediction

with respect P_{t1} during the time interval $dt = t2 - t1$. The θ variation $\Delta\theta$, from (1), can be written as:

$$\Delta\theta = v * sin(\phi_{t2})/L - v * sin(\phi_{t1})/L \ . \tag{2}$$

From (2) $\Delta\theta \cong 0$, then can be written: $\theta_{t2} \cong \theta_{t1} + \Delta\theta \cong \theta_{t1}$. Then, from (1) ($\Delta$X,$\Delta$Y) can be calculated as (prediction began exactly when t2≅t1):

$$\Delta X = v * cos(\phi_{t1}) * cos(\theta_{t1}) * ds \ . \tag{3}$$

$$\Delta Y = v * cos(\phi_{t1}) * sin(\theta_{t1}) * ds \ . \tag{4}$$

where ds (used to control $|\Delta P|$) depend on dt, v and environment complexity. Since the value of $\cos(\phi_{t1})$, $\sin(\theta_{t1})$ and $\cos(\theta_{t1})$ are known, the computation of (3) and (4) are fast and well suited for real-time computation.

The sensor range is typically a cone (whose amplitude≅ 30^o) centered on its pointing direction which has a fixed angular offset δ with respect to θ. Then an obstacle can be approximated with a point placed (on the pointing direction) at the minimum distance (LateralDist) reported by the sensor (see Fig. 2.a). Its position, detected, as example, by the right (left is analogue) side sensor can be calculated, with respect P_t, as (see Fig. 2.b):

$$\Delta Xo_{Right} = SensorX_R + LateralDist_{Right} * cos(\theta + \delta_R) \ . \tag{5}$$

$$\Delta Yo_{Right} = SensorY_R + LateralDist_{Right} * sin(\theta + \delta_R) \ . \tag{6}$$

where ($SensorX_R$, $SensorY_R$) is the right sensor positions with respect P_t.

Remarks: A) Since the (3), (4) and (5), (6) are increments with respect P_t, for the OA, the P_t given from GPS is not necessary. It is a good feature of the OA since in narrow places the GPS signal can be lost. B) Happens that unrecognized obstacle (because it is out of the front sensor range) collide with the vehicle back. For this motivation, the obstacle data are stored in order to consider also the dimension/length of the vehicle using reliable data generating realistic obstacle position previsions. As example in Fig. 2.b an obstacle detected since t0, at the time t4 is unrecognized by the left sensor. In this case, the last known obstacle position, at t3, can be used to decide were we go on time t5.

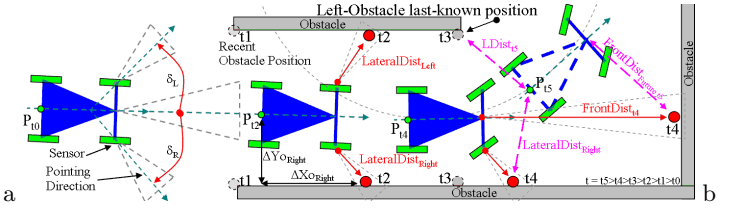


Fig. 2. Vehicle sensor specifications (a) and its employment with obstacle (b)

3 Intelligent Guidance System

The GS must control the vehicle so that it follows the roadmap. If near the vehicle there are obstacles the FCD calculates ϕ (phi) and v (ws) which allows waypoint-following while avoiding obstacles. Better driving capability are obtained processing the distances of the vehicle from the W_i and the obstacles, given by the Distance Computer (DC), in order to obtain simple inputs for the FCD. The GS, at time t, works as follow (see the schema on Fig. 3):

- it calculates the estimated position of obstacles (see Fig. 2);
- it calculates the future position P_{t+dt} of the vehicle (see Fig. 1);
- are calculated the distances, which the sensors would recognize in P_{t+dt};
- it decides which side of the vehicle is more critical;
- the FCD, calculates phi and ws optimizing the future position of the vehicle;
- the LOS calculates the required ϕ_d for the obstacle-free path;
- the curvature are optimized by the CO;

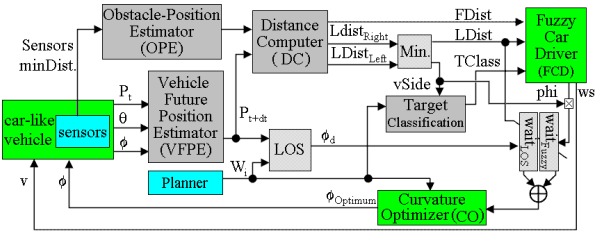


Fig. 3. Guidance System schema

Remark: The time interval dt vary depending on the situation (ex. it increase if there are no obstacles and decrease vice versa. It also depends on the dynamism of the environment (dt is great if the obstacles are slow, and small vice versa. Conversely, if the environment is highly dynamic, dt must be small in order to consider the obstacles stationary.

The LOS guidance scheme [11] calculates the steering ϕ_d required by the vehicle to reach the desired vehicle position $W_i(x_d, y_d)$. The ϕ_d is computed as (see Fig. 4.a):

$$\phi_d = k * (\psi_d - \theta) = k * \Delta\phi \ . \tag{7}$$

where k is a constant and ψ_d is the desiderate angle between W_i and P_t(x, y) (i.e. the desired vehicle direction) given by: $\psi_d = \tan^{-1}$ [(y_d-y) / (x_d-x)].

The classic waypoints following, with arc of circle, produce a path made of constant-curvature elementary trajectory that drives the vehicle from W_{i-1} to W_i. Unfortunately, it does not take in consideration the next-waypoint (W_{i+1}) approach and then the vehicle orientation (near W_i) can be not well suited to reach W_{i+1} (see circular trajectory from W_1 to W_2 of Fig. 4.b). An alternative trajectory that reaches the W_i with an optimal vehicle orientation that minimizes curvature discontinuity can be obtained considering W_{i+1} approach. The

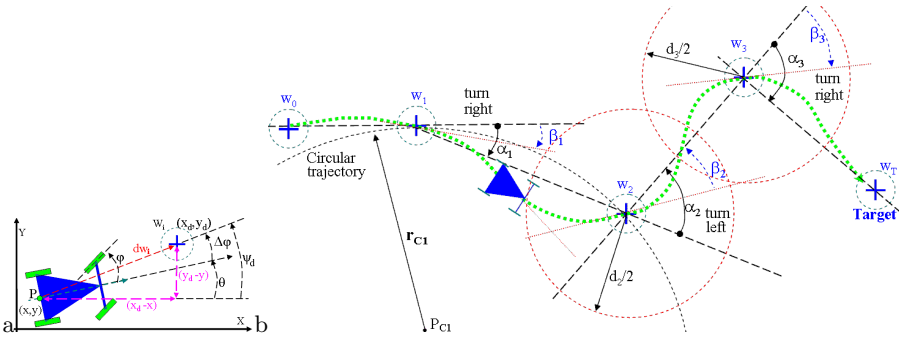
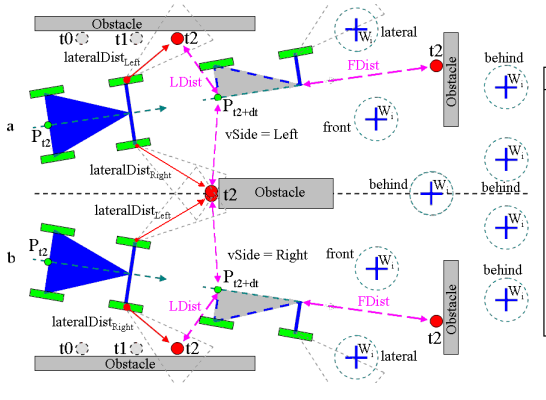


Fig. 4. LOS and CO. The trajectory i=2 (radius r_{C1}) have a bad approach for W_{i+1}.

CO basic idea is: knowing if W_{i+1} requires a steering to the right or left, alter the trajectory of the vehicle so as to facilitate the arrival to W_{i+1} before W_i. The trajectory is altered, near the W_i, by adding to ϕ_d the angle β_i if $dw_i > d_i/2$ or -β_i if $dw_{i+1} < d_i/2$ (where β_i=$\alpha_i/2$, α_i is the angle of rotation led by 3 consecutive W_i, d_i is the distance $W_{i-1} \leftrightarrow W_i$ and dw_i the distance $P_t \leftrightarrow W_i$).

Figure 5 a/b shows two cases in witch the vehicle sensors find obstacles. If the vehicle side (vSide) that has the biggest risk to collide (i.e. the side where there are smaller distances from the obstacles) is known, the OA can be simplified focusing on this side since gives rise to more dangerous situations. Then the GS will avoid the nearest side-obstacles and go where it has a greater margin of movement (i.e. lowest risk). The FCD consider only the half vehicle sensors data of vSide: the front plus the vSide instead of all. Then, with N sensors, the numbers of FCD inputs decrease from N to approximately N/2. Consequently, the rules to be evaluated decrease significantly in number and complexity and their evaluation can be very fast. The vSide and the distance LDist are obtained considering the minimum with a small threshold in order to avoid oscillation:



FCD Rule set

FDist	LDist	TClas	phi	ws
not low	not low	front	front	high
low	not low	front	front	low
not low	low	front	lateral	med
low	low	front	veryLateral	low
not low	not low	lateral	lateral	high
low	not low	lateral	lateral	low
not low	low	lateral	front	med
low	low	lateral	lateral	low
not low	not low	behind	behind	high
low	not low	behind	veryBehind	low
not low	low	behind	behind	med
low	low	behind	veryBehind	low

c

Fig. 5. Future vehicle position, target classification and FCD specification

if (LateralDist$_{Right}$ <LateralDist$_{Left}$ + threshold)$|_{t+dt}$
then vSide = Right and LDist = LateralDist$_{Right}$
else vSide = Left and LDist = LateralDist$_{Left}$

The current target position W_i, with respect to P_{t+dt}, is also considered (see Fig. 5) using the Target-Classification (TClass) variable. W_i position can be: *lateral*, *front* or *behind* (with respect the obstacle) then:

if (vSide=Right) { if (TargeIsOnFront) then TClass = front;
else if (TargeIsOnRight) then TClass = lateral;
else TClass = behind; }
else if (vSide=Left) { if (TargeIsOnFront) then TClass = front;
else if (TargeIsOnRight) then TClass = behind;
else TClass = lateral; }

Using as inputs the distances LDist, FrontDist (FDist) and the TClass variable the FCD calculate the values of phi and ws. Since classical FIS is used the fuzzy rules are in the classical form: if FDist is [ ] and LDist is [ ] and TClas is [ ] then phi is [ ] and ws is [ ]. The chosen FCD inputs/outputs and MFs are:

- FDist: zero, low, med, high.
- LDist: zero, low, med, high.
- TClass: lateral, front, behind.
- phi: veryLateral, lateral, front, behind, veryBehind.
- ws: zero, low, med, high.

Using the knowledge of the car guidance and reasoning as in Fig. 5.a (i.e. lateral=left and behind=right) since the case .b is symmetric a very small rule set can be written (see Fig. 5.c). It is obtained by fusion of many rules in order to speed up the Fuzzy evaluation. After the fuzzy inference, the sigh of phi must be changed in order to agree with vSide: if (vSide=Right) $\Rightarrow phi = -phi$.

4 Vehicle Guidance System Test with VMM

VVM is a client/server modular environment for vehicle dynamics control and simulation. Its great advantage is that the same control system is used both for the simulation and, as it is and with the same software, for the vehicle control. The GS is implemented extending the built-in VVM modules and also using the A.I. module for the FCD. In order to perform tests with a realistic car model, the ODE simulation module is extended with the equation (1) and its limitations.

Two types of test are performed: curvature optimization in simple environment and OA in complex environment. The GS performances are compared with a continuous curvature planner that generates trajectory (tagged as "circular trajectory") between consecutive W_i as a curve having the form of a portion of a circle or straight line.

In Fig. 6, simulated data obtained with VVM, are painted. The vehicle (see lines tagged as "GS trajectory") reaches all waypoints of the path (that can generate curvature discontinuity even if vehicle steering angle and steering speed are bounded) avoiding moving obstacles (see "other vehicles" in Fig.6.b) with a waypoint approach that avoids curvature discontinuity. The GS is then tested

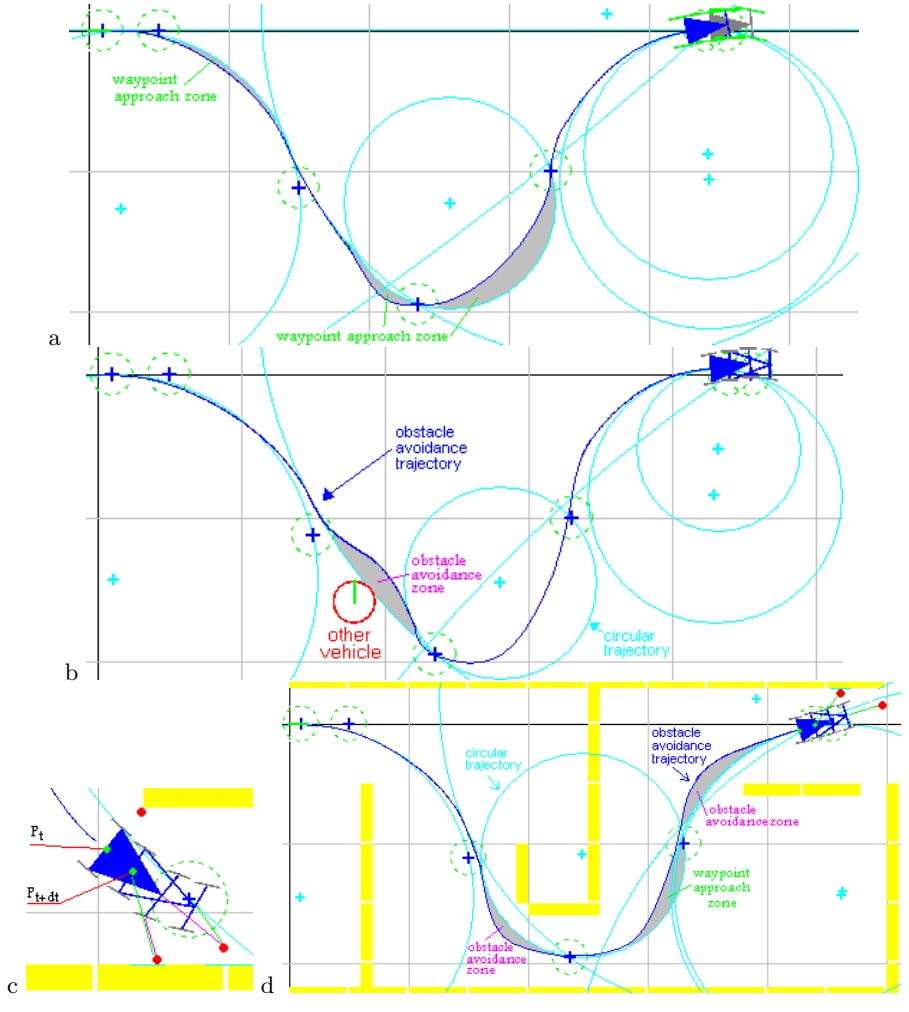


Fig. 6. GS experiments comparing its performance with classical guidance system in different situations

with simple path (see Fig. 6.c) and complex unstructured environment (see Fig. 6.d). These experiments demonstrates the proposed GS and also its VVM implementation performances since the vehicle reaches all the waypoints with good precision avoiding all unknown obstacles.

5 Conclusions

In this paper, a synergic combination of a curvature optimization algorithm and an obstacle avoidance system designed for four wheels car-like vehicles is presented. The curvature optimization system deforms the vehicle trajectory to

obtain a better waypoint approach. The obstacle avoidance is based on a fuzzy controller that use the prevision of the vehicle position and sensor data of the side of the vehicle closed to the obstacle. The GS is also well suited for long vehicle since the obstacle avoidance takes in consideration the front and the rear of the vehicle simultaneously.

References

1. Fraichard, T., Scheuer, A.: From Reeds and Shepp's to Continuous-Curvature Paths. IEEE Trans. on Robotics 20(6), 1025–1035 (2004)
2. Lu, J., Sekhavat, S., Laugier, C.: Fuzzy variable-structure control for nonholonomic vehicle path tracking. In: Proc. IEEE/IEEJ/JSAI Int. Conf. on Intel. Transp. Systems, Tokyo, Japan, pp. 465–470 (1999)
3. Erfu, Y., Dongbing, G., Tsutomu, M., Hu, H.: Nonlinear Tracking Control of a Car-like Mobile Robot via Dynamic Feedback Linearization. In: Proc. of the Control 2004 Conf., Bath, United Kingdom, September 2-9 (2004) ISBN 0 86197
4. Raimondi, F.M., Ciancimino, L.S.: Virtual Vehicles Manager: a java virtual reality environment for distributed multi vehicles dynamics control and simulation. In: Proc. of IEEE Int. Conf. EuroCon 2007, Warsaw, Poland, September 9-12 (2007)
5. Moustris, G., Tzafestas, S.G.: A robust Fuzzy-Logic path tracker for non-holonomic mobile robots. Int. Journal on Art. Int. Tools 14(6), 935–965 (2005)
6. Lindhe, M., Ogren, P., Johansson, K.H.: Flocking with obstacle avoidance: a new distributed coordination algorithm based on Voronoi partitions. In: IEEE Conf. on Robotics and Aut., April 26-May 1 (2004)
7. Stentz, A.: The focused D* algorithm for real-time replanning. In: International Joint Conference on Artificial Intelligence (August 1995)
8. Geraerts, R., Overmars, M.H.: A Comparative Study of Probabilistic Roadmap Planners. In: Proc. Workshop on the Algorithmic Foundations of Robotics (WAFR 2002), pp. 43–57 (2002)
9. Kobialka, H., Becanovic, V.: Speed-dependent obstacle avoidance by dynamic active regions. In: Proc. RoboCup 2003 Symp., Padova, Italy, July 2-11, 2003, pp. 534–542 (2003)
10. Balluchi, A., Bicchi, A., Souéres, P.: Path-following with a bounded-curvature vehicle: a hybrid control approach. Int. Jou. of Control 78(15), 1228–1247 (2005)
11. Guo, J.: A waypoint-tracking controller for a biomimetic autonomous underwater vehicle. Elsevier Ocean Engineering, March 09 (2006)
12. Brock, O., Khatib, O.: High speed navigation using the global dynamic window approach. In: IEEE Int. Conference on Robotics and Automation, May 10-15, 1999, vol. 1, pp. 341–346 (1999)
13. Ulrich, I., Borenstein, J.: VFH*: local obstacle avoidance with lookahead verification. In: IEEE Int. Conf. on Rob. and Aut., San Francisco, CA, April 24-28, 2000, pp. 2505–2511 (2000)

Extending Fuzzy Sets with New Evidence for Improving a Sign Language Recognition System

Christian Vogler[1] and Athena Tocatlidou[2]

[1] Institute for Language and Speech Processing, Athens, Greece
cvogler@ilsp.gr
[2] Informatics Laboratory, Agricultural University of Athens, Athens, Greece
atocat@aua.gr

Abstract. We report an application for extending a fuzzy set with new information to improve recognition rates in a sign language recognition system. The fuzzy sets in the rule base are provided by experts, based on linguistic models of sign languages, which are then extended by fuzzy sets estimated from actual data. The extension algorithm unites an initial knowledge entity and a piece of new information, which is iteratively incorporated until convergence is reached. Experiments show that combining prior information and new evidence improves recognition rates beyond what can be achieved using either body of knowledge by itself.

Keywords: Uncertain knowledge updating, Information fusion, Fuzzy sets.

1 Introduction

Contemporary sign language recognition systems [1,2], based on machine learning approaches, have achieved impressive results, but they suffer from a seemingly irreconcilable conflict between incorporating linguistic knowledge and learning everything from data, as entities learned from data often have no direct linguistic counterpart and cannot easily be manipulated by humans. Unless we solve this problem, it is doubtful that the field of sign language recognition can advance much beyond the current state of the art.

At the core we have the following problem: Given prior knowledge — such as linguistic information —, how can we extend it with additional evidence from data? To this end, in this paper we represent information with fuzzy sets, which have the advantage that the initial body of knowledge can be designed more easily by human experts than it could with other methods for representing uncertainty. Then incorporating information from data becomes equivalent to extending a fuzzy set with new evidence.

Frequently, the extension of prior information is realized through an aggregation procedure involving fuzzy sets. Many proposals for this procedure have appeared, which focus mainly on the mathematical requirements of the operators, and on examining the optimistic-pessimistic character of the process [3,4]. In this paper we take a different approach that focuses on the set-theoretic and pragmatic aspects of extending fuzzy sets [5,6].

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 344–351, 2009.

We demonstrate and test this approach within a sign language recognition system. Such systems provide an ideal test bed, for several reasons: First, we use large amounts of data, making experimental results more representative. Second, there exists a wealth of material on sign languages linguistics, so having human experts devise the initial models and fuzzy sets is a natural approach. Moreover, there are discrepancies between what humans perceive and what happens physically during the execution of signs, so human-devised models should benefit greatly from being refined with information taken from data.

2 Fuzzy Sets

Knowledge representation in the sign language recognition system is based on fuzzy sets [7]; that is, classes of objects in which the transition of membership is gradual rather than abrupt. Thus, if $\mathcal{F}$ is a fuzzy set in a universe $\mathcal{U}$ of discourse, then every member of $\mathcal{U}$ has a grade of membership in $\mathcal{F}$ between 0 and 1, where 0 represents non-membership and 1 represents full membership. $\mathcal{F}$ is characterized by a membership function $\mu^{\mathcal{F}} : \mathcal{U} \rightarrow [0,\ 1]$, which associates with each element $x \in \mathcal{U}$ a number $\mu_{\mathcal{F}}(x) \in [0,\ 1]$.

A discrete fuzzy set is the set defined over a discrete domain. Given a fuzzy set $\mathcal{F}$, its support set is defined as the crisp set

$$S_{\mathcal{F}} := \{x \in \mathcal{U} :\ \mu_{\mathcal{F}}(x) \neq 0\}\,. \tag{1}$$

2.1 Extending Fuzzy Sets with New Evidence

The knowledge updating process enriches the initial fuzzy set with elements from the new evidence, thereby producing a more general fuzzy set. It can be compared with the initial one within a probabilistic framework, since the probability distribution over the initial set of elements is enlarged to include new data. This process manifests itself in a change of the shape of the initial fuzzy set.

More specifically, consider two discrete fuzzy sets $\mathcal{F}$ and $\mathcal{G}$, which represent existing uncertain information and new evidence, respectively. Let $\mathcal{T}$ denote the set of elements in $\mathcal{U}$ that belong to the support set of $\mathcal{G}$, but not to the support set of $\mathcal{F}$:

$$\mathcal{T} := \{x \in \mathcal{U} :\ \mu_{\mathcal{G}}(x) \neq 0 \wedge \mu_{\mathcal{F}}(x) = 0\}\,. \tag{2}$$

The updating algorithm extends $\mathcal{F}$ to a new fuzzy set $\mathcal{H}$, containing all the elements in $\mathcal{F}$ plus elements from $\mathcal{T}$, such that for its support set $S_{\mathcal{H}}$ the following holds:

$$S_{\mathcal{H}} = S_{\mathcal{F}} \cup \mathcal{T}' \text{ where } \oslash \neq \mathcal{T}' \subseteq \mathcal{T}. \tag{3}$$

This process implements a monotonic information gain for the initial set.

Given the fuzzy set $\mathcal{F}$, and the corresponding possibility distribution $\{F_i, m_i\}$, we assign the respective basic probability masses $m_1, m_2, \ldots, m_n$to the focal elements $F_1 \subseteq F_2 \subseteq \ldots \subseteq F_n$ [8]. Similarly, let $G_1 \subseteq G_2 \subseteq \ldots \subseteq G_q$ be the corresponding family of focal elements for the fuzzy set $\mathcal{G}$. For the first focal,

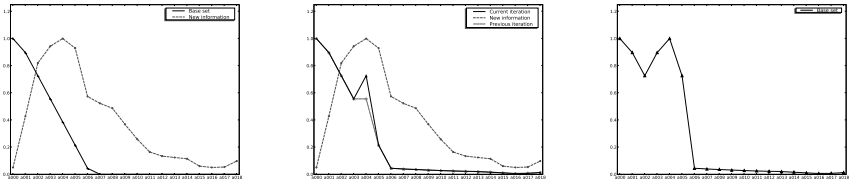


Fig. 1. Updating a fuzzy set from the sign language recognition application with new evidence. Left to right: Human-generated set (solid) and set generated from data (dashed); Snapshot of the updating process; Converged fuzzy set after 4 iterations.

which contains all elements with full membership in $\mathcal{G}$, we can locate a focal element F_k, such that the two sets are disjoint. More formally, there exists an index $k \in \{1, \ldots, n\}$, such that for some F_k in $\{F_1, \ldots, F_n\}$ the following holds:

$$G_1 \cap F_k = \oslash \text{ and } G_1 \cap F_{k+1} \neq \oslash. \tag{4}$$

Subsequently, we create the family of focals $\{F_1, F_2, \ldots, F_k \cup G_1, F_{k+1} \cup G_2, \ldots, F_n \cup G_q\}$ resulting from the union of the focal elements of the fuzzy set $\mathcal{G}$ with the subset of the focals from the fuzzy set $\mathcal{F}$ starting from F_k. To this family we assign the basic probability masses $m_1, \ldots, m_n$ from $\mathcal{F}$.

If $n - k \neq q - 1$; that is, if the families of focal elements are not aligned, either we repeat the final focal element F_n of the fuzzy set $\mathcal{F}$ with its mass m_n accordingly divided, or we repeat the final focal $\mathcal{G}_q$. The outcome is a new fuzzy set $\mathcal{H}$, which represents an extension of $\mathcal{F}$ toward the support set of $\mathcal{G}$ that also preserves the existing information in $\mathcal{F}$. The following holds for the support sets:

$$S_{\mathcal{F}} \subseteq S_{\mathcal{H}} \text{ and } S_{\mathcal{G}} \subseteq S_{\mathcal{H}}. \tag{5}$$

$\mathcal{H}$ represents an information gain over $\mathcal{F}$, induced by the new evidence in $\mathcal{G}$. This process can be continued by repeatedly extending $\mathcal{H}$ (in place of $\mathcal{F}$) with $\mathcal{G}$, until it converges to the point where further updates have no effect (cf. Fig. 1).

The above-defined number $k \in \{1, \ldots, n\}$ indicates the number of permissible repetitions. The number of these iterations can be considered as an indication of the *conceptual distance* of the entities represented by the two fuzzy sets — the greater the number of repetitions, the more dissimilar are the two concepts [5,6].

2.2 Obtaining New Evidence from Data

Given the possibility distribution $\{F_i, m_i\}$ for $\mu_{\mathcal{F}}(x_k)$, we calculate the probability distribution

$$Pr_{\mathcal{F}}(x_k) = \sum_{i=k}^{n} \frac{m_i}{|F_i|}, \tag{6}$$

which with the one-to-one correspondence via the recurrence

$$m_n = |F_n| Pr_{\mathcal{F}}(x_n) \text{ and } m_k = |F_k| \left(Pr_{\mathcal{F}}(x_k) - Pr_{\mathcal{F}}(x_{k+1})\right) \tag{7}$$

shows that $Pr_{\mathcal{F}}(x_k)$ is a dual representation of the membership function $\mu_{\mathcal{F}}(x_k)$. Given a list of data points $\mathcal{P} = \{p_1, \ldots, p_n | p_i \in \mathcal{U}\}$, we estimate its representative fuzzy set by calculating $Pr_{\mathcal{F}}$, via counting how often each x_k occurs in $\mathcal{P}$:

$$Pr_{\mathcal{F}}(x_k) = \frac{|\{p_j \in \mathcal{P} | p_j = x_k\}| \, w_k}{\sum_{i=1}^{n} |\{p_l \in \mathcal{P} | p_l = x_i\}| \, w_i}, \tag{8}$$

where $w_i = \frac{1}{f_i}$ is a weight that normalizes each estimate by the overall frequency f_i with which x_i occurs in $\mathcal{U}$. An example of a human-generated fuzzy set, new evidence from sign language data, and the generalization toward a new fuzzy set is shown in Fig. 1.

2.3 Reasoning with Support Intervals

In the sign language recognition application, the fuzzy sets are applied as part of a rule-based reasoning system implemented in the programming language Fril [9]. Formally, the format of each rule is

((property of_X **is** f) **if**
(feature_1 of_X **is** f_1) **weight** w_1 **and**
...
(feature_n of_X **is** f_n) **weight** w_n): **Pos** $[n_1, p_1]$, **Neg** $[n_2, p_2]$.

The intervals after the keywords **Pos** and **Neg** are *probability intervals* that describe the likelihood of the rule and its negation, respectively. For any specific case on which the rule is evaluated, the degree to which the case satisfies the prototypical concept represented by the rule, is calculated in three steps:

Each condition of the rule is matched against the relevant data, yielding a partial probability interval for each respective characteristic. These are combined into a collective interval for the entire rule through a weighted sum. Finally, the collective interval is applied to the **Pos** and **Neg** intervals in the definition, giving the final probability interval for the satisfaction of the rule.

More formally, during the evaluation of the rule, a case's characteristic is matched against the relevant rule, thus giving a partial probability interval, *support pair*, $[\alpha_i, \beta_i]$ [9]. In this way, the rule body (set of conditions) is assigned a probability interval, computed from the formula

$$[\alpha, \beta] = \left[S\left(\sum_{i=1}^{n} w_i \alpha_i\right), S\left(\sum_{i=1}^{n} w_i \beta_i\right) \right], \tag{9}$$

where S is a user-defined function, $S : [0, 1] \rightarrow [0, 1]$, which acts as a filter. This allows the combination of the conditions of the rule to take intermediate values; that is, values between those of their disjunction and conjunction.

The degree of rule satisfaction is indicated by a final interval (support pair) $[\gamma_1, \gamma_2]$ that denotes the probability for the rule to be true, and is calculated as

$$\gamma_1 = \begin{cases} n_1\beta + n_2(1-\beta) \text{ if } n_1 \leq n_2 \\ n_1\alpha + n_2(1-\alpha) \text{ if } n_1 > n_2 \end{cases}, \quad \gamma_2 = \begin{cases} p_1\alpha + p_2(1-\alpha) \text{ if } p_1 \leq p_2 \\ p_1\beta + p_2(1-\beta) \text{ if } p_1 > p_2 \end{cases}. \tag{10}$$

(a) Extended **(b)** Bent **(c)** Spread **(d)** Opposed

Fig. 2. Examples of hand configuration primitives. A handshape consists of the composition of multiple primitives.

3 Application: Sign Language Recognition System

We now describe a sign language recognition system to test the above algorithm. The task is to recognize sequences of handshapes from continuous sentences.

3.1 Hand Configuration Model

The hand configurations are recognized from their underlying joint angle representation. Note that, in principle, a similar approach can be taken when joint angles are replaced with other features, such as the ones extracted from video.

Our underlying assumption is that, although joint angles are continuous, a small set of discrete finger primitives is sufficient for describing the full range of hand configurations. This assumption is supported by linguistic research [10].

Hand Configuration Primitives. The primitives belong to four orthogonal groups that describe the finger states, the finger and thumb abductions (spread), and the thumb rotations (Fig. 2). A complete hand configuration composes five finger states, three finger abductions, and thumb abduction and rotation.

Each primitive is classified through the values of the involved joint angles, which in turn are represented by discrete labels, each of which corresponds to a continuous fuzzy set. As our generalization algorithm requires discrete fuzzy sets (Sec. 2.1), we quantize the joint angles (which fall between $[-\frac{\pi}{4}, \pi]$) into a number of clusters, ranging from 8 to 22, depending on the affected joint.

In order to classify a primitive, we subject the joint angle measurements and corresponding fuzzy labels to a set of rules. Consider the rule for a bent finger:

((finger X **is bent**)
(metacarpophalangeal joint_of X **is** *extended_or_bent*) **and**
(proximal interphalangeal joint_of X **is** *bent*) **and**
(distal interphalangeal joint_of X **is** *bent*)) : **Pos** $[n_1, p_1]$, **Neg** $[n_2, p_2]$

where *extended_or_bent* and *bent* are fuzzy sets. Applying these rules to each data frame yields the support intervals that measure the possibility of each respective primitive occurring at that particular frame. In summary, constructing the hand configuration in terms of primitives based on fuzzy joint angle labels allows us to deal with uncertain or noisy joint angles.

Hand Configuration Rules. The next layer of rules expresses hand configurations in terms of finger states, with weighted clauses, as discussed in Sec. 2.3. The weights, at present, need to be assigned by a human expert, chosen such that they favor the most prominent and most discriminative features of a handshape; for example, non-curled fingers have larger weights than curled fingers:

$((X$ **is_a 1-handshape**)
(index finger_of X **is extended**) **weight** 0.3
(middle finger_of X **is curled**) **weight** 0.15
$\ldots$: **Pos** $[n_1, p_1]$

3.2 Sequences of Handshapes

The rules described in the previous sections calculate the support intervals of all possible handshapes for each frame. Before making a decision on each handshape in each frame, the recognizer has to apply any constraints on the possible sequences of handshapes.

For this paper we implemented only a constraint that aims at eliminating fast oscillations between handshapes. To this end, we let $Pr(H_q) = \frac{\alpha_q + \beta_q}{2}$ where α_q and β_q are the endpoints of the support interval associated with the handshape. Then we calculate the sequence of handshapes of length T :

$$H_1 H_2 \ldots H_n := \arg\max_{H_q} \prod_{t=1}^{T} Pr(H_{q,t}|H_{q,t-1}) \text{ where} \tag{11}$$

$$Pr(H_{q,t}|H_{q,t-1}) = \begin{cases} Pr(H_{q,t}) \text{ if } H_{q,t} = H_{q,t-1} \\ \xi\, Pr(H_{q,t}) \text{ otherwise} \end{cases}, \tag{12}$$

where ξ is a probability less than one that penalizes a change of handshapes.

4 Experiments and Results

The purpose of the experiments was to compare the recognition accuracy using models provided by a human expert, models purely generated from data, and the human-generated models augmented with evidence from data.

4.1 Setup

The data set consisted of 499 sentences, between 2 and 7 signs long, and a total of 1604 signs from the 22-sign vocabulary that was used by Vogler and Metaxas [11]. These data were collected from the right hand with a Virtual Technologies Cyberglove$^{\text{TM}}$, which records the joint and abduction angles of the fingers, at 60 frames per second.

To evaluate the proposed approach, we followed Vogler and Metaxas [2] and split the data into 400 training sentences and 99 test sentences. No part of the test set was used for obtaining the fuzzy sets with the new evidence; conversely, no part of the training set was used for the recognition experiments. The data were labeled, which allowed us to assign each frame to the correct respective fuzzy sets, and then to calculate their probability duals (cf. Sec. 2.2).

Table 1. Recognition results. H, D, S, I, N denote the number of correctly recognized handshapes, deletion errors, substitution errors, insertion errors, and total number of handshapes in the data set, respectively.

Type of experiment	Correct	Accuracy	Details
Sets designed by human	57.78%	36.68%	H=219, D=58, S=102, I=80, N=379
Sets derived from data	52.54%	37.20%	H=198, D=64, S=117, I=57, N=379
Sets designed by human extended with sets derived from data	63.32%	39.05%	H=240, D=51, S=88, I=92, N=379

4.2 Evaluation Criteria

The main evaluation criterion was the percentage of correctly recognized handshapes. We distinguish among three different types of recognition errors. Assume that the correct sequence of handshapes is denoted by *A B C*. Then a substitution error, such as *A S C,* consists of a confusion of one handshape with another. An insertion error, such as *A B I C,* consists of the insertion of an incorrect handshape into a sequence of correct ones. A deletion error, such as *A C,* consists of the deletion of a handshape from a correct sequence.

The *word accuracy* constitutes the overall percentage of handshapes that the recognizer handles correctly. Let $H = N - S - D$, where H is the number of correctly spotted handshapes, N is the total number of handshapes in the test set, S and D are the number of substitution errors and deletion errors, respectively. Then $Acc = H - I$, where I is the number of insertion errors.

4.3 Results

The experimental results are given in Table 1. Overall, they show that in this application, enhancing the fuzzy sets provided by human experts with the ones derived from the actual data does better than either method by itself. At the same time, recognition rates are not yet competitive with pure machine learning-based methods [11,1].

The recognition rates of the experiment with sets derived from data, in the second row of the table, suggest that the main problem lies not so much with the data or the extension algorithm, but rather with the chosen representation of the handshapes in terms of primitives. Therefore, the most promising step toward improving results likely consists of changing or replacing the primitives for each handshape, such that the linguistics-based conceptual description of a handshape matches the actual data better.

5 Conclusions

The algorithm reported in this paper exhibits some interesting properties: It implements information extension as an aggregation of fuzzy information, thereby

preserving existing information. Furthermore, it is based on set theory, which provides for a certain robustness in its behavior; namely, avoiding point-wise updating and the problems caused by outliers. Finally, it is realized through a gradual process, which adds flexibility to the decision maker's choices.

All these properties make it ideal for combining expert knowledge with information derived from data. If results can be improved to the point where they become competitive with pure machine machine learning-based approaches, it will mean a giant step forward for the entire field of sign language recognition, which currently suffers from a severe lack of integration of linguistic knowledge into recognition systems.

Acknowledgments

This work has been partially funded by the European Union, in the framework of the Greek national project DIANOEMA (GSRT, M3.3, id 35).

References

1. von Agris, U., Zieren, J., Canzler, U., Bauer, B., Kraiss, K.F.: Recent developments in visual sign language recognition. Springer Universal Access in the Information Society 6(4), 323–362 (2008)
2. Vogler, C., Metaxas, D.: A framework for recognizing the simultaneous aspects of American Sign Language. Computer Vision and Image Understanding (CVIU) 81, 358–384 (2001)
3. Dubois, D., Prade, H.: A review of fuzzy set aggregation connectives. Information Sciences 36, 85–121 (1985)
4. Zimmerman, H., Zysno, P.: Latent connectives in human decision making. Fuzzy Sets and Systems 4(1), 37–51 (1980)
5. Tocatlidou, A.: Learning based similarity for fuzzy sets. International Journal of Intelligent Systems 13, 193–220 (1998)
6. Tocatlidou, A., Ruan, D., Kaloudis, S., Lorentzos, N.: Uncertain knowledge association through information gain. In: Ruan, D., Chen, G., Kerre, E., Wets, G. (eds.) Intelligent Data Mining: Techniques and Applications. Springer, Heidelberg (2005)
7. Zadeh, L.: Fuzzy sets. Information and Control 8, 338–353 (1965)
8. Klir, G., Yuan, B.: Fuzzy Sets and Fuzzy logic: Theory and Applications. Prentice Hall, Upper Saddle River (1995)
9. Baldwin, J., Martin, T., Pilsworth, B.: FRIL Fuzzy and Evidential Reasoning in A.I. Research Studies Press (1995)
10. Sandler, W.: Representing handshapes. In: Edmondson, W., Wilbur, R. (eds.) International Review of Sign Linguistics, vol. 1, pp. 115–158. Lawrence Erlbaum Associates, Inc., Mahwah (1996)
11. Vogler, C., Metaxas, D.: Handshapes and movements: Multiple-channel ASL recognition. In: Volpe, G., et al. (eds.) GW 2003. LNCS, vol. 2915, pp. 247–258. Springer, Heidelberg (2004)

General Fuzzy Answer Set Programs

Jeroen Janssen[1,⋆], Steven Schockaert[2,⋆⋆], Dirk Vermeir[1], and Martine De Cock[2]

[1] Dept. of Computer Science, Vrije Universiteit Brussel
{jeroen.janssen,dvermeir}@vub.ac.be
[2] Dept. of Applied Mathematics and Computer Science, Universiteit Gent
{steven.schockaert,martine.decock}@ugent.be

Abstract. A number of generalizations of answer set programming have been proposed in the literature to deal with vagueness, uncertainty, and partial rule satisfaction. We introduce a unifying framework that entails most of the existing approaches to fuzzy answer set programming. In this framework, rule bodies are defined using arbitrary fuzzy connectives with monotone partial mappings. As an approximation of full answer sets, k–answer sets are introduced to deal with conflicting information, yielding a flexible framework that encompasses, among others, existing work on valued constraint satisfaction and answer set optimization.

Keywords: Answer Set Programs, Fuzzy Logic, Valued Constraint Satisfaction.

1 Introduction

Answer set programming [1] (ASP) is a form of non-monotonic reasoning which is based on the stable model semantics for logic programming [2]. Intuitively, in calculating answer sets of a logic program (rule base), we are interested in what can be derived from given facts by applying rules (forward chaining). This corresponds to a form of skeptical reasoning, where we are only interested in well-motivated models. When there are no occurrences of negation-as-failure in the rules, there is exactly one answer set, which corresponds to the unique minimal model of the rule base. In general, with negation-as-failure, there may be several answer sets which are defined using stable model semantics.

Various extensions of ASP have been proposed to deal with different facets of imperfect information, most notably probabilistic extensions to deal with uncertainty and fuzzy extensions to deal with vagueness. Consider, for example, the following rules:

$$r_1 : bad\_weather \leftarrow rainy$$
$$r_2 : bad\_weather \leftarrow \sim sunshine$$
$$r_3 : bbq \leftarrow \sim bad\_weather \wedge hungry$$
$$r_4 : sunshine \leftarrow true$$

⋆ Funded by a joint Research Foundation–Flanders (FWO) project.
⋆⋆ Research Assistant of the Research Foundation - Flanders (FWO).

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 352–359, 2009.

where $\sim$ is used to denote negation-as-failure. These rules encode that we would like to have a barbecue, provided that the weather is not bad and we are hungry. Clearly, concepts such as rainy (dry – drizzle – sprinkle – shower – downpour), sunshine (open sky – partially clouded – overcast) and hungry can be a matter of degree. This can be taken into account, by allowing propositions to be true to a certain degree in $[0, 1]$, interpreting logical connectives using fuzzy logic operators, and generalizing the notion of an answer set in an appropriate way [3,4,5,6]. To indicate that it is partially cloudy, the fact r_4 could then, for instance, be replaced by

$$r_4 : sunshine \leftarrow 0.6$$

Most existing approaches are limited to standard fuzzy connectives such as t–norms, negators, and t–conorms. This, however, severely limits the expressiveness of the approach, and often more subtle aggregation strategies are desired.

A second problem with existing approaches concerns the treatment of partial rule satisfaction. While ideally all rules should be completely satisfied, in practice, available knowledge is often inconsistent (overconstrained). In this case, we are interested in models that satisfy the available rules to the maximal extent possible. One possibility, which is, among others, adopted in [4,7], is to assign a score to each of the rules, indicating to what extent they should *at least* be satisfied. This, however, introduces the problem of choosing optimal values for these weights, which may be far from trivial, limiting the flexibility thus obtained.

In this paper, we propose a generalization of existing work on fuzzy ASP, addressing the aforementioned shortcomings. Specifically, we define an answer set semantics for rules whose body may contain arbitrary fuzzy connectives with monotone partial mappings. To cope with the problems of using weights for partial rule satisfaction, we propose a solution which uses variables, rather than constants, as rule weights, combined with an aggregator expression defining a preference ordering on solutions. Furthermore we introduce the notion of a k–answer set, which resembles a similar concept from the field of valued constraint satisfaction problems (VCSPs) [8].

2 General Fuzzy Programs

Definitions. In the following, let $(\mathcal{L}, \leq)$ be a bounded, complete lattice of truth values, and let $(\wedge, \rightarrow)$ be a residual pair on $\mathcal{L}$, i.e. $\wedge$ is a t–norm and $\rightarrow$ an implicator such that $a \wedge b \leq c \Leftrightarrow a \leq b \rightarrow c$ for all a, b, and c in $\mathcal{L}$. A function $f : \mathcal{L}^n \mapsto \mathcal{L}$ is called acceptable iff for all $1 \leq i \leq n$, f is monotonically increasing or decreasing in its i^{th} argument. We will consider rules over a signature $S = \langle \mathcal{L}, F, \mathcal{V} \rangle$, where $F = \bigcup_{0 \leq i \leq n} F_i$ is a finite set of acceptable functions such that $\forall 0 \leq i \leq n \cdot \forall f \in F_i \cdot f : \mathcal{L}^n \mapsto \mathcal{L}$, and $\mathcal{V}$ is a countable set of *propositions* (or variables). We require F_2 to contain at least the operators $\wedge$ and $\rightarrow$, but it may also contain other residual pairs, as well as other types of connectives. Throughout this paper, we will use $\wedge_m$, $\wedge_p$ and $\wedge_l$ to denote the minimum, product, and Łukasiewicz t–norms, when $\mathcal{L} = [0, 1]$, and $\rightarrow_g$, $\rightarrow_p$ and $\rightarrow_l$ to denote their residual implicators.

A V-*valuation*, $V \subseteq \mathcal{V}$, is an $\mathcal{L}$-fuzzy set on V, i.e. a $V \mapsto \mathcal{L}$ mapping. The set $\mathcal{E}(S)$ of *expressions* over the signature S is defined as usual. A subexpression e' of e is said to *occur positively* in e if either $e' = e$ or e' occurs positively in an increasing argument, or negatively in a decreasing argument of $e = f(\ldots)$. Negative occurrences of subexpressions are defined similarly. The *Herbrand base* $\mathcal{B}_e$ of an expression $e \in \mathcal{E}(S)$ is the set of propositions appearing in e; if all propositions from $\mathcal{B}_e$ appear only positively in e then e is called *positive*. A $\mathcal{B}_e$-valuation is called an *interpretation* of e. The *value* $[e]_I \in \mathcal{L}$ of an expression e w.r.t. an interpretation I is defined by:

$$[e]_I = \begin{cases} I(e) & \text{if } e \in \mathcal{B}_e \\ e & \text{if } e \in F_0 \\ f([e_1]_I, \ldots, [e_m]_I) & \text{if } e = f(e_1, \ldots, e_m) \end{cases}$$

Rules. General fuzzy programs will consist of *rules*, which are expressions of the form $r = a \leftarrow \alpha$, with $\leftarrow$ a residual implicator, $a \in \mathcal{L} \cup \mathcal{V}$ and $\alpha \in \mathcal{E}(S)$[1]. We will refer to a as the head (consequent) of the rule and to α as the body (antecedent); we use $\wedge_r$ to denote the t-norm forming a residual pair with $\leftarrow$. Note that the head of a rule can either be a variable or a constant. In the latter case, the rule is often called a *constraint*. As rules are expressions, the interpretation of a rule is the interpretation of the corresponding expression, as defined above. We extend this to sets of rules and define an interpretation of a set of rules $\{r_1, \ldots, r_n\}$ (commonly referred to as a *rulebase*) as being a $\mathcal{B}_{r_1} \cup \ldots \cup \mathcal{B}_{r_n}$-valuation.

Aggregator Expressions. While we are, in principle, mainly interested in interpretations that satisfy all of the available rules, practical considerations may lead to weaker requirements. Time restrictions, for instance, may cause us to prefer interpretations that satisfy the rules to a suboptimal degree, if they can be found significantly faster. Second, and perhaps more fundamentally, when there exist no perfect interpretations, we may still be interested in interpretations that satisfy the rules to the best extent possible.

For example, consider a weighted graph coloring problem, where edges of a graph are given a weight in $[0, 1]$ indicating how important it is that the nodes they connect are given a different color. Let e_{ij} be the weight of the edge connecting nodes i and j; for each color c and each node i, we introduce a variable c_i to encode whether i is given color c. Let $\mathcal{R}$ be given by (n being the number of

[1] Note that we only consider residual implicators to implement rules. However, S–implicators, for instance, could easily be added as syntactic sugar. Indeed, for any De Morgan triplet $(\wedge, \vee, \sim)$, where $\sim$ is an involutive negator, we can show that ($k \in \mathcal{L}$)

$$[a \leftarrow_s \alpha]_I \geq k \Leftrightarrow [a \leftarrow_r \sim (\sim k \leftarrow_r \alpha)]_I \geq 1$$

where $\leftarrow_s$ and $\leftarrow_r$ are the S–implicator and residual implicator induced by $\wedge$ and $\sim$, respectively. This observation essentially means that we get S–implicators for free when we have residual implicators.

nodes, and C the set of all available colors for which an arbitrary total ordering $\leq$ is defined):

$$\begin{aligned}\mathcal{R} =&\{r_{ci} : c_i \leftarrow_l 1 | 1 \leq i \leq n, c \in C\}\\ &\cup \{l_i : 0 \leftarrow_l 1 - \mathit{atLeastOne}(\{c_i | c \in C\}) | 1 \leq i \leq n\}\\ &\cup \{ex_{icc'} : 0 \leftarrow_l c_i \wedge_m c'_i | 1 \leq i \leq n, (c, c') \in C^2, c < c'\}\\ &\cup \{a_{cij} : 0 \leftarrow_l c_i \wedge_l c_j \wedge_l e_{ij} | 1 \leq i < j \leq n, c \in C\}\end{aligned}$$

where, for $A \subseteq [0, 1]$, $\mathit{atLeastOne}(A)$ is true iff $\sum_{a \in A} a \geq 1$. Obviously this function is monotonically increasing and therefore also acceptable. Note that in this particular example, due to the rules l_i and $ex_{icc'}$, each variable c_i is only allowed to take values from $\{0, 1\}$. We now want to be able to either find the optimal (but not necessarily perfect) graph coloring, or the optimal solution that can be found in a given time frame. To denote the appropriateness of a solution, we use the concept of an *aggregator expression.*

Formally, we define an aggregator expression as a positive expression over a signature $\mathcal{S}_\mathcal{A} = \langle \mathcal{P}, \mathcal{F}_\mathcal{A}, \mathcal{R}_\mathcal{A} \rangle$, where $\mathcal{P} = (O_A, \leq_{O_A})$ is a partially ordered set (preference ordering), $\mathcal{F}_\mathcal{A}$ is a set of operators on O_A, and $\mathcal{R}_\mathcal{A}$ is a set of variables which are called rule propositions. When there is no cause for confusion, we will usually write $\leq$ instead of $\leq_{O_A}$.

In the graph coloring example, one possible choice of aggregator is

$$\mathcal{A} = (\prod_i l_i) \cdot (\prod_{i,c,c'} ex_{icc'}) \cdot (\sum_{c,i,j} a_{cij})$$

where we do not care at all to what degree the choice rules r_{ci} are satisfied. Furthermore we insist that the hard constraints l_i and $ex_{icc'}$ are all satisfied to degree 1. Indeed, these rules can only be satisfied to degree 0 and 1. As soon as one of them is satisfied to degree 0, $\mathcal{A}$ evaluates to 0. Finally, according to $\mathcal{A}$, the best solution is the one that maximizes the sum of the degrees to which the rules a_{cij} are satisfied. Note that in this case, $\mathcal{P} = ([0, +\infty[, \leq)$, and in particular, our preference ordering is total. In general, a large number of alternative strategies to select Pareto–optimal solutions are available, only some of which induce preference orderings that are total. Such strategies have been extensively studied in the context of valued constraint satisfaction problems; see, for instance, [8,9].

A *(general fuzzy) program* (short: **gfasp**) P is then defined as a tuple $P = \langle \mathcal{A}, \mathcal{R}, \phi, \gamma \rangle$, where $\mathcal{A}$ is an aggregator expression over the signature $\mathcal{S}_\mathcal{A}$, $\mathcal{R}$ is a set of rules, ϕ a bijective mapping from the rule propositions in $\mathcal{R}_\mathcal{A}$ to rules from $\mathcal{R}$ and γ an order-preserving mapping from $(\mathcal{L}, \leq)$ to $(O_A, \leq)$. When the mappings ϕ and γ are clear from the context, we will usually omit them. In particular, we write $r : a \leftarrow \alpha$ to denote that the rule $a \leftarrow \alpha$ corresponds to rule proposition r. Furthermore, when the aggregator expression $\mathcal{A}$ corresponds to the minimum of all rule propositions, we will usually identify a program with its set of rules $\mathcal{R}$. The *Herbrand Base* of such a gfasp–program is denoted as $\mathcal{B}_P$ and defined as $\mathcal{B}_P = \bigcup_{r \in \mathcal{R}} \mathcal{B}_r$.

A mapping ρ from $\mathcal{R}$ to $\mathcal{L}$ is called a *rule interpretation* of a program $P = \langle \mathcal{A}, \mathcal{R}, \phi, \gamma \rangle$. Note that any interpretation I of the rule base $\mathcal{R}$ induces a rule interpretation ρ_I, defined by $\forall r \in \mathcal{R} \cdot \rho_I(r) = [r]_I$. For the ease of presentation, we will also write ρ to denote the induced mapping from $\mathcal{R}_\mathcal{A}$ to O_A, i.e. we define $\forall r \in \mathcal{R}_\mathcal{A} \cdot \rho(r) = \gamma(\rho(\phi(r)))$. A *$\rho$-rule model*, with ρ a rule interpretation, is any interpretation I of $\mathcal{R}$ satisfying $\forall r \in \mathcal{R} \cdot [r]_I \geq \rho(r)$. Finally, an interpretation I of $\mathcal{R}$ is called a *k-model* of $\mathcal{R}$, $k \in O_A$, if $[\mathcal{A}]_{\rho_I} \geq k$. Obviously, all ρ_1-rule models of a program are also ρ_2-rule models when $\rho_2 \leq \rho_1$ ($\leq$ being the pointwise extension of $\leq_{O_A}$). Similarly, all k-models of $\mathcal{R}$ are also m-models, for any $m \leq k$.

3 Answer Sets for General Fuzzy Programs

Not all models correspond to our intuition about reasoning using rules. For example, while $\{a^i\}$ is a 1-model of the program $\{a \leftarrow_g .5\}$ as soon as $i \geq 0.5$, the idea of parsimonious rule application only justifies deriving $a^{.5}$. Similarly, the program $\{a \leftarrow_g b, b \leftarrow_g a\}$ admits $\{a^i, b^i\}$ as a 1-model for any $i \in \mathcal{L}$ but only $\{a^0, b^0\}$ is intuitively acceptable. In general, we insist that a nonzero value of a proposition in a model must be motivated by a non-cyclic rule, i.e. if a^l is to be accepted ($l > 0$), it must be implied by a rule $a \leftarrow \alpha$ for which no element in the rule body depends (either directly or indirectly) on a. This consideration also clarifies the role of constraints: since a constraint's consequent is a constant, it cannot be used to motivate a valuation of a proposition; rather it constrains the acceptable valuations of the propositions appearing in its antecedent.

These considerations are captured by the *minimal support* of a rule, defined for a rule $r\colon a \leftarrow \alpha$, a $\mathcal{B}_r$-valuation I and a truth value μ as

$$I_s(r, \mu) = \inf\{y \in \mathcal{L} \mid (y \leftarrow [\alpha]_I) \geq \mu\}$$

Clearly, the minimal support of a rule corresponds to the minimal degree to which the head must be assumed, given the required satisfaction degree μ. Note that from the residuation principle, we straightforwardly find that $I_s(r, \mu) = [\alpha]_I \wedge_r \mu$. The *immediate consequence operator* maps interpretations to the interpretations that result after applying each of the rules once. Specifically, for a set of rules $\mathcal{R}$ and a rule interpretation $\rho : \mathcal{R} \mapsto \mathcal{L}$, we define

$$\Pi_{\mathcal{R},\rho}(I)(a) = \sup\{I_s(r, \rho(r)) \mid (r\colon a \leftarrow \alpha) \in \mathcal{R}\}$$

Note that a similar consequence operator was introduced in [7]. One can (easily) verify that the fixpoints of $\Pi_{\mathcal{R},\rho}$ are ρ-rule models of P, P being an arbitrary program with rule base $\mathcal{R}$. To define acceptable models, called *answer sets* of programs, we will first deal with the simplest case: *positive* (or *simple*) programs, i.e. programs without constraints such that the bodies of all rules are positive expressions.

3.1 Simple Programs

It is easy to see that for simple programs, the immediate consequence operator is monotonically increasing over interpretations, i.e. if $I(a) \leq I'(a)$ for every

proposition a, then also $\Pi_{\mathcal{R},\rho}(I)(a) \leq \Pi_{\mathcal{R},\rho}(I')(a)$ for every proposition a. From [10], we then know that the smallest fixpoint $\Pi^{\star}_{\mathcal{R},\rho}$ of $\Pi_{\mathcal{R},\rho}$ exists and can be obtained by repeated application of $\Pi_{\mathcal{R},\rho}$, starting from the minimal interpretation $\mathcal{I}_0$ defined by $\forall a \in \mathcal{B}_P \cdot \mathcal{I}_0(a) = 0$. Note, however, that this fixpoint may not be reachable in a finite number of steps [5,7].

Definition 1. *Let $P = \langle \mathcal{A}, \mathcal{R}, \phi, \gamma \rangle$ be a simple program. An interpretation M is a* k-answer set *of P iff $[\mathcal{A}]_{\rho_M} \geq k$ and $M = \Pi^{\star}_{\mathcal{R},\rho_M}$.*

Thus answer sets correspond to the maximal amount of knowledge derivable by applying rules (i.e. the immediate consequence operator), starting without any prior assumptions (i.e. the zero interpretation $\mathcal{I}_0$). Note that all k-answer sets are k-models since $[\mathcal{A}]_{\rho_M} \geq k$. As the next proposition reveals, there is a close connection between answer sets and minimal models.

Proposition 1. *Let $P = \langle \mathcal{A}, \mathcal{R}, \phi, \gamma \rangle$ be a simple program with an interpretation M. Then M is an $[\mathcal{A}]_{\rho_M}$-answer set of P iff M is a minimal ρ_M-rule model of P.*

3.2 General Programs

To define answer sets of general programs, first note that any program can be transformed to a program without constraints, by introducing a proposition c_t for each truth value t occurring in the heads of the constraints, and adding rules to ensure that c_t is always interpreted by t; due to space restrictions, we omit the details. Therefore, we will only consider programs without constraints below. To define answer sets for constraint–free programs, we generalize the Gelfond-Lifschitz transformation from [2]. Similar to [4], we reduce the program P to a simple program P^M, called the reduct of P w.r.t. a candidate answer set M.

Definition 2. *Let I be an interpretation. The* simple reduct *of a rule $r\colon a \leftarrow \alpha$, w.r.t. I, denoted r^I, is defined by $r^I = a \leftarrow \alpha^I_-$ where α^I_- is obtained from α by replacing all negative occurrences of propositions from*

$\mathcal{B}_\alpha$ by their value in I. Similarly, the simple reduct of a program P w.r.t. I, denoted P^I, is obtained by replacing each rule r by its reduct r^I.

Obviously, if P is constraint-free, P^I is a simple program as all negative occurrences of propositions are replaced by a value from $\mathcal{L}$. It is easy to see that this is indeed a generalisation of the Gelfond-Lifschitz transformation: in traditional logic programming the only negative occurrence of a proposition a in a rule body is by the use of negation-as-failure.

Definition 3. *Let P be a program without constraints. An interpretation M is a k-answer set of P iff M is a k-answer set of P^M.*

Although the immediate consequence operator for constraint-free programs is, in general, not monotonic (consider e.g. the program $a \leftarrow_g \sim b$), it turns out that an answer set of such a program is still a minimal fixpoint.

Proposition 2. *Let P be a constraint-free program with a k-answer set M. Then M is a minimal fixpoint of $\Pi_{P,M}$.*

Note, however, that the converse of Proposition 2 does not hold.

4 Related Work

Over the past years, there have been various proposals for many-valued logic programming, most of which can be seen as special cases of our approach.

For example, the proposals in [5,6,11] contain general expressions as rule bodies and many-valued predicates, although they do not feature partial rule fulfillment. Such approaches can be readily integrated into ours by using a crisp aggregator expression of the form $r_1 \geq 1 \wedge r_2 \geq 1 \wedge \cdots \wedge r_n \geq 1$.

Other approaches (e.g. [4,7]) do feature partial rule fulfillment, but use weights to express the relative importance of rules. Furthermore, in the proposal of Lukasiewicz & Straccia [4] rule bodies are restricted to applications of t–norms and negators, whereas Damasio et al. [7] only allow monotonically increasing operators in rule bodies. These proposals are special cases of our approach, in which a rule of the form $a \overset{w}{\leftarrow} \alpha$ is transformed to a gfasp rule $a \leftarrow \alpha \wedge_r w$ and the aggregator expression is of the form $\forall r \in P \cdot r = 1$. Due to this observation, our approach inherits the modelling power of [7], and thus also generalizes, among others, hybrid probabilistic logic programs [12].

There is also a strong link between our general fuzzy answer set programming framework and valued constraint satisfaction problems in the sense of [8]. Specifically, it is easy to see that such problems correspond to general fuzzy programs involving only choice and constraint rules. Finally, note that the underlying idea of aggregation expressions also bears resemblance to preference-based frameworks in classical ASP (e.g. [13]). We can show that using suitable constraint rules, such preference-based frameworks can be encoded in our framework, thus yielding a generalization of approaches such as [13] to fuzzy ASP.

5 Conclusions

We introduced general fuzzy answer set programs as an extension of fuzzy answer set programs in the sense of [3]. Rather than restricting rule bodies to be variadic applications of t-norms on (negated) propositions, we allow rule bodies to be defined in terms of arbitrary operators with monotone partial mappings. This is made possible by a generalization of the well-known Gelfond-Lifschitz transformation, whose soundness was demonstrated by linking (partial) answer sets to minimal fixpoints of the immediate consequence operator. Furthermore, we decoupled the order structure used in the aggregator expression from the truth lattice used to interpret the rules, opening up various possibilities for modelling preference amongst rules. In particular, this allows us to define optimal partial answer sets (i.e. k-answer sets), when full answer sets (i.e. answer sets fulfilling all rules fully) do not exist, or cannot be found within a reasonable time frame. The resulting framework turns out to generalize most of the current approaches to many-valued logic programming, including fuzzy and hybrid probabilistic extensions to ASP, as well as various approaches to answer set optimization and valued constraint satisfaction. While we have only discussed our framework from a theoretical perspective, we intend to explore the use of expressive fuzzy SAT solvers to find k-answer sets in practice, generalizing a technique proposed in [14].

References

1. Baral, C.: Knowledge Representation, Reasoning and Declarative Problem Solving. Cambridge University Press, Cambridge (2003)
2. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programming. In: Proceedings of the Fifth International Conference and Symposium on Logic Programming (ICLP/SLP 1988), pp. 1081–1086. ALP, IEEE, The MIT Press (1988)
3. Van Nieuwenborgh, D., De Cock, M., Vermeir, D.: An introduction to fuzzy answer set programming. Annals of Mathematics and Artificial Intelligence 50(3-4), 363–388 (2007)
4. Lukasiewicz, T., Straccia, U.: Tightly integrated fuzzy description logic programs under the answer set semantics for the semantic web. In: Marchiori, M., Pan, J.Z., de Sainte Marie, C. (eds.) RR 2007. LNCS, vol. 4524, pp. 289–298. Springer, Heidelberg (2007)
5. Straccia, U.: Annotated answer set programming. In: Proceedings of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2006) (2006)
6. Damásio, C.V., Pereira, L.M.: Antitonic logic programs. In: Eiter, T., Faber, W., Truszczyński, M. (eds.) LPNMR 2001. LNCS, vol. 2173, pp. 379–392. Springer, Heidelberg (2001)
7. Damásio, C.V., Medina, J., Ojeda-Aciego, M.: Sorted multi-adjoint logic programs: termination results and applications. In: Alferes, J.J., Leite, J. (eds.) JELIA 2004. LNCS, vol. 3229, pp. 260–273. Springer, Heidelberg (2004)
8. Schiex, T., Fargier, H., Verfaillie, G.: Valued constraint satisfaction problems: hard and easy problems. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI 1995), pp. 631–637 (1995)
9. Dubois, D., Fortemps, P.: Computing improved optimal solutions to max–min flexible constraint satisfaction problems. European Journal of Operational Research 118, 95–126 (1999)
10. Tarski, A.: A lattice theoretical fixpoint theorem and its application. Pacific Journal of Mathematics 5, 285–309 (1955)
11. Damasio, C.V., Pereira, L.M.: An encompassing framework for paraconsistent logic programs. Journal of Applied Logic 3, 67–95 (2003)
12. Damásio, C.V., Pereira, L.M.: Hybrid probabilistic logic programs as residuated logic programs. In: Brewka, G., Moniz Pereira, L., Ojeda-Aciego, M., de Guzmán, I.P. (eds.) JELIA 2000. LNCS, vol. 1919, pp. 57–72. Springer, Heidelberg (2000)
13. Brewka, G., Niemelä, I., Truszczyński, M.: Answer set optimization. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence, pp. 867–872. Morgan Kaufmann Publishers, San Francisco (2003)
14. Janssen, J., Heymans, S., Vermeir, D., De Cock, M.: Compiling fuzzy answer set programs to fuzzy propositional theories. In: Proceedings of the 24th International Conference on Logic Programming (ICLP 2008). Springer, Heidelberg (2008) (to appear)

Reverse Engineering of Regulatory Relations in Gene Networks by a Probabilistic Approach

Michele Ceccarelli[1,2], Sandro Morganella[1,2], and Pietro Zoppoli[2]

[1] Research Centre on Software Technologies-RCOST, University of Sannio, Benevento, Italy
ceccarelli@unisannio.it
[2] Biogem s.c.a.r.l., Ariano Irpino (Avellino), Italy
{morganella,zoppoli}@biogem.it

Abstract. In the last years microarray technology has revolutionised the fields of genetics, biotechnology and drug discovery. Due to its high parallelity, different analyses can be accomplished in one single experiment to generate vast amounts of data. In this paper we propose a new approach to solve the reverse engineering of regulatory relations task into gene networks from high-throughput data. We develop an Inference of Regulatory Interaction Schema (**IRIS**) algorithm that uses an iterative method to map gene expression profile values (steady-state and time-course) into discrete states, so that, a probabilistic approach can be used to infer gene interaction rules. IRIS provides two different descriptions of each regulatory relation: the description in which interactions are described as conditional probability tables (CPT-like) and descriptions in which regulations are truth tables (TT-like). We test IRIS on two synthetic networks and on real biological data showing its accuracy and efficiency.

At URL http://bioinformatics.biogem.it a Matlab implementation of IRIS is available.

1 Introduction

Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Studying which genes are active and which are inactive in different kinds of cells helps scientists understand more about how these cells work and about what happens when the genes in a cell don't work properly. Recent years witnessed an information revolution, following the advent of novel high-throughput experimentation methods which encompass biological system on a new scale. Most notable of these methods are transcription profiling using oligonucleotide chips and DNA microarrays. With the development of these technologies, scientists can now examine thousands of genes at the same time.

In this scenario, molecular genetics and biology are rapidly evolving into a quantitative science, and such as, it is increasingly relying on engineering and physics to make sense of high-throughput data. Development of reliable data

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 360–367, 2009.

analysis methods to infer complex networks of a living system based on high-throughput data, is a major issue in current bioinformatics research. Algorithms that solve this task are called reverse engineering algorithms [1]. We can distinguish two kinds of reverse engineering algorithms: algorithms inferring gene network topologies and algorithms learning regulation rules.

In literature several approaches to infer gene networks exist. Each proposed method is based on a particular mathematical formalism as information-theoretic [2,3], ordinary differential equations [4,5,6] and Bayesian networks [7,8,9]. These algorithms provide either none or few informations of regulation rules which represent another fundamental aspect in the complex orchestration of a cell. Gat-Viks *et al.* [10] recently proposed a method that can be used for this purpose. The major disadvantage of their approach is that a set of initial regulation functions must be user-defined which are then integrated with experimental observations to compute refined regulations. In this way the results are strongly influenced by the initial user-defined regulations. As an alternative to manual setting of regulation rules, methods based on Bayesian networks and Maximum likelihood estimation can be used to infer automatically parameters into directed acyclic graphs [11,12]. Here we propose an iterative approach to learn regulatory relations from gene expression profiles. In our approach only the topology of a network is required, which can be represented by any direct graph including graphs with cycles.

2 Model for Biological Networks

A biological network can be modeled by a direct graph $G(V, E)$. Each node $v \in V$ represents a gene, which may attain a discrete state $D = \{0, 1\}$ to represent inactive and active state respectively. If a gene $v \in V$ has at least one parent then it is called *regulatee*, we define as R_v the set of its parents (*regulators*). If a gene $v \in V$ has none parent, then it is a *stimulator* and we define as V_s the set of all stimulators. Our biological model follows the notation described in [13]. In addition a matrix $M = n \times m$ is used to represent experimental data (n is the number of genes and m is the number of performed experiments). For each $1 \leq i \leq n$ and $1 \leq j \leq m$ $M[i, j]$ is defined as:

$$M[i,j] \begin{cases} \in \mathbb{R} & \text{if the gene } i \text{ is measured into experiment } j \\ = NaN & \text{if the gene } i \text{ is not measured into experiment } j \end{cases} \quad (1)$$

3 IRIS Algorithm

In this section we describe our approach to infer regulatory relations in gene networks from gene expression data. IRIS needs of a network topology G defining relationships between genes, a gene expression profile matrix data M and a constant α:

We notice that regulation functions are inferred only for genes that have at least one regulator (line 1).

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
A	a_1	a_2	NaN	a_4	a_5	NaN	a_7	a_8	a_9	a_{10}
B	b_1	NaN	b_3	b_4	b_5	NaN	b_7	NaN	b_9	b_{10}
C	c_1	NaN	NaN	c_4	c_5	c_6	c_7	NaN	c_9	c_{10}

(a)

	e_1	e_4	e_5	e_7	e_9	e_{10}
A	a_1	a_4	a_5	a_7	a_9	a_{10}
B	b_1	b_4	b_5	b_7	b_9	b_{10}
C	c_1	c_4	c_5	c_7	c_9	c_{10}

(b)

Fig. 1. (a) Hypothetical experimental matrix data for three genes A, B and C, where A and B are regulators of C. (b) The experimental sub-matrix data after *compacting* action.

Algorithm 1.1 IRIS($G = (V, E)$, M, α)

1: **for** *each gene* $g \in V \setminus V_s$ **do**
2: INITIALIZATION
3: DISCRETIZATION
4: LEARNING_REGULATION_FUNCTION
5: **end for**

3.1 Initialization

Aim of this step is both to handle missing data and to calculate specific parameters for each gene. Handling of missing data is done by a *compacting* action. In this step we store for each regulation only the experiments in M where all involved genes are measured (Fig. 1). We refer to this new matrix as sub-matrix M_s. This compacting action allows to use both steady-state and time-course expression profiles. We note that a sub-matrix for each regulatee gene exists. For each row i of the current sub-matrix we also compute the following parameters: mean μ_i, standard deviation σ_i and two thresholds:

$$up_i = \mu_i + \alpha\ \sigma_i \quad \text{and} \quad down_i = \mu_i - \alpha\ \sigma_i \tag{2}$$

3.2 Discretization

In this step we map real-valued measurements into discrete states. We fix the values that can be considered with high likelihood as an active/inactive state using the thresholds defined in (2) and the derivative concept. Let M_s be a sub-matrix, we compute the discretized derivative as $M'_s[i,j] = M_s[i,j] - M_s[i,j-1]$ for $j = 2, \ldots, m_s$ where m_s is the number of columns of M_s. Let S be the matrix of discrete states ($S[i,j]$ contains the discretized state of the real value $M_s[i,j]$). We compute a first discretization step using the rule:

$$S[i,j] = \begin{cases} 0 & \text{if } M'_s[i,j] \leq 0 \text{ AND } M_s[i,j] \in (-\infty, down_i] \\ 1 & \text{if } M'_s[i,j] \geq 0 \text{ AND } M_s[i,j] \in [up_i, +\infty) \\ NaN & \text{Otherwise} \end{cases} \tag{3}$$

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}
A	1	NaN	1	NaN	NaN	0	NaN	0	0	0	NaN
B	1	NaN	1	1	1	NaN	1	NaN	1	NaN	1
C	1	1	1	0	0	0	0	NaN	0	NaN	0

Fig. 2. Example of matrix S

with $2 \leq j \leq m_s$. Figure 2 shows an example of matrix S for three genes and for a sub-matrix M_s with 11 experiments.

This discretization rule is very strict so it provides a matrix S where the values identified as active/inactive can be considered as fixed points. However, there can be many uncertain values, so we need to recover some of these. We perform the recovery step as follows:

$$S[i,j] = \begin{cases} 0 \text{ if } S[i,j-1] = 0 \text{ AND } S[i,j+1] = 0 \text{ AND } M_s[i,j] \leq \mu_i \\ 1 \text{ if } S[i,j-1] = 1 \text{ AND } S[i,j+1] = 1 \text{ AND } M_s[i,j] \geq \mu_i \end{cases} \tag{4}$$

$$S[i,j] = \begin{cases} 1 \text{ if } S[i,j-1] = 1 \text{ AND } S[i,j+1] = NaN \text{ AND } M'_s[i,j] \geq 0 \\ 1 \text{ if } S[i,j-1] = NaN \text{ AND } S[i,j+1] = 1 \text{ AND } M'_s[i,j] \geq 0 \\ 0 \text{ if } S[i,j-1] = 0 \text{ AND } S[i,j+1] = NaN \text{ AND } M'_s[i,j] \leq 0 \\ 0 \text{ if } S[i,j-1] = NaN \text{ AND } S[i,j+1] = 0 \text{ AND } M'_s[i,j] \leq 0 \end{cases} \tag{5}$$

with $2 \leq j \leq m_s - 1$. We use an iterative approach that reapplies (4) and (5) until either all values have been assigned to a valid active/inactive state or in the last iteration no recovery action has been performed.

3.3 Learning_Regulation_Function

In this step we use the matrices S to learn the conditional probability tables (CPTs) for all regulatee genes. We compute for each regulatee gene also a regulation function TT-like. To infer the regulation functions CPT-like we use the theory of the relative frequency [14]. Let us consider a gene v and the set of its regulators R_v then the matrix S will contain several state assignments for the genes in R_v and v itself. Let Γ_v be the set of all possible state assignments of the variables in R_v, then we have:

$$\begin{aligned} &freq(\{r_v, v = 0\}) = |\{r_v, v = 0\}| \\ &\qquad\qquad\qquad\qquad \text{with } r_v \in \Gamma_v \\ &freq(\{r_v, v = 1\}) = |\{r_v, v = 1\}| \end{aligned} \tag{6}$$

where $|\{r_v, v = s\}|$ are the occurrences number in S of the state assignment $\{r_v \cup v = s\}$. Let $P(\{v = s\}|\{r_v\})$ be the conditional probability of gene v to attain the state $s \in D$ given the state assignment $r_v \in \Gamma_v$, then we have:

$$P(\{v=0\}|\{r_v\}) = \frac{freq(\{r_v, v=0\})}{freq(\{r_v, v=0\}) + freq(\{r_v, v=1\})}$$
$$P(\{v=1\}|\{r_v\}) = \frac{freq(\{r_v, v=1\})}{freq(\{r_v, v=0\}) + freq(\{r_v, v=1\})} \quad \forall\, r_v \in \Gamma_v \text{ in } S \quad (7)$$

Using the (7) we compute the truth table for each gene, as:

$$T(\{r_v\}) = \begin{cases} 0 & \text{if } P(\{v=0\}|\{r_v\}) > P(\{v=1\}|\{r_v\}) \\ 1 & \text{if } P(\{v=1\}|\{r_v\}) > P(\{v=0\}|\{r_v\}) \\ -1 & \text{Otherwise} \end{cases} \qquad \forall\, r_v \in \Gamma_v \quad (8)$$

where $T(\{r_v\})$ represents the state response of the regulatee gene v to state assignment r_v of all its regulators.

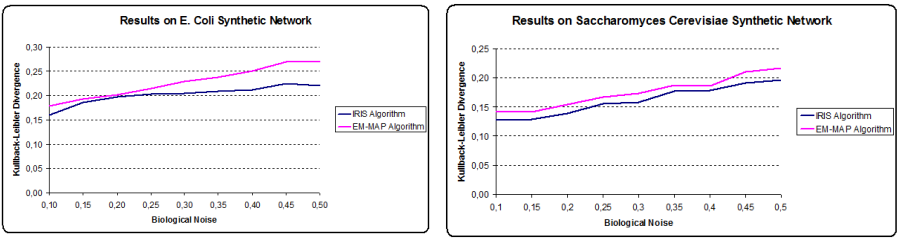
Note if $P(\{v=0\}|\{r_v\}) = P(\{v=1\}|\{r_v\}) = 0.50$ we cannot distinguish between active and inactive state, so we have a undefined response. Into truth tables this situation is indicates as -1.

4 Results

In this section the results of our method are shown. In detail, we apply our approach both on two synthetic networks and on a real biological dataset. The synthetic networks are generated by SynTReN [15] which allows to generate both well-known networks and gene expression profile dataset. For both synthetic networks we generate nine datasets using nine different biological noise levels: 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50. To validate the accuracy of our approach we use the Expectation Maximization Maximum a Posteriori (EM-MAP) algorithm [12] as comparison method. The Kullback-Leibler Divergence (KLD) [16] is used to compare the CPTs generated by IRIS with the ones generated by EM-MAP. The results are depicted in Fig. 3. Both methods use as input the discrete data computed by IRIS discretization step using as constant $\alpha = 0.1$[1], so they differ just in the CPT computation method.

The first observation is that both methods have KLDs close to zero, which represent KLD of two equal probability distributions. In this way we show that the proposed discretization method maps with a good accuracy real values into discrete states. In addition, for each biological noise level IRIS provides CPTs that better approximates true probability distribution then EM-MAP. We also apply IRIS algorithm to solve the regulation function reverse engineering task on real gene expression profiles [17] for the yeast mitotic cell-cycle. The target gene network, depicted in Fig. 4, has been extracted from the work of Noman and Iba [18]. We obtain a description of all regulatory relations for this network from the literature [19] [20]. This relations can be considered to be true tables for this

[1] The value $\alpha = 0.1$ has been assessed by a tuning process.



(a) (b)

Fig. 3. Graphical results on synthetic networks (a) *E. coli* (b) *S. cerevisiae*

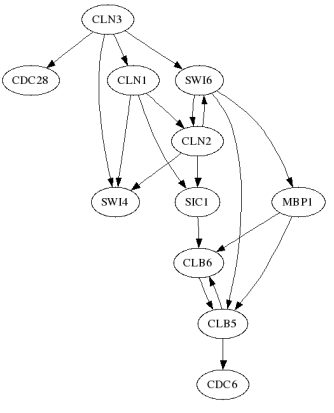


Fig. 4. Cell-cycle network of *S. cerevisiae*

pathway with an high degree of confidence. We compare the true tables with the regulatory relations computed by IRIS using again $\alpha = 0.1$ and we obtain: 79.55% of correct states, 11.36% of incorrect states and 9.09% of undefined states. The percentage of the correct valuations increases if we don't consider the undefined states (which cannot be considered neither correct nor incorrect states), so we have: 87.50% of correct states and 12.50% of incorrect states. Using the regulation functions TT-like obtained by IRIS we can also compute the steady states [13] of the network and two well-known steady states [20] are obtained (Tab. 1). Integrating the inferred regulation functions CPT-like in factor graph probabilistic graphical model and using an inference algorithm, known as sum-product [21], we can compute the posterior distribution of hidden

Table 1. Well-known steady states obtained using the regulation functions TT-like computed by IRIS algorithm

CLN3	*MBP1*	*SWI6*	*SWI4*	*CLN1*	*CLN2*	*CLB5*	*CLB6*	*SIC1*
0	0	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0

variables given the observed ones. Using this approach we found well-known features of the network [20,22,23,24,25,26].

5 Conclusions

In this paper a new approach to infer regulatory relations into gene networks from microarray data has been proposed. This approach provides an iterative method to translate gene expression profiles into two discrete states (active and inactive). It allows to describe each gene interaction as CPT and truth table.

We demonstrated on two synthetic networks that our method has a good accuracy. In detail, The results show that our algorithm (IRIS) provides more accurate results than EM-MAP algorithm. Moreover, we tested our approach on real gene expression profiles for the yeast mitotic cell-cycle showing that IRIS allows to infer regulatory relations with an accuracy of 79.55% and to obtain findings well-known in literature.

References

1. Bensal, M., Bistro, V., Imposable, A.A., Bernardo, D.D.: How to Infer Gene Networks from Expression Profiles. Molecular System Biology 3, 78 (2007)
2. Basso, K., Margolin, A., Stolovitzky, G., Klein, U., Favera, R.D., Califano, A.: Reverse Engineering of Regulatory Networks in Human B Cells. Nature Genetics 37(4), 382–390 (2005)
3. Basso, K., Margolin, A., Stolovitzky, G., Klein, U., Favera, R.D., Califano, A.: Inferring Gene Regulatory Networks Using Differential Evolution with Local Search Heuristics. IEEE Transaction on Computational Biology and Bioinformatics 4(4), 634–647 (2007)
4. Gardner, T., Bernardo, D.D., Lorenz, D., Collins, J.: Inferring Genetic Networks and Identifying Compound Mode of Activation Via Expression Profiling. Science 301(5629), 102–105 (2003)
5. Bernardo, D.D., Thompson, M., Gardner, T., Chobot, S., Eastwood, E., Wojtovich, A., Elliott, S., Schaus, S., Collins, J.: Chemogenomic Profiling on a Genome-Wide Scale Using Reverse-Engineered Gene Networks. Nature Biotechnology 39(23), 377–383 (2005)
6. Bensal, M., Gatta, G.D., Bernardo, D.D.: Inference of Gene Regulatory Networks and Compound Mode of Action from Time Course Gene Expression Profiles. Bioinformatics 22(7), 815–822 (2006)
7. Yu, J., Smith, V., Wang, P., Hartemink, A., Jarvis, E.: Advances to Bayesian Network Inference of Generating Casual Networks from Observational Biological Data. Bioinformatics 20, 3594–3603 (2004)
8. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using Bayesian Networks to Analyze Expression Data. In: Proc. Fourth Annual Int. Conf. on Computational Molecular Biology, pp. 127–135. ACM Press, New York (2000)
9. Pe'er, D., Regev, A., Elidan, G., Friedman, N.: Inferring Subnetowrks from perturbated Expression Profile. Bioinformatics 1(1), 1–9 (2001)
10. Ulitsky, I., Gat-Viks, I., Shamir, R.: MetaReg: A Platform for Modeling, Analysis and Visualization of Biological Systems Using Large-Scale Experimental Data. Genome Biology 9 (2008)

11. Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society 39(1), 1–38 (1977)
12. Lauritzen, S.: The EM Algorithm for Graphical Association Models with Missing Data 19 (1995)
13. Gat-Viks, I., Tanay, A., Shamir, R.: Modeling and Analysis of Heterogeneous Regulation in Biological Networks. In: Eskin, E., Workman, C. (eds.) RECOMB-WS 2004. LNCS (LNBI), vol. 3318, pp. 98–113. Springer, Heidelberg (2004)
14. Frassen, B.V.: Relative Frequencies. Synthese 34(2), 133–166 (2004)
15. den Bulcke, T.V., Leemput, K.V., Naudts, B., van Remortel, P., Ma, H., Verschoren, A., Moor, B.D., Marchal, K.: SynTReN: a Generator of Synthetic Gene Expression Data for Design and Analysis of Structure Learning Algorithms. BMC Bioinformatics 7 (2006)
16. Kullback, S., Leibler, R.: On Information and Sufficiency. Annals of Mathematical Statistics (22), 79–86 (1951)
17. Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botsein, D., Futcher, B.: Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Molecular Biology of the Cell 9(12), 3273–3297 (1998)
18. Noman, N., Iba, H.: Inferring Gene Regulatory Networks Using Differential Evolution with Local Search Heuristics. IEEE Transaction on Computational Biology and Bioinformatics 4, 634–647 (2007)
19. KEGG: (Kyoto Encyclopedia of Genes and Genomes), http://www.genome.ad.jp/kegg/
20. Li, F., Long, T., Lu, Y., Tao, C.: The Yeast Cell Cycle Network is Robustly Designed. PNAS 101(14), 4781–4786 (2004)
21. Kschischang, F.R., Brendan, J.F., Loeliger, H.A.: Factor Graphs and the Sum-Product Algorithm. IEEE Transactions on Information Theory 47(2), 498–519 (2001)
22. Schwob, E., Nasmyth, K.: CLB5 and CLB6, a new Pair of B Cyclins Involved in DNA Replication in Saccharomyces Cerevisiae. Genes and Development 7, 1160–1175 (1993)
23. Di Como, C.J., Chang, H., Arndt, K.T.: Activation of CLN1 and CLN2 G1 cyclin gene expression by BCK2. Molecular and Cellular Biology 15(4), 1835–1846 (1995)
24. Nugorho, T.T., Mendenhall, M.D.: An Inhibitor of Yeast Cyclin-dependent Protein Kinase Plays an Important Role in Ensuring the Genomic Integrity of Daughter Cells. Molecular and Cellular Biology 14(5), 3320–3328 (1994)
25. Verma, R., Annan, R.S., Huddleston, M.J., Carr, S.A., Reynard, G., Deshaies, R.J.: Phosphorylation of Sic1p by G1 Cdk Required for Its Degradation and Entry into S Phase. Science 278(5337), 455–460 (1997)
26. Amon, A., Tyers, M., Futcher, B., Nasmyth, K.: Mechanisms that Help the Yeast Cell Cycle Clock Tick: G2 Cyclins Transcriptionally Activate G2 Cyclins and Repress G1 Cyclins. Cell 74(6), 993–1007 (1993)

Temporal Features in Biological Warfare

Silvana Badaloni and Marco Falda

Dept. of Information Engineering - University of Padova
via Gradenigo, 6 - 35131 Padova, Italy
{silvana.badaloni,marco.falda}@unipd.it

Abstract. No matter how prepared a population may be, bioterrorism cannot be prevented: the first clues will always be given by ill people. Temporal analysis applied to this type of scenarios could be an additional tool for limiting disruption among civilians allowing for recognizing typical temporal progression and duration of symptoms in first infected people. We propose the application of a fuzzy temporal reasoning system we have developed for biomedical temporal data analysis in different scenarios after a hypothetical attack. The system is able to handle both qualitative and metric temporal knowledge affected by vagueness and uncertainty, taking into account in this way the vagueness of patients reports expressed in natural language.

1 Introduction

In case of biological attacks, the effects of a deliberate release will be obvious if a large number of troops become ill with similar symptoms at the same time. It may be less clear in a civilian population [1], supposed to be in a period of peace. For this reason establishing a diagnosis is critical to the public health response to a bioterrorism-related epidemic, since the diagnosis will guide the use of vaccinations, medications, and other interventions [2]. Moreover, new or reemerging infectious diseases have relevant implications: during the past 20 years, over 30 new lethal pathogens have been identified; for example the emergence of Severe Acute Respiratory Syndrome (SARS) in Southeast Asia rapidly spread to 29 countries in less than 90 days [3]. Emerging disease outbreaks may be difficult to distinguish from the intentional introduction of infectious diseases for nefarious purposes, when considering that Genetic Engineering of biological warfare agents can alter their pathogenicity, incubation periods, or even the clinical syndromes they cause. For this reason, it is important to develop automatic Syndromic Surveillance Systems [2] able to notify as soon as possible the early manifestations of bioterrorism-related diseases from population monitoring. A preliminary step towards the design of a component of such a System could be based on the use of temporal reasoning techniques in order to identify typical temporal progression of diseases.

Taking into account that medical data relative to temporal evolution of diseases are often affected by vagueness and uncertainty, the temporal reasoning model that seems to be more adequate for such real application could be the Fuzzy Temporal Reasoning System (FTR in the following) that we developed in a previous research [4].

The system is based on the integration of temporal information both qualitative and metric represented as fuzzy constraints in a network and extends a previously proposed

V. Di Gesù, S.K. Pal, and A. Petrosino (Eds.): WILF 2009, LNAI 5571, pp. 368–375, 2009.

system [5] that dealt with fuzzy qualitative temporal reasoning. We have applied our System in several diagnostic problems. In [6], we applied it for discriminating exanthematic diseases from temporal patterns of patient symptoms and in [7] we studied how our system could represent temporal evolutions of symptoms in different patients affected by SARS, thus making possible to deduce characteristic periods of a new disease.

Dealing with the study of biological warfare, we address the problem of the automated analysis of temporal medical data in order to obtain information useful for early detection of biological attacks. In particular, we will start from the temporal evolution of five NIAID (National Institute of Allergy and Infectious Diseases) diseases represented as fuzzy constraint temporal networks. Then we will check the consistency of temporal data relative to a set of ten patients reports [8] with respect to the previously considered diseases; we will use the algorithmic methodologies for checking temporal consistency offered by the FTR system. Two are the main objectives:

- to find the most plausible disease and, once found it,
- to exploit the information in order to infer the possible contagion.

The paper is organized as follows. Section 2 describes the problem of identifying biological attacks while Section 3 is dedicated to a brief presentation of the FTR System. In Section 4 the considered diseases are summarized and in Section 5 the results about temporal analysis of patients data is shown.

2 Identifying Biological Attacks

Early symptoms of disease induced by a biological warfare agent may be non-specific or difficult to recognize, for example a simple febrile illness; the disease itself could affect individuals living in widely dispersed areas, who may then present to several different healthcare providers [1]; once the disease has been diagnosed, appropriate prophylaxis, treatment, and other measures to decrease spreading, such as quarantine (for a contagious illness) would be adopted.

As said before, many diseases caused by bioterrorism present with relatively common features, such as fever or headache, but there are several considerations that can ease the identification of a Bioterrorism-related scenario [9]:

symptoms: a number of patients that abruptly present to care providers or emergency rooms manifesting similar and unexpected symptoms;

zoonoses: most of the agents used in biological warfare are diseases that affect animals, for this reasons sudden deaths between animals can anticipate diffusion among humans;

unexplained factors: whenever an unusual pattern is detected a biological attack may be suspected: unexplained deaths for an usually mild disease, unusual exposure routes for a pathogen, for a geographical area, for a season;

diffusion patterns: higher symptoms manifestations in certain areas, for example buildings, or in short time periods. The abrupt onset and single peak of cases would implicate a point-source exposure without secondary transmission [10].

In this paper only considerations about time will be taken into account.

3 The Fuzzy Temporal Reasoning System

In the present section we will summarize the main characteristics of the FTR System (for a more detailed description cfr. [4]).

In Allen's Interval Algebra [11] the temporal qualitative knowledge is represented as a binary relation between a pair of intervals in terms of *atomic relations*:

$$I_1 \ (rel_1, \ldots, rel_m) \ I_2$$

where each rel_i is one of the 13 mutually exclusive atomic relations that may exist between two intervals (such as *equal*, *before*, *meets* etc.).

To deal with vague and uncertain temporal information Allen's Interval Algebra has been extended in [5] with the Possibility Theory by assigning to every atomic relation rel_i a degree α_i, which indicates the *preference degree* of the corresponding assignment among the others

$$I_1 \ R \ I_2 \text{ with } R = (rel_1[\alpha_1], \ldots, rel_{13}[\alpha_{13}])$$

where α_i is the preference degree of rel_i $(i = 1, \ldots, 13)$; preferences can be defined in the interval $[0, 1]$. If we take the set $\{0, 1\}$ the classic approach is obtained.

Intervals are interpreted as ordered pairs $(x, y) : x \leq y$ of $\Re^2$, and soft constraints between them as fuzzy subsets of $\Re^2 \times \Re^2$ in such a way that the pairs of intervals that are in relation rel_k have membership degree α_k.

Temporal metric constraints have been extended to the fuzzy case starting from the traditional TCSPs [12] in many ways [13,14,15]. To represent fuzzy temporal metric constraints we adopt trapezoidal distributions [4], since they seem enough expressive and computationally less expensive than general semi-convex functions [16].

Each trapezoid is represented by a 4-tuple of values describing its four characteristic points plus a degree of consistency α_i denoting its height.

$$T_k = \ll a_k, b_k, c_k, d_k \gg [\alpha_k]$$

with $a_k, b_k \in \Re \cup \{-\infty\}$, $c_k, d_k \in \Re \cup \{+\infty\}$, $\alpha_k \in (0, 1]$, $\ll$ is either `(` or `[` and $\gg$ is either `)` or `]`. The points b_k and c_k determine the interval of those temporal values which are likely, whereas a_k and d_k determine the interval out of which the values are absolutely impossible. The generalized definition of trapezoid extreme increases the expressiveness of the language.

As far as integration is concerned, we have defined the fuzzy extensions PA^{fuz}, PI^{fuz}, IP^{fuz} and IA^{fuz} of the corresponding classical algebras PA, PI, IP and IA referring to point-point, point-interval, interval-point and interval-interval relations [17,4,5], we have extended the composition operation and the transitivity table [18]. In the integrated framework we can manage temporal networks where nodes can represent both points and intervals, and where edges are accordingly labeled by qualitative and quantitative fuzzy temporal constraints.

Path-Consistency and Branch & Bound algorithms have been generalized to the fuzzy case adding some relevant refinements that improve their efficiency. Path-consistency has a polynomial computing time and it is used to prune the search space in the

Branch & Bound algorithm; however for real world applications tractable subsets of relations such as those belonging to the Convex Pointizable Algebra SA_c should be used, since in that case Path-consistency is sufficient to find the minimal network [19].

4 Biological Agents

Before presenting the application of the Temporal Reasoning system it is useful to briefly describe the biological agents that will be considered.

4.1 NIAID "Category A" Main Diseases

NIAID (National Institute of Allergy and Infectious Diseases) is the primary Institute at NIH, the US National Institute of Health, for emerging infectious diseases research, including research on agents of bioterrorism. This institute has grouped biological agents in three categories according to their ease of use for a biological attack; the most dangerous are in Category A (Table 1) and are agents that can be easily disseminated or transmitted person to person, that have high mortality and can cause public panic and social disruption, therefore needing special action for public preparedness.

Table 1. Category A agents

• *Bacillus anthracis* (Anthrax)
• *Clostridium botulinum* toxin (Botulism)
• *Yersinia pestis* (Plague)
• *Variola major* (Smallpox) and other pox viruses
• *Francisella tularensis* (Tularemia)
• *Viral hemorrhagic fevers* (VHF)

We have considered the timelines of five diseases: Anthrax, Tularemia, Smallpox, Plague and Ebola. These timelines can be obtained from temporal characteristics of the diseases themselves and are reported in Figure 1. In the following just Anthrax and Plague are described.

Anthrax (*Bacillus anthracis*). Anthrax is one of the most serious diseases: when inhaled it can be quite lethal [9].

Most of the early symptoms of inhalation Anthrax are similar to those for other infectious diseases, making a differential diagnosis difficult during flu season, for example [20]. The distribution of the incubation period for inhalational Anthrax can be relatively broad as observed in Sverdlovsk (2-43 days); in any case, it does not extend more than 60 days. The clinical presentation has been described as a 2-phases illness: the nonspecific prodrome for Anthrax may last from several hours to several days [2]. The second phase develops abruptly, with sudden fever, dyspnea, diaphoresis, and shock.

Case fatality rates of 80% or more, with nearly half of all deaths occurring within 24 to 48 hours, is highly likely to be Anthrax or pneumonic plague. A temporal constraint network for modelling Anthrax can be composed by four vertices: the contagion (1), the first symptom (2), the worsening phase (3) and the death or recovery.

The constraints, deduced from the previous description, are expressed in hours.

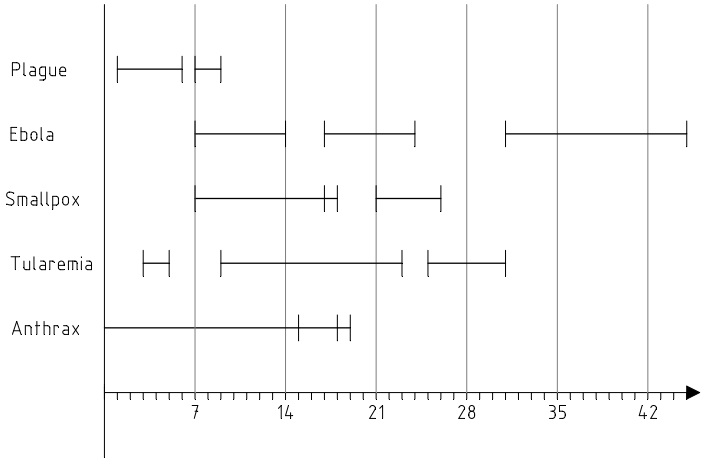


Fig. 1. Timelines for the considered diseases; for each disease the incubation period, the worsening period and the death have been represented (in days)

- Incubation lasts no more than 60 days: $1\ \{(-\infty, -\infty, 1440, 1440]\}\ 2$
- First phase lasts from several hours to several days: $2\ \{(6, 12, 24, 96)\}\ 3$
- Death occurs within 24 to 48 hours: $3\{(12, 24, 48, 60)\}\ 4$

Plague (*Yersinia pestis*). Plague is of great concern in a biological attack scenario, since it is available around the world, it is easy to produce and disseminate it through aerosolization; moreover, it causes high fatality rates and can rapidly spread during an epidemic [21]. Vaccine has limited efficacy following aerosol dispersion [9].

A pneumonic plague outbreak would result with symptoms initially resembling those of other severe respiratory illnesses. Exposure to aerosolized *Y. pestis* results in pneumonic plague, which has a typical incubation period of 2 to 4 days (range 1-6 days).

The fatality rate of patients with pneumonic plague when treatment is not commenced within 24 hours of symptoms onset is extremely high [21]. In modelling plague constraint network notice that second phase is almost immediate, therefore, assuming that the same vertices are used for all diseases, the constraint between vertex 3 and 4 could be:

$$3\{before[0.5], meets[1.0]\}4$$

Notice that in the description of Anthrax fuzzy metric constraints were used, while here also a qualitative fuzzy temporal constraint has been specified. This shows that a user can represent the temporal knowledge as it is available.

5 Temporal Analysis

To develop a general framework for automated temporal analysis of biological warfare data different aspects can be considered. Here first we apply the solver to match temporal data coming from patients with the typical evolution of the five diseases previously

cited in order to identify the most plausible disease. Second, when the disease has been selected its characteristic development is used to infer the contagion period.

5.1 Patients Reports

We consider a set of medical data concerning 10 patients reports [8]. These descriptions contain temporal information that can be modelled using a temporal constraint network according to the FTR representation system.

The timetable of all the ten patients is shown in Figure 2. In the following we report as a detailed example the description and the modelling of the first patient.

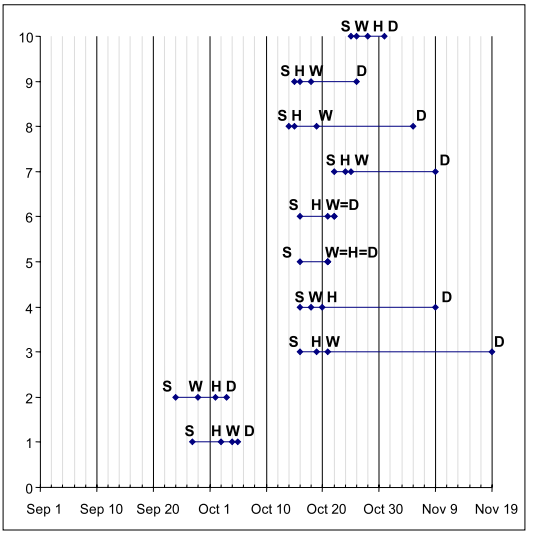


Fig. 2. Timelines for patients (S = first symptom, H = hospitalization, W = worsening, D = death/discharge)

Patient 1. *On October 2, 2001*, a 63-year-old Caucasian person awoke early with nausea, vomiting, and confusion and was taken to a local emergency room for evaluation. His illness, which started *on September 27* was characterized by malaise, fatigue, fever, chills, anorexia and sweats. [...] *On hospital day 2*, penicillin G, levofloxacin, and clindamycin were begun. He remained febrile and became unresponsive to deep stimuli. His condition progressively deteriorated, with hypotension and worsening renal insufficiency. The patient died *on October 5*.

A temporal constraint network for modelling, for instance, Patient1 can be composed by five vertices:

1. the origin of time t_0;
2. first symptom (S);
3. worsening (W);
4. hospitalization (H);
5. death / discharge from hospital (D).

The constraints, deduced from the previous reports and expressed in hours from Jan 1 (t_0), can be represented as:

- 1 $[6456, 6456, 6480, 6480]$ 2 (on Sep 27)
- 1 $[6576, 6576, 6600, 6600]$ 4 (on Oct 2)
- 1 $[6624, 6624, 6648, 6648]$ 3 (on Oct 4)
- 1 $[6648, 6648, 6672, 6672]$ 5 (on Oct 5)

Now, to find the most plausible disease we combine the patients networks with the network of each agent. In this way, by means of a consistency analysis, we can have an

idea of the disease that has the highest compatibility with the considered scenario and then infer the contagion period.

Assuming that the outbreak is located in a single source, all patients should become ill within the incubation period. Applying the FTR system it results that Anthrax is the only disease among the 5 considered which is consistent with all patients; for example the Plague incubation period is too short to fully accommodate a range of 1 month between the appearing of the symptoms in the patients. This inference confirms the hypothesis about Anthrax found by laboratory tests [8].

Then, taking into account that Anthrax incubation lasts no more than 60 days and that symptoms in all patients appeared from September 24 to October 26, the FTR system can deduce that contagion of all these patients could have occurred from the end of July to few days before September 22.

6 Conclusions

In this paper we have studied how to develop a Temporal Reasoner for an automatic Syndromic Surveillance System able to notify as soon as possible the early manifestations of bioterrorism-related diseases from population monitoring. To this aim the detection of temporal characteristic features become an important aspect that we have addressed using the Fuzzy Temporal Reasoning System. This system has allowed inferring information about possible contagion period in an Anthrax attack scenario happened in U.S. in 2001.

As future directions are concerned, we intend to enrich the analysis capabilities of the FTR system for example to identifying clusters in contagion dynamics. In this way we aim to develop a more sophisticated system to face this global threat.

References

1. Beeching, N.J., Dance, D.A.B., Miller, A.R.O., Spencer, R.C.: Biological warfare and bioterrorism. BMJ 324, 336–339 (2002)
2. Buehler, J.W., Berkelman, R.L., Hartley, D.M., Peters, C.J.: Syndromic surveillance and bioterrorism-related epidemics. Emerg. Infect. Dis. 9, 1197–1204 (2003)
3. Kortepeter, M., Christopher, G., Cieslak, T., Culpepper, R., Darling, R.: USAMRIID's Medical Management of Biological Casualties Handbook. 6th edn. U.S. Army Medical Research Institute of Infectious Diseases (2005)
4. Badaloni, S., Falda, M., Giacomin, M.: Integrating quantitative and qualitative constraints in fuzzy temporal networks. AI Communications 17, 183–272 (2004)
5. Badaloni, S., Giacomin, M.: The algebra IA^{fuz}: a framework for qualitative fuzzy temporal reasoning. Artificial Intelligence 170, 872–908 (2006)
6. Badaloni, S., Falda, M.: Discriminating exanthematic diseases from temporal patterns of patient symptoms. In: Miksch, S., Hunter, J., Keravnou, E.T. (eds.) AIME 2005. LNCS, vol. 3581, pp. 33–42. Springer, Heidelberg (2005)
7. Badaloni, S., Falda, M.: Temporal characterization of ill-known diseases. In: Proc. of the International Workshop on Intelligent Data Analysis in Biomedicine and Pharmacology, Verona, Italy, pp. 17–22 (2006)

8. Jernigan, J.A., Stephens, D.S., Ashford, D.A., Omenaca, C., Topiel, M.S., Galbraith, M., Tapper, M., Fisk, T.L., Zaki, S., Popovic, T., Meyer, R.F., Quinn, C.P., Harper, S.A., Fridkin, S.K., Sejvar, J.J., Shepard, C.W., McConnell, M., Guarner, J., Shieh, W.J., Malecki, J.M., Gerberding, J.L., Hughes, J.M., Perkins, B.A.: Bioterrorism investigation team. bioterrorism-related inhalational anthrax: the first 10 cases reported in the united states. Emerg Infect Dis. 7, 933–944 (2001)
9. Bellamy, R.J., Freedman, A.R.: Bioterrorism. QJ Med. 94, 227–234 (2001)
10. Dennis, D.T., Inglesby, T.V., Henderson, D.A., et al.: Tularemia as a biological weapon: medical and public health management. Journal of American Medical Association (JAMA) 285, 2763–2773 (2001)
11. Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM 26, 832–843 (1983)
12. Dechter, R., Meiri, I., Pearl, J.: Temporal constraint networks. Artificial Intelligence 49, 61–95 (1991)
13. Dubois, D., Lang, J., Prade, H.: Timed possibilistic logic. Fundamenta Informaticae 15, 211–234 (1991)
14. Marín, R., Cárdenas, M.A., Balsa, M., Sanchez, J.L.: Obtaining solutions in fuzzy constraint network. International Journal of Approximate Reasoning 16, 261–288 (1997)
15. Godo, L., Vila, L.: Possibilistic temporal reasoning based on fuzzy temporal constraints. In: Proc. of IJCA 1995, pp. 1916–1922 (2001)
16. Khatib, L., Morris, P.H., Morris, R.A., Rossi, F.: Temporal constraint reasoning with preferences. In: Proc. IJCAI 2001, Seattle, WA, pp. 322–327 (2001)
17. Badaloni, S., Giacomin, M.: Fuzzy extension of interval-based temporal sub-algebras. In: Proc. of IPMU 2002, Annecy, France, pp. 1119–1126 (2002)
18. Meiri, I.: Combining qualitative and quantitative constraints in temporal reasoning. Artificial Intelligence 87, 343–385 (1996)
19. van Beek, P., Cohen, R.: Exact and approximate reasoning about temporal relations. Computational Intelligence 6, 132–144 (1990)
20. Wilkening, D.A.: Sverdlovsk revisited: Modeling human inhalation anthrax. In: Proc. of the National Academy of Sciences (PNAS), vol. 102 (20), pp. 7589–7594 (2006)
21. Inglesby, T.V., Dennis, D.T., Henderson, D.A., et al.: Plague as a biological weapon: medical and public health management. Journal of American Medical Association (JAMA) 283, 2281–2290 (2000)

Author Index