Anna Maria Fanelli
Witold Pedrycz
Alfredo Petrosino (Eds.)

LNAI 6857

Fuzzy Logic and Applications

9th International Workshop, WILF 2011
Trani, Italy, August 2011
Proceedings

Springer

Lecture Notes in Artificial Intelligence 6857

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

LNAI Founding Series Editor

Anna Maria Fanelli Witold Pedrycz
Alfredo Petrosino (Eds.)

Fuzzy Logic and Applications

9th International Workshop, WILF 2011
Trani, Italy, August 29-31, 2011
Proceedings

Springer

Volume Editors

Anna Maria Fanelli
University of Bari
Computer Science Department
via E. Orabona 4, 70126, Bari, Italy
E-mail: fanelli@di.uniba.it

Witold Pedrycz
University of Alberta
Department of Electrical and Computer Engineering
Edmonton, AB Alberta T6G 2V4, Canada
E-mail: pedrycz@ee.ualberta.ca

Alfredo Petrosino
University of Naples Parthenope
Department of Applied Science
Centro Direzionale Is. C4, 80143 Naples, Italy
E-mail: petrosino@uniparthenope.it

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-23712-6 e-ISBN 978-3-642-23713-3
DOI 10.1007/978-3-642-23713-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011935118

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

LNCS Sublibrary: SL 7 – Artificial Intelligence

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 9th International Workshop on Fuzzy Logic and Applications, WILF 2011, held in Trani (Italy), during August 29–31, 2011, covered topics related to theoretical and experimental areas of fuzzy sets and systems with emphasis on different applications.

This event represents the pursuance of an established tradition of biannual interdisciplinary meetings. The previous editions of WILF have been held, with an increasing number of participants, in Naples (1995), Bari (1997), Genoa (1999), Milan (2001), Naples (2003), Crema (2005), Camogli (2007), and Palermo (2009). Each event aimed to highlight connections and to stress synergies of fuzzy sets theory with soft 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 and foster multidisciplinary research.

WILF 2011 received more than 50 paper submissions from all over the world. A rigorous peer-review selection process was applied to ultimately select about 30 high-quality manuscripts to be published in this volume. In addition to regular papers, WILF 2011 included two tutorials:

- Francesco Marcelloni (University of Pisa, Italy), "Multi-Objective Evolutionary Fuzzy Systems"
- Javier Montero (Complutense University, Madrid, Spain), "Uncertainty and Ignorance in Decision Modelling"

There were also three keynote talks:

- Janusz Kacprzyk (Polish Academy of Sciences), "Towards Human/Social-Inspired Computation: The Role of Computing with Words and Fuzzy Logic"
- Witold Pedrycz (University of Alberta, Edmonton, Canada), "From Fuzzy Models to Granular Fuzzy Models"
- Luis Magdalena (European Centre for Soft Computing, Mieres, Spain), "Some Open Questions in Fuzzy Rule-Based Systems Design"

Among them, the present volume includes contributions by Witold Pedrycz (University of Alberta, Canada) and Francesco Marcelloni (University of Pisa, Italy).

The success of this workshop is to be credited to the effort of many people, in particular to the Program Committee members for their commitment to providing high quality, constructive reviews, to the local Organizing Committee for the support in the organization of all the workshop events and to all the authors for their valuable presentations.

Finally, and mainly, we would like to express our deepest gratitude to Giovanna Castellano for her truly superb contributions to the organization of this very special workshop.

June 2011

Anna Maria Fanelli
Witold Pedrycz
Alfredo Petrosino

Organization

Conference Chairs

Anna Maria Fanelli	University of Bari, Italy
Witold Pedrycz	University of Alberta, Canada
Alfredo Petrosino	University of Naples Parthenope, Italy

Program Committee

Andrzej Bargiela	University of Nottingham, Malaysia Campus, Malaysia
Isabelle Bloch	ENST-CNRS, Paris, France
Piero P. Bonissone	GE CRD, Schenectady, USA
Patrick Bosc	Université de Rennes I, IRISA/ENSSAT, France
Ashish Ghosh	Indian Statistical Institute, Kolkata, India
Francisco Herrera	University of Granada, Spain
Katsuhiro Honda	Osaka Prefecture University, Japan
Eyke Hllermeier	Marburg University, Germany
Janusz Kacprzyk	Polish Academy of Sciences, Poland
Etienne E. Kerre	University of Ghent, Belgium
Erich Peter Klement	Johannes Kepler University, Linz, Austria
Malay Kumar Kundu	Indian Statistical Institute, Kolkata, India
Vincenzo Loia	University of Salerno, Italy
Luis Magdalena	European Centre for Soft Computing, Mieres, Spain
Francesco Marcelloni	University of Pisa, Italy
Sushmita Mitra	Indian Statistical Institute, Kolkata, India
Sankar Kumar Pal	Indian Statistical Institute, Kolkata, India
Enrique Ruspini	European Centre for Soft Computing, Mieres, Spain
Settimo Termini	University of Palermo and CNR, Italy
Enric Trillas	European Centre for Soft Computing, Mieres, Spain
Ronald Yager	Iona College, New Rochelle, USA

Steering Committee

Andrea Bonarini	Politechnic Institute of Milan, Italy
Vito Di Gesù	University of Palermo, Italy
Antonio Di Nola	University of Salerno, Italy
Francesco Masulli	University of Genoa, Italy
Gabriella Pasi	University of Milan Bicocca, Italy
Alfredo Petrosino	University of Naples Parthenope, Italy

Scientific Secretary

Giovanna Castellano	University of Bari, Italy
Ciro Castiello	University of Bari, Italy
Corrado Mencar	University of Bari, Italy
Maria A. Torsello	University of Bari, Italy

Sponsoring Institutions

Dipartimento di Informatica (DIB), Università di Bari, Italy
Università degli Studi di Bari, Italy
Università degli Studi di Napoli "Parthenope", Italy
European Society for Fuzzy Technology Logic and Technology (EUSFLAT)
Italian Group of Italian Researchers in Pattern Recognition (GIRPR)

Under the patronage of

Provincia di Barletta - Andria - Trani, Italy
Città di Trani, Italy

Acknowledgment

The work was supported by MIUR- FIRB Project IntelliLogic (cod. RBIPO6MMBW) funded by Minister of Education, University and Research of Italy, and partly by SOAF and AMIAV Projects, funded by Misura 3.17, Campania Region of Italy.

Table of Contents

Advances in Theory of Fuzzy Sets

Advances in Fuzzy Systems

Advances in Classification and Clustering

Advanced Applications

Solutions of Equation $I(x, y) = I(x, I(x, y))$ for Implications Derived from Uninorms

Sebastia Massanet and Joan Torrens

Department of Mathematics and Computer Science
University of the Balearic Islands
07122 Palma de Mallorca, Spain
{s.massanet,dmijts0}@uib.es

Abstract. Uninorms are one of the most studied classes of aggregation functions and with more applications in the field of the aggregation of information. Their conjunctive or disjunctive behaviour is essential for their use as logical connectives and for obtaining fuzzy implications derived from uninorms. In this communication, we want to analyse which fuzzy implications derived from uninorms satisfy the iterative equation $I(x, y) = I(x, I(x, y))$. This equation comes from $p \to q \equiv p \to (p \to q)$, a tautology in classical logic, and it is related with the law of importation respect to the minimum $I(\min\{x, y\}, z) = I(x, I(y, z))$.

Keywords: Fuzzy implication, uninorm, functional equation.

1 Introduction

In fuzzy logic, the fuzzy implication is one of the most important operations as the classical implication is in the classical logic. These operators are essential in fuzzy control and approximate reasoning, as well as in many the fields where these theories are applied. This is because they are used not only to model fuzzy conditionals, but also to make inferences in any fuzzy rule based system ([6]) through the Modus Ponens and Modus Tollens. However, the fact that many authors have focused their interest in the theoretical study of fuzzy implications as highlight the survey [10] and also the recent book [2], exclusively devoted to fuzzy implications is based on a wider range of applications.

There exist many equations with fuzzy implications involved, often derived from equivalences in classical logic, that has been studied deeply (see [2] and [10]), to obtain fuzzy implications with concrete properties. One of these equations is the following one

$$I(x, y) = I(x, I(x, y)) \quad \text{for all} \quad x, y \in [0, 1], \tag{1}$$

where I denotes a fuzzy implication. This equation comes from the tautology in classical logic

$$p \to (p \to q) \equiv p \to q,$$

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 1–8, 2011.

and it has been already studied for R, (S,N)-implications and QL-operators in [15] and for D-operators in [16]. In the case of D-operators, some remarks and clarifications on the results has been made in [11]. Although fuzzy implications are usually built from the basic logical connectives, i.e., t-norms, t-conorms and negations, more recently after the introduction of the uninorms in [17] as generalizations of t-norms and t-conorms, these ones have become a rich source from which to build fuzzy implications ([2], [4], [9], [14]). In addition, uninorms have applications in many fields like information aggregation, neural networks, mathematical morphology, logical connectives and fuzzy logic ([3,4,5], [7], [9],[13,14], [17]). The goal of this communication is to make a similar study about the satisfaction of Equation (1), similar to the one made until now, for implications derived from uninorms. This study will allow to investigate more in this equation that is related to the law of importation with respect to the minimum

$$I(\min\{x,y\},z) = I(x,I(y,z)).$$

It is obvious that this law of importation implies the satisfaction of the equation we want to study. Recall that the law of importation is stronger than the exchange principle in general and equal to when it holds $I(x,0) = N(x)$, with N a continuous negation ([12]). These properties on fuzzy implications are among the most important ones in fuzzy logic.

The communication is organized as follows. After some preliminaries introducing the necessary notation, Equation (1) is studied in Section 3, structured depending on the class of implications derived from uninorms considered. Finally, in Section 4 there are some conclusions and future work we want to share.

2 Preliminaries

We will suppose the reader to be familiar with the theory of t-norms and t-conorms (all necessary results and notations can be found in [7]) and uninorms (see [5] and Chapter 5 in [2]). Some classes of uninorms have been characterized like uninorms of $U_{\min}$ and $U_{\max}$ ([5]), idempotent uninorms ([3], [8]) and representable ones ([5]). Let us recall this last class that will be used more extensively in this paper.

Definition 1. *A uninorm U with neutral element $e \in (0,1)$ is representable if there exists a continuous and strictly increasing function $h : [0,1] \to [-\infty,+\infty]$ (called additive generator of U), with $h(0) = -\infty$, $h(e) = 0$ and $h(1) = +\infty$ such that U is given by*

$$U(x,y) = h^{-1}(h(x) + h(y))$$

for all $(x,y) \in [0,1]^2 \setminus \{(0,1),(1,0)\}$ and or $U(0,1) = U(1,0) = 0$ or $U(0,1) = U(1,0) = 1$.

The most accepted definition of fuzzy implication is the following one.

Definition 2. *A binary operator $I : [0,1]^2 \to [0,1]$ is a fuzzy implication if satisfies:*

(I1) $I(x,z) \geq I(y,z)$ *when* $x \leq y$, *for all* $z \in [0,1]$.
(I2) $I(x,y) \leq I(x,z)$ *when* $y \leq z$, *for all* $x \in [0,1]$.
(I3) $I(0,0) = I(1,1) = 1$ *and* $I(1,0) = 0$.

Note that, from the definition, it follows that $I(0,x) = 1$ and $I(x,1) = 1$ for all $x \in [0,1]$ whereas the symmetrical values $I(x,0)$ and $I(1,x)$ are not derived from it. Some specially interesting properties of implications are:

– The *exchange principle*,

$$I(x, I(y,z)) = I(y, I(x,z)), \quad \text{for all } x,y,z \in [0,1]. \tag{EP}$$

– The *ordering property* for uninorms,

$$x \leq y \iff I(x,y) \geq e, \quad \text{for all } x,y \in [0,1], \text{ where } e \in (0,1). \tag{$\mathbf{OP_U}$}$$

Fuzzy implications can be derived of several classes of aggregation functions. In the following sections, we will use those most used derived from uninorms:

– (U,N)-implications derived from a disjunctive uninorm U and a negation N,

$$I_{U,N}(x,y) = U(N(x), y), \quad \text{for all } x,y \in [0,1].$$

– RU-implications derived from a uninorm U such that $U(x,0) = 0$ for all $x < 1$,

$$I_U(x,y) = \sup\{z \in [0,1] \mid U(x,z) \leq y\}, \quad \text{for all } x,y \in [0,1].$$

– QL-operators derived from a conjunctive uninorm U with neutral element e, a disjunctive uninorm $U^{'}$ with neutral element e' and a negation N (usually strong),

$$I_Q(x,y) = U^{'}(N(x), U(x,y)), \quad \text{for all } x,y \in [0,1].$$

(U,N) and RU-implications are always implications in the sense of Definition 2, while QL-operators are not implications in general. Finally recall the characterization of RU-implications and their structure when the uninorm is idempotent.

Theorem 1. ([1]) *Let $I : [0,1]^2 \to [0,1]$ be a function. Then I is a RU-implication derived from a left-continuous uninorm U with neutral element $e \in (0,1)$, if and only if, I satisfies (I2), (OP_U), (EP) and I right-continuous on the second variable.*

Moreover, in this case the uninorm U must be conjunctive and is given by

$$U(x,y) = \inf\{z \in [0,1] | I(x,z) \geq y\}.$$

Theorem 2. ([13]) *Let $U = (e,g)$ be any idempotent uninorm with neutral element e and g its associated decreasing function with $g(0) = 1$. The residual implication I_U is given by:*

$$I_U(x,y) = \begin{cases} \min(g(x), y) & \text{if } y < x, \\ \max(g(x), y) & \text{if } y \geq x. \end{cases}$$

3 Main Results

In this section, we want to characterize which implications I derived from uninorms from the classes introduced in Section 2 satisfy Equation (11).

3.1 (U, N) and RU-Implications

Theorem 3. *A (U, N)-implication I generated from a disjunctive uninorm U and a negation N satisfies Equation (7) if and only if the range of N is a subset of the idempotent elements of U.*

Proof. It is similar to the case of (S, N)-implications (see Theorem 10 in [15]) taking into account that in this case, $I(x, e) = N(x)$ holds. □

The following result is an immediate consequence.

Corollary 1. *A (U, N)-implication I generated from a disjunctive uninorm U and a continuous negation N satisfies Equation (7) if and only if U is an idempotent uninorm.*

From the study of Equation (11) for this class of implications, a new characterization of RU-implications derived from left-continuous idempotent uninorms is obtained.

Theorem 4. *A binary function $I : [0, 1]^2 \to [0, 1]$ satisfies (I2), (EP), (OP_U), right-continuity on the second argument and Equation (7) if and only if I is a RU-implication derived from a left-continuous idempotent uninorm.*

Proof. Let I be a binary function satisfying (I2), (EP), (OP_U), right-continuity on the second argument and Equation (11). By Theorem 1, the first four properties ensure that a conjunctive left-continuous uninorm can be generated trough:

$$U(x, y) = \inf\{t \in [0, 1] | I(x, t) \geq y\},$$

for all $x, y \in [0, 1]$, and if we denote $I_U(x, y) = \sup\{t \in [0, 1] | U(x, t) \leq y\}$, then $I = I_U$.

We will prove that U must be idempotent. For all $x \in [0, 1]$, $U(x, x) = \inf\{t \in [0, 1] | I(x, t) \geq x\}$. Let us suppose that there exists $t_0 \in [0, x)$ such that $I(x, t_0) \geq x$. Then by (OP_U), $I(x, I(x, t_0)) \geq e$. Since we know that I satisfies Equation (11), $I(x, t_0) = I(x, I(x, t_0)) \geq e$ and by (OP_U), $x \leq t_0$ and we have a contradiction. This proves $U(x, x) \geq x$. Now, using Equation (11) and (OP_U), we have $I(x, I(x, x)) = I(x, x) \geq e \Rightarrow x \leq I(x, x)$ and therefore,

$$U(x, x) = \inf\{t \in [0, 1] | I(x, t) \geq x\} \leq x$$

for all $x \in [0, 1]$. Thus, U is an idempotent uninorm.

Reciprocally, if I is a RU-implication derived from a left-continuous idempotent uninorm, it satisfies the first four properties by Theorem 1. We need only prove Equation (11). By Theorem 2, we obtain two cases:

– If $I(x,y) = y$ then Equation (1) holds trivially.
– Otherwise, $I(x,y) = g(x)$ and Equation (1) also holds since in this case, we have $I(x, I(x,y)) = I(x, g(x)) = g(x)$. □

Corollary 2. *A RU-implication I generated from a left-continuous uninorm U satisfies Equation (1) if and only if U is idempotent.*

Proof. Let I be a RU-implication generated from a left-continuous idempotent uninorm U. Then by the previous theorem, I satisfies Equation (1). Reciprocally, a RU-implication I satisfies (I2), (EP), (OP_U) and is right-continuous on the second variable, therefore if Equation (1) holds, by the previous theorem, U must be idempotent. □

3.2 *QL*-Operators

In this case, we will start with a general result for this type of operators. After that, the QL-implications satisfying Equation (1) are studied and finally, some particular cases depending on the class of uninorm chosen to generate the QL-operator are analysed. From now on, e and e' denote the neutral elements of U and U' respectively.

Proposition 1. *Let I be a QL-operator derived from a disjunctive uninorm U', a conjunctive uninorm U and a negation N, that satisfies Equation (1). Then $U'(N(e), N(e)) = N(e)$. Moreover, if N is strong, then $U(N(e'), N(e')) = N(e')$.*

Proof. Taking $x = e$ and $y = e'$ in Equation (1), we obtain that $N(e)$ must be an idempotent element of U' and now taking $x = N(e')$ and $y = e$ and applying $N(N(e')) = e'$ since N is strong, we obtain the same property with $N(e')$. □

***QL*-implications.** In this case, we will conclude that the only solutions of Equation (1) are QL-implications derived from t-norms and t-conorms and among them, there is only one.

Proposition 2. *Let I be a QL-implication derived from a disjunctive uninorm U' with continuous associated t-norm and t-conorm, a conjunctive uninorm U and a negation N, satisfying Equation (1). Then U' must be an Archimedean t-conorm and U a t-norm.*

Proof. In [9] it is proved that if $I(x,y) = U'(N(x), U(x,y))$ is an implication then U' must be an Archimedean t-conorm. Now, taking $x = e$ and $y = 0$ in Equation (1) we have $S(N(e), N(e)) = N(e)$ and consequently, $N(e) = 0$. Since $N \geq N_S$ and N_S is a strong negation, $e = 1$. Thus, U is a t-norm. □

Thus, this case reduces to the case of QL-operators derived from Archimedean t-conorms and t-norms studied in [15], see Theorem 15. However, the concrete case of QL-implications is not studied separately and as the following theorem proves, there is only one QL-implication that satisfies Equation (1).

Theorem 5. *Let $S = (S_L)_\varphi$ be a nilpotent t-conorm, $T = (T_L)_\phi$ a nilpotent t-norm and N a strong negation such that the corresponding QL-operator I satisfies Equation (1). Then the following statements are equivalent:*

1. *I is an implication.*
2. *$(N_C)_\varphi = N = (N_C)_\phi$.*
3. *I is the Kleene-Dienes implication.*

Proof. First we prove 1. $\Rightarrow$ 2.. Let I be an implication. Then, by Theorem 15 in [15], we have $(N_C)_\varphi \leq N \leq (N_C)_\phi$. In addition,

$$T(x,y) = 0 \iff x \leq (N_C)_\phi(y). \tag{2}$$

Now, let us suppose that there exists some $N(x_0)$ such that $(N_C)_\varphi(N(x_0)) < N(N(x_0))$. Since T is continuous there exists $y_0 > 0$ such that $T(x_0, y_0) = (N_C)_\varphi(N(x_0))$ and then, by Theorem 15 in [15], $I(x_0, y_0) = 1$. Meanwhile, by the same theorem and Equation (2), $I(x, y_0) = N(x)$ for all $x \leq (N_C)_\phi(y_0)$ violating condition (I1). This is, $N = (N_C)_\varphi$.

Now, let us suppose that there exists x_0 such that $(N_C)_\phi(x_0) > N(x_0)$. Let y_0 be such that $N(x_0) < y_0 \leq (N_C)_\phi(x_0)$. Then $I(x_0, y_0) = N(x_0)$ since $y_0 \leq (N_C)_\phi(x_0)$ (see again Theorem 15 in [15]), but $I(1, y_0) = y_0$, violating condition (I1). This is, $N = (N_C)_\phi$.

After that 2. $\Rightarrow$ 3. is a straightforward calculation from the expression of the QL-operator in Theorem 15 in [15] and 3. $\Rightarrow$ 1. is trivial. □

Case of QL-operators with U' representable. Now, we return to the general case of QL-operators depending on the uninorm chosen to generate them we want to solve Equation (1). In this paper we reduce our study to the case of U' a representable uninorm. We will start when U' is representable.

Proposition 3. *Let I be a QL-operator derived from a disjunctive representable uninorm U', a conjunctive uninorm U and a strong negation N. If I satisfies Equation (1), then necessarily $e = N(e')$.*

Proof. If U' is representable, it does not have non-trivial idempotent elements and by Proposition 1, we obtain $N(e) = 0, 1, e'$, i.e., $e = 0, 1, N(e')$. Since U is conjunctive, $e \neq 0$ and therefore, two cases are possible: $e = 1$ or $e = N(e')$. Let us prove that U cannot be a continuous t-norm T. Let h' be the additive generator of U' and let us suppose that I satisfies Equation (1). So,

$$h'^{-1}(h'(N(x)) + h'(T(x, U'(N(x), T(x,y))))) = h'^{-1}(h'(N(x)) + h'(T(x,y)))$$

must hold and consequently, when $T(x,y) \neq 0, 1$ and $x \neq 0, 1$,

$$T(x, U'(N(x), T(x,y))) = T(x,y) \tag{3}$$

must hold. Taking x, y such that $0 < y < N(e') < x < 1$ by Proposition 1, $T(x,y) = \min\{x,y\} = y$ and since $x > N(e')$ we have $N(x) < e'$ and $h'(N(x)) < 0$, therefore

$$U'(N(x), y) = h'^{-1}(h'(N(x)) + h'(y)) < y.$$

On the other hand, from Equation (3), we obtain $U'(N(x), y) = y$, a contradiction. So, necessarily $e = N(e')$. □

Consequently, we will analyse the case $e = N(e')$. Thus, cases are considered depending on the class of uninorm to which the conjunctive uninorm belongs. The proof of the following result is not included due to lack of space.

Proposition 4. *Let U' be a disjunctive representable uninorm with additive generator h', U a conjunctive uninorm and N a strong negation such that $e = N(e')$. Then the corresponding QL-operator satisfies Equation (1) if and only if U is representable with additive generator h such that $f = h' \circ h^{-1}$ is given by*

$$f(x) = cx + h'(e)$$

with $c \in \mathbb{R}^+$ constant. In this case, the resulting QL-operator is given by

$$I(x,y) = \begin{cases} 1 \text{ if } (x = 0) \text{ or } (x = 1, y > 0), \\ y \text{ otherwise.} \end{cases}$$

Remark 1. Note that the previous result ensures that, when we consider the disjunctive uninorm U' representable, there are only solutions if the conjunctive one U is also representable. There are no solutions when U belongs to $U_{\min}$ or is idempotent. However, all the conjunctive representable uninorms satisfying the conditions of the previous proposition generate the same implication. So, we only obtain one solution in this case.

4 Conclusions and Future Work

In this work, the iterative equation $I(x,y) = I(x, I(x,y))$ has been studied for the major classes of implications derived from uninorms. Thus, the (U,N) and RU-implications satisfying the studied equation have been completely characterized, obtaining a new axiomatic characterization based on Equation (1) of RU-implications derived from left-continuous idempotent uninorms. In addition, it has been proved that the only QL-implication satisfying the studied equation is the Kleene-Dienes implication. In the case of QL-operators in general, the study when the disjunctive uninorm is representable has been completed, determining that we only obtain solutions when the conjunctive uninorm is also representable.

It remains to study the cases when the disjunctive uninorm is idempotent or belongs to $U_{\max}$. Finally, it would be interesting to study the relationship between the obtained solutions for this equation and the ones of the law of importation with $T = \min$.

Acknowledgements. This paper has been partially supported by the Spanish Grant MTM2009-10320 with FEDER support.

References

1. Aguiló, I., Suñer, J., Torrens, J.: A characterization of residual implications derived from left-continuous uninorms. Information Sciences 180(20), 3992–4005 (2010)
2. Baczyński, M., Jayaram, B.: Fuzzy Implications. Springer, Heidelberg (2008)
3. De Baets, B.: Idempotent uninorms. European Journal of Operational Research 118(3), 631–642 (1999)
4. De Baets, B., Fodor, J.C.: Residual operators of uninorms. Soft Computing 3, 89–100 (1999)
5. Fodor, J.C., Yager, R.R., Rybalov, A.: Structure of uninorms. Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems 5, 127–411 (1997)
6. Kerre, E., Huang, C., Ruan, D.: Fuzzy Set Theory and Approximate Reasoning. Wu Han University Press, Wu Chang (2004)
7. Klement, E., Mesiar, R., Pap, E.: Triangular norms. Kluwer Academic Publishers, Dordrecht (2000)
8. Martin, J., Mayor, G., Torrens, J.: On locally internal monotonic operators. Fuzzy Sets and Systems 137, 27–42 (2003)
9. Mas, M., Monserrat, M., Torrens, J.: Two types of implications derived from uninorms. Fuzzy Sets and Systems 158, 2612–2626 (2007)
10. Mas, M., Monserrat, M., Torrens, J., Trillas, E.: A survey on fuzzy implication functions. IEEE Transactions on Fuzzy Systems 15(6), 1107–1121 (2007)
11. Massanet, S., Torrens, J.: Some remarks on the solutions to the functional equation $I(x,y) = I(x, I(x,y))$ for D-operations. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods. Communications in Computer and Information Science, vol. 80, pp. 666–675. Springer, Heidelberg (2010)
12. Massanet, S., Torrens, J.: The law of importation versus the exchange principle on fuzzy implications. Fuzzy Sets and Systems 168(1), 47–69 (2011)
13. Ruiz, D., Torrens, J.: Residual implications and co-implications from idempotent uninorms. Kybernetika 40, 21–38 (2004)
14. Ruiz-Aguilera, D., Torrens, J.: R-implications and S-implications from uninorms continuous in $]0,1[^2$ and their distributivity over uninorms. Fuzzy Sets and Systems 160, 832–852 (2009)
15. Shi, Y., Ruan, D., Kerre, E.: On the characterization of fuzzy implications satisfying $I(x,y) = I(x, I(x,y))$. Information Sciences 117, 2954–2970 (2007)
16. Xie, A., Qin, F.: Solutions of the functional equation $I(x,y) = I(x, I(x,y))$ for a continuous D-operation. Information Sciences 180, 2487–2497 (2010)
17. Yager, R., Rybalov, A.: Uninorm aggregation operators. Fuzzy Sets and Systems 80, 111–120 (1996)

On the Behavior of WOWA Operators

Bonifacio Llamazares

University of Valladolid, PRESAD Research Group, 47011 Valladolid, Spain
boni@eco.uva.es
http://www2.eco.uva.es/boni

Abstract. In this paper we analyze the behavior of WOWA operators, a class of functions that simultaneously generalize weighted means and OWA operators. Moreover, we introduce functions that also generalize both operators and characterize those satisfying a condition imposed to maintain the relationship among the weights.

Keywords: WOWA operators, OWA operators, weighted means.

1 Introduction

Weighted means and ordered weighted averaging (OWA) operators (Yager [8]) are aggregation functions widely used in the literature. Weighted means allow to weight each information source in relation to their reliability while OWA operators allow to weight the values according to their ordering. The need to combine both functions has been reported by several authors (see, among others, Torra [4] and Torra and Narukawa [6]). In [4], Torra introduces the weighted OWA (WOWA) operator, a new aggregation function that allows to combine both weights.

The aim of this paper is to analyze the behavior of WOWA operators. Moreover, since, in some cases, the results provided by these operators may be questionable, we propose to use functions that maintain the relationship among the weights of a weighting vector when the non-zero components of the other weighting vector are equal. In this way, we obtain a class of functions that have been previously introduced by Engemann *et al.* [1] in a different framework.

The paper is organized as follows. In Section 2 we introduce weighted means, OWA operators and WOWA operators. Section 3 shows some questionable behaviors of WOWA operators. In Section 4 we propose a condition to maintain the relationship among the weights and characterize the functions that satisfy this condition. The paper concludes in Section 5.

2 Preliminaries

Throughout the paper we will use the following notation: vectors will be denoted in bold; $\boldsymbol{\eta}$ will denote the vector $(1/n, \ldots, 1/n)$; $\boldsymbol{x} \geq \boldsymbol{y}$ will mean $x_i \geq y_i$ for all $i \in \{1, \ldots, n\}$; given σ a permutation of $\{1, \ldots, n\}$, $\boldsymbol{x}_\sigma$ will denote the vector $(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 9–16, 2011.

In the following definition we present some well-known properties usually demanded to the functions used in the aggregation processes.

Definition 1. *Let $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function.*

1. *F is symmetric if $F(\boldsymbol{x}_\sigma) = F(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and for all permutation σ of $\{1, \ldots, n\}$.*
2. *F is monotonic if $F(\boldsymbol{x}) \geq F(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ such that $\boldsymbol{x} \geq \boldsymbol{y}$.*
3. *F is idempotent if $F(x, \ldots, x) = x$ for all $x \in \mathbb{R}$.*
4. *F is compensative if $\min(\boldsymbol{x}) \leq F(\boldsymbol{x}) \leq \max(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*
5. *F is homogeneous of degree 1 if $F(\lambda \boldsymbol{x}) = \lambda F(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and for all $\lambda > 0$.*

2.1 Weighted Means and OWA Operators

Weighted means and OWA operators are defined by vectors with non-negative components whose sum is 1.

Definition 2. *A vector $\boldsymbol{\mu} \in \mathbb{R}^n$ is a weighting vector if $\boldsymbol{\mu} \in [0,1]^n$ and $\sum_{i=1}^{n} \mu_i = 1$.*

Definition 3. *Let $\boldsymbol{p}$ be a weighting vector. The weighted mean associated with $\boldsymbol{p}$ is the function $F_{\boldsymbol{p}} : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by*

$$F_{\boldsymbol{p}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} p_i x_i.$$

The weighted means are monotonic, idempotent, compensative and homogeneous of degree 1 functions.

In [8], Yager introduced OWA operators as a tool for aggregation procedures in multicriteria decision making.

Definition 4. *Let $\boldsymbol{w}$ be a weighting vector. The OWA operator associated with $\boldsymbol{w}$ is the function $F^{\boldsymbol{w}} : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by*

$$F^{\boldsymbol{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i x_{\sigma(i)},$$

where σ is a permutation of $\{1, \ldots, n\}$ such that $x_{\sigma(1)} \geq \cdots \geq x_{\sigma(n)}$.

OWA operators are symmetric, monotonic, idempotent, compensative and homogeneous of degree 1 functions.

One of the most important issues in the theory of OWA operators is the determination of associated weights (see, for instance, Xu [7] and Fullér [2]). In [9], Yager relates the OWA operators weights to quantifiers.

Definition 5. *A quantifier is a non-decreasing function $Q : [0,1] \longrightarrow [0,1]$ that satisfies $Q(0) = 0$ and $Q(1) = 1$.*

Given a quantifier Q, the OWA operator weights can be obtained from the expression $w_i = Q\left(i/n\right) - Q\left(\left(i-1\right)/n\right), \quad i = 1, \ldots, n$, (Yager [9]). From this relation follows:

$$Q\left(\frac{i}{n}\right) = \sum_{j=1}^{i} w_j, \quad i = 1, \ldots, n;$$

i.e., the same weighting vector can be obtained through any quantifier interpolating the points $\left(i/n, \sum_{j=1}^{i} w_j\right)$, $i = 1, \ldots, n$.

2.2 WOWA Operators

WOWA operators were introduced by Torra [4] in order to consider situations where both the importance of information sources and the importance of values had to be taken into account.

Definition 6. *Let $\boldsymbol{p}$ and $\boldsymbol{w}$ be two weighting vectors. The WOWA operator associated with $\boldsymbol{p}$ and $\boldsymbol{w}$ is the function $W_{\boldsymbol{p}}^{\boldsymbol{w}} : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by*

$$W_{\boldsymbol{p}}^{\boldsymbol{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} \mu_i x_{\sigma(i)},$$

where σ is a permutation of $\{1, \ldots, n\}$ such that $x_{\sigma(1)} \geq \cdots \geq x_{\sigma(n)}$ and the weight μ_i is defined as

$$\mu_i = f\left(\sum_{j=1}^{i} p_{\sigma(j)}\right) - f\left(\sum_{j=1}^{i-1} p_{\sigma(j)}\right),$$

where f is a non-decreasing function that interpolates the points $\left(i/n, \sum_{j=1}^{i} w_j\right)$ together with the point $(0, 0)$. Moreover, f is the identity when the points can be interpolated in this way.

Any quantifier generating the weighting vector $\boldsymbol{w}$ satisfies the required properties of the function f in the previous definition (under the assumption that the quantifier is the identity when $\boldsymbol{w} = \boldsymbol{\eta}$). For this reason, it is possible to give an alternative definition of WOWA operators using quantifiers (Torra and Godo [5]).

Definition 7. *Let $\boldsymbol{p}$ be a weighting vector and let Q be a quantifier. The WOWA operator associated with $\boldsymbol{p}$ and Q is the function $W_{\boldsymbol{p}}^{Q} : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by*

$$W_{\boldsymbol{p}}^{Q}(x_1, \ldots, x_n) = \sum_{i=1}^{n} \mu_i x_{\sigma(i)},$$

where σ is a permutation of $\{1, \ldots, n\}$ such that $x_{\sigma(1)} \geq \cdots \geq x_{\sigma(n)}$ and the weight μ_i is defined as

$$\mu_i = Q\left(\sum_{j=1}^{i} p_{\sigma(j)}\right) - Q\left(\sum_{j=1}^{i-1} p_{\sigma(j)}\right).$$

WOWA operators are monotonic, idempotent, compensative and homogeneous of degree 1 functions. Moreover, $W_{\boldsymbol{p}}^{\boldsymbol{\eta}} = F_{\boldsymbol{p}}$ and $W_{\boldsymbol{\eta}}^{\boldsymbol{w}} = F^{\boldsymbol{w}}$ (Torra [4]).

3 Analysis of WOWA Operators

In this section we show with examples some questionable behaviors of WOWA operators.

Example 1. Suppose we have five sensors to measure a certain physical property. The sensors are of different quality and precision, so they are weighted according to the weighting vector $\boldsymbol{p} = (0.3, 0.2, 0.2, 0.2, 0.1)$. Moreover, to prevent a faulty sensor alter the measurement, we take the weighting vector $\boldsymbol{w} = (0, 1/3, 1/3, 1/3, 0)$; thus, the maximum and minimum values are not considered.

Given $\boldsymbol{w} = (0, 1/3, 1/3, 1/3, 0)$, we have to choose a quantifier interpolating the points $(0, 0)$, $(0.2, 0)$, $(0.4, 1/3)$, $(0.6, 2/3)$, $(0.8, 1)$ and $(1, 1)$. We consider the quantifier depicted in Figure 1, which is given by

$$Q(x) = \begin{cases} 0 & \text{if } x \leq 0.2, \\ \frac{5}{3}x - \frac{1}{3} & \text{if } 0.2 < x < 0.8, \\ 1 & \text{if } x \geq 0.8. \end{cases}$$

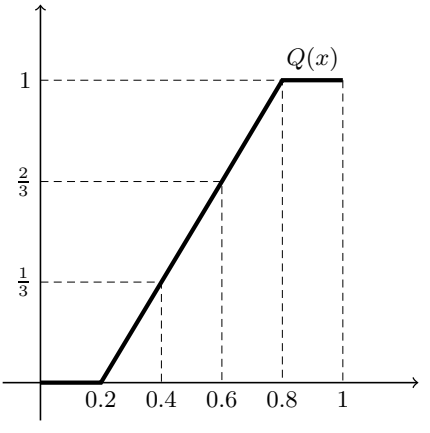


Fig. 1. Quantifier associated to the weighting vector $\boldsymbol{w} = (0, 1/3, 1/3, 1/3, 0)$

Suppose the values obtained by the sensors are $\boldsymbol{x} = (10, 4, 5, 6, 3)$. If σ is a permutation ordering these values in a decrease way, then, in this case, $\boldsymbol{p}_\sigma = \boldsymbol{p} = (0.3, 0.2, 0.2, 0.2, 0.1)$. The weighting vector $\boldsymbol{\mu}$ is

$$\begin{aligned} \mu_1 &= Q(0.3) - Q(0) = 1/6, & \mu_2 &= Q(0.5) - Q(0.3) = 1/3, \\ \mu_3 &= Q(0.7) - Q(0.5) = 1/3, & \mu_4 &= Q(0.9) - Q(0.7) = 1/6, \\ \mu_5 &= Q(1) - Q(0.8) = 0, \end{aligned}$$

and $W_{\boldsymbol{p}}^{\boldsymbol{w}}(10, 4, 5, 6, 3) = 10/6 + 2 + 5/3 + 4/6 = 6$.

However, our intention is not to consider the maximum and minimum values and only take into account the values 4, 5 and 6; which have been provided by sensors with the same weight. Therefore, it seems logical to make the average of these values, in which case we would get 5 as final value.

Example 2. Consider again the situation of the previous example and suppose now that $\boldsymbol{p} = (0.4, 0.2, 0.2, 0.1, 0.1)$ and $\boldsymbol{x} = (10, 3, 5, 6, 7)$. If σ is a permutation ordering these values from the largest to the smallest element, then $\boldsymbol{p}_\sigma = (0.4, 0.1, 0.1, 0.2, 0.2)$. The weighting vector $\boldsymbol{\mu}$ is

$$\begin{aligned} \mu_1 &= Q(0.4) - Q(0) \quad = 1/3, & \mu_2 &= Q(0.5) - Q(0.4) = 1/6, \\ \mu_3 &= Q(0.6) - Q(0.5) = 1/6, & \mu_4 &= Q(0.8) - Q(0.6) = 1/3, \\ \mu_5 &= Q(1) \quad - Q(0.8) = 0, \end{aligned}$$

and $W_{\boldsymbol{p}}^{\boldsymbol{w}}(10, 3, 5, 6, 7) = 10/3 + 7/6 + 1 + 5/3 = 43/6$.

As in the previous example, we do not want to consider the maximum and minimum values and to aggregate the remaining ones, in this case the values 5, 6, and 7. However, the WOWA operator returns a value greater than the three aggregate values because it weights the maximum (10 in this case) with 1/3. On the other hand, there are other interesting properties that the WOWA operator does not satisfy:

1. The value returned by the WOWA operator does not always lie between the values returned by the weighted mean and the OWA operator:

$$\begin{aligned} F^{\boldsymbol{w}}(10, 3, 5, 6, 7) &= 7/3 + 2 + 5/3 = 6, \\ F_{\boldsymbol{p}}(10, 3, 5, 6, 7) &= 4 + 0.6 + 1 + 0.6 + 0.7 = 6.9, \end{aligned}$$

   but $W_{\boldsymbol{p}}^{\boldsymbol{w}}(10, 3, 5, 6, 7) = 43/6$.
2. The value returned by the WOWA operator does not always coincide with the values of the weighted mean and the OWA operator when both are equal:

$$\begin{aligned} F^{\boldsymbol{w}}(8, 2.5, 5, 6, 7) &= 7/3 + 2 + 5/3 = 6, \\ F_{\boldsymbol{p}}(8, 2.5, 5, 6, 7) &= 3.2 + 0.5 + 1 + 0.6 + 0.7 = 6, \end{aligned}$$

   but $W_{\boldsymbol{p}}^{\boldsymbol{w}}(8, 2.5, 5, 6, 7) = 8/3 + 7/6 + 1 + 5/3 = 6.5$.

4 Choosing Functions to Maintain the Relationship among the Weights

As we have seen in the previous section, the results provided by WOWA operators may be questionable. If we consider again Example 2, we want to aggregate the values 5, 6, and 7, which are the values given by the sensors with weights 0.2, 0.1, and 0.1, respectively. One possibility is to weight these values by means of the weighting vector $(0.5, 0.25, 0.25)$. In this way, it is possible to maintain the relationship among the initial weights. The returned value in this case is 23/4.

According to the above remarks, we look for a function $F_{\boldsymbol{p}}^{\boldsymbol{w}} : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by

$$F_{\boldsymbol{p}}^{\boldsymbol{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} \rho_i x_{\sigma(i)},$$

where σ is a permutation of $\{1, \ldots, n\}$ such that $x_{\sigma(1)} \geq \cdots \geq x_{\sigma(n)}$ and the weight ρ_i is defined as

$$\rho_i = \frac{f(w_i, p_{\sigma(i)})}{\sum_{j=1}^{n} f(w_j, p_{\sigma(j)})},$$

where $f : [0,1]^2 \longrightarrow [0,1]$. In this way the weights ρ_i depend on the weights w_i and $p_{\sigma(i)}$.

In order to maintaining the relationship among the weights of a vector ($\boldsymbol{p}$ or $\boldsymbol{w}$) when the non-zero components of the other vector are equal, it is necessary that f satisfies the following condition:

$$f(tx, y) = f(x, ty) = tf(x, y),$$

for all $x, y \in [0,1]$ and $t \in [0, \infty)$ such that $tx, ty \in [0,1]$. In the next proposition we characterize the functions that satisfy this condition.

Proposition 1. *Let $f : [0,1]^2 \longrightarrow [0,1]$ be a function such that $f(tx, y) = f(x, ty) = tf(x, y)$ for all $x, y \in [0,1]$ and $t \in [0, \infty)$ such that $tx, ty \in [0,1]$. Then $f(x, y) = cxy$, where $c \in [0,1]$.*

Proof. Given $x, y \in [0,1]$, $f(x, y) = f(x \cdot 1, y \cdot 1) = xf(1, y \cdot 1) = xyf(1,1)$. □

If $f(x, y) = cxy$, with $c \in [0,1]$, then $\rho_i = \dfrac{w_i p_{\sigma(i)}}{\sum_{j=1}^{n} w_j p_{\sigma(j)}}$; that is,

$$F_{\boldsymbol{p}}^{\boldsymbol{w}}(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} w_i p_{\sigma(i)} x_{\sigma(i)}}{\sum_{j=1}^{n} w_j p_{\sigma(j)}}.$$

It is worth noting that this function has been used by Engemann *et al.* [1] in a framework of decision making under risk and uncertainty (in this case, $\boldsymbol{p}$ is the vector of probabilities of the states of nature).

In order to ensure that $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ is well defined, we need that $w_j p_{\sigma(j)}$ be non-zero for some $j \in \{1, \ldots, n\}$. This requirement is guaranteed by any of the following conditions:

1. The number of non-zero weights in each vector $\boldsymbol{p}$ and $\boldsymbol{w}$ is greater than $n/2$.
2. All the components of $\boldsymbol{p}$ are non-zero.

In addition to this, $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ has another problem in your definition: sometimes, the vector $\boldsymbol{p}_\sigma$ is not unique and $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ may return different values according to the vector $\boldsymbol{p}_\sigma$ used. This fact is illustrated in the following example.

Example 3. Consider $\boldsymbol{p} = (0.5, 0.2, 0.3)$, $\boldsymbol{w} = (0.3, 0.4, 0.3)$, and $\boldsymbol{x} = (7, 5, 7)$. When $\boldsymbol{x}$ is ordered from greatest to least, then we have the vector $(7, 7, 5)$. In this vector, the first component can be associated to the weight 0.5 or 0.3. In the first case, the weighting vector $\boldsymbol{\rho}$ is $(5/11, 4/11, 2/11)$ and

$$F_{\boldsymbol{p}}^{\boldsymbol{w}}(7, 5, 7) = \frac{35}{11} + \frac{28}{11} + \frac{10}{11} = \frac{73}{11}.$$

In the second case, the weighting vector $\boldsymbol{\rho}$ is $(9/35, 4/7, 6/35)$ and

$$F_{\boldsymbol{p}}^{\boldsymbol{w}}(7, 5, 7) = \frac{9}{5} + 4 + \frac{6}{7} = \frac{233}{35}.$$

A similar problem arises in the IOWA operators, introduced by Yager and Filev [10]. The solution proposed by these authors, applied to our framework, is to replace the weights associated to equal values by the average of them. In the previous example we replace the weights $p_1 = 0.5$ and $p_3 = 0.3$ by 0.4. In this case the weighting vector $\boldsymbol{\rho}$ is $(6/17, 8/17, 3/17)$ and

$$F_{\boldsymbol{p}}^{\boldsymbol{w}}(7, 5, 7) = \frac{42}{17} + \frac{56}{17} + \frac{15}{17} = \frac{113}{17}.$$

With regard to the properties satisfied by $F_{\boldsymbol{p}}^{\boldsymbol{w}}$, it is easy to check that $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ is idempotent, compensative, homogeneous of degree 1, and that satisfies $F_{\boldsymbol{p}}^{\boldsymbol{\eta}} = F_{\boldsymbol{p}}$ (Engemann *et al.* [1]) and $F_{\boldsymbol{\eta}}^{\boldsymbol{w}} = F^{\boldsymbol{w}}$.

Nevertheless, as noted by Liu [3], $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ is not monotonic. Moreover, similar to WOWA operators, it does not satisfy other interesting properties as we show in the next example.

Example 4. Consider again the weighting vectors $\boldsymbol{p} = (0.4, 0.2, 0.2, 0.1, 0.1)$ and $\boldsymbol{w} = (0, 1/3, 1/3, 1/3, 0)$. Then, we have:

1. The value returned by $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ does not always lie between the values returned by the weighted mean and the OWA operator:

   $$F^{\boldsymbol{w}}(10, 3, 5, 6, 7) = 6, \quad F_{\boldsymbol{p}}(10, 3, 5, 6, 7) = 6.9, \quad \text{but} \quad F_{\boldsymbol{p}}^{\boldsymbol{w}}(10, 3, 5, 6, 7) = 23/4.$$

2. The value returned by $F_{\boldsymbol{p}}^{\boldsymbol{w}}$ does not always coincide with the value returned by the weighted mean and the OWA operator when both values are the same:

   $$F^{\boldsymbol{w}}(8, 2.5, 5, 6, 7) = 6, \quad F_{\boldsymbol{p}}(8, 2.5, 5, 6, 7) = 6, \quad \text{but} \quad F_{\boldsymbol{p}}^{\boldsymbol{w}}(8, 2.5, 5, 6, 7) = 23/4.$$

5 Concluding Remarks

In this paper we have analyzed WOWA operators and we have shown that, in some applications, these operators do not always provide the expected result. Due to the questionable behavior of these operators, we have imposed a condition to maintain the relationship among the weights and we have characterized the functions that satisfy this condition. However, the obtained functions are not monotonic. So, we can conclude that none of the analyzed functions is fully convincing.

Acknowledgments. This work is partially supported by the Junta de Castilla y León (Consejería de Educación y Cultura, Project GR99), the Spanish Ministry of Science and Innovation (Project ECO2009-07332), and ERDF.

References

1. Engemann, K.J., Filev, D.P., Yager, R.R.: Modelling decision making using immediate probabilities. International Journal of General Systems 24, 281–294 (1996)
2. Fullér, R.: On obtaining OWA operator weights: a short survey of recent developments. In: Fifth IEEE International Conference on Computational Cybernetics (ICCC 2007), Gammarth, Tunisia, pp. 241–244 (2007)
3. Liu, X.: Preference solutions of probability decision making with RIM quantifiers. International Journal of Intelligent Systems 20, 1253–1271 (2005)
4. Torra, V.: The weighted OWA operator. International Journal of Intelligent Systems 12, 153–166 (1997)
5. Torra, V., Godo, L.: Averaging continuous distributions with the WOWA operator. In: Second European Workshop on Fuzzy Decision Analysis and Neural Networks for Management, Planning and Optimization (EFDAN 1997), Dortmund, Germany, pp. 10–19 (1997)
6. Torra, V., Narukawa, Y.: Modeling Decisions: Information Fusion and Aggregation Operators. Springer, Berlin (2007)
7. Xu, Z.S.: An overview of methods for determining OWA weights. International Journal of Intelligent Systems 20, 843–865 (2005)
8. Yager, R.R.: On ordered weighted averaging operators in multicriteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190 (1988)
9. Yager, R.R.: Families of OWA operators. Fuzzy Sets and Systems 59, 125–148 (1993)
10. Yager, R.R., Filev, D.P.: Induced ordered weighted averaging operators. IEEE Transactions on Systems, Man, and Cybernetics 29, 141–150 (1999)

Measuring the Amount of Knowledge for Atanassov's Intuitionistic Fuzzy Sets

Eulalia Szmidt, Janusz Kacprzyk, and Paweł Bujnowski

[1] Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01–447 Warsaw, Poland
[2] Warsaw School of Information Technology, ul. Newelska 6, 01-447 Warsaw, Poland
{szmidt,kacprzyk,bujnowski}@ibspan.waw.pl

Abstract. We address the problem of how to measure amount of knowledge conveyed by an Atanassov's intuitionistic fuzzy set (A-IFS for short). The problem is useful from the point of view of a specific purpose, notably related to decision making. An amount of knowledge is strongly linked to its related amount of information. We pay particular attention to the relationship between the positive and negative information and a lack of information expressed by the hesitation margin.

1 Introduction

In case of data represented by a fuzzy set information on them is expressed by a membership function. Knowledge is basically related to information considered in a particular context under consideration. The transformation of information into knowledge is critical from the practical point of view (cf. Stewart [9]) - a notable example may here be the omnipresent problem of decision making.

Here we consider information conveyed by a piece of data represented by an A-IFS and its related knowledge that is placed in a context considered. Information that is conveyed by an A-IFS, may be considered just as some generalization of information conveyed by a fuzzy set, and consists from the two terms present in the definition of an A-IFS, i.e., the membership and non-membership functions ("responsible" for the positive and negative information, respectively). But for practical purposes it seems necessary to also take into account a so called hesitation margin (cf. Szmidt and Kacprzyk [13], [14], [17], [15], [20]), [21], Bustince et al. [4], [5], Szmidt and Kukier [19], [25], [26], etc.).

Entropy is often viewed as a dual measure of the amount of knowledge. In this paper we show that the entropy alone (cf. Szmidt and Kacprzyk [15], [20]) may not be a satisfactory dual measure of knowledge useful from the point of view of decision making in the A-IFS context. The reason is that an entropy measure answers the question about the fuzziness but does not consider reasons for the fuzziness. So, the two situations, one with the maximal entropy for a membership function equal to a non-membership function (e.g. both equal to 0.5), and another when we know absolutely nothing (i.e. both equal to 0), are equal from the point of view of the entropy measure (in terms of the A-IFSs). However, from the point of view of decision making the two situations are

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 17–24, 2011.

clearly different. This is the motivation of this paper as we propose here a new measure of knowledge for the A-IFSs which is meant to complement the entropy measure to be able to capture additional features which may be relevant when making decisions. The new measure of the amount of knowledge is tested on a simple example taken from the source Quinlan's paper but solved using different tools than therein. This example, which is simple judging by appearances, is a challenge to many classification and machine learning methods, and its solution which we have proposed can be an inspiration to the solution of many real world problems.

2 Brief Introduction to the A-IFSs

One of the possible generalization of a fuzzy set in X (Zadeh [27]), given by

$$A^{'} = \{< x, \mu_{A'}(x) > | x \in X\} \tag{1}$$

where $\mu_{A'}(x) \in [0,1]$ is the membership function of the fuzzy set $A^{'}$, is an A-IFS (Atanassov [1], [2]) A is given by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\} \tag{2}$$

where: $\mu_A : X \rightarrow [0,1]$ and $\nu_A : X \rightarrow [0,1]$ such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{3}$$

and $\mu_A(x)$, $\nu_A(x) \in [0,1]$ denote a degree of membership and a degree of non-membership of $x \in A$, respectively. (Two approaches to the assigning memberships and non-memberships for A-IFSs are proposed by Szmidt and Baldwin [12]).

An additional concept for each A-IFS in X, that is not only an obvious result of (2) and (3) but which is also relevant for applications, we will call (Atanassov [2])

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{4}$$

a *hesitation margin* of $x \in A$ which expresses a lack of information of whether x belongs to A or not (cf. Atanassov [2]). It is obvious that $0 \leq \pi_A(x) \leq 1$, for each $x \in X$.

The hesitation margin turns out to be important while considering the distances (Szmidt and Kacprzyk [13], [14], [17], entropy (Szmidt and Kacprzyk [15], [16], [20]), similarity (Szmidt and Kacprzyk [21]) for the A-IFSs, ranking (Szmidt and Kacprzyk [22], [23]) etc. i.e., the measures that play a crucial role in virtually all information processing tasks. Hesitation margins turn out to be relevant for applications - in image processing (cf. Bustince et al. [4], [5]) and classification of imbalanced and overlapping classes (cf. Szmidt and Kukier [19], [25], [26]), group decision making, negotiations, voting and other situations (cf. Szmidt and Kacprzyk papers).

As we will use three term representation of A-IFSs, each element x will be described via a triplet: (μ, ν, π), i.e., by the membership μ, non-membership ν, and hesitation margin π.

In our further considerations we will use the notion of distances. In Szmidt and Kacprzyk [14], [17], Szmidt and Baldwin [10], [11], it is shown why in the calculation of distances between A-IFSs one should use all three terms describing A-IFSs. In

this paper we use the normalized Hamming distance between any two A-IFSs A and B in $X = \{x_1, x_2, \ldots, x_n\}$, namely:

$$l_{IFS}(A,B) = \frac{1}{2n}\sum_{i=1}^{n}(|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \tag{5}$$

The distance is from the interval [0,1] and fulfills all the conditions of a metric.

Also the notation of a complement set A^C will be used

$$A^C = \{\langle x, \nu_A(x), \mu_A(x), \pi_A(x)\rangle | x \in X\} \tag{6}$$

2.1 Entropy of the A-IFSs

We will verify here if entropy of the A-IFSs may be a reliable measure of the amount of knowledge from the point of view of decision making. Entropy examined here is a non-probabilistic-type entropy measure for the A-IFSs in the sense of De Luca and Termini's [6] axioms which are intuitive and have been widely employed in the fuzzy literature. The axioms were properly reformulated for the A-IFSs (see Szmidt and Kacprzyk [15]) and are discussed in length by Szmidt and Kacprzyk in [15], [16] and [20]. Here we remind only the basic idea.

Entropy, as is considered here, answers the question: how fuzzy is a fuzzy set? In other words, entropy $E(x)$ measures the missing information which may be necessary to say if an element x described by (μ, ν, π) fully belongs or fully does not belong to our set.

Definition 1. *A ratio-based measure of fuzziness i.e., entropy of an (intuitionistic fuzzy) element x is given in the following way [15]:*

$$E(x) = \frac{a}{b} \tag{7}$$

where: a is a distance(x, x_{near}) *from x to the nearer element x_{near} among the elements: $M(1,0,0)$ and $N(0,1,0)$, and b is the* distance(x, x_{far}) *from x to the farer element x_{far} among the elements: $M(1,0,0)$ and $N(0,1,0)$.*

Different ways of expressing entropy $E(x)$ (7) are presented in Szmidt and Kacprzyk in [15], [16] and [20]. Formula (7) describes the degree of fuzziness for a single element belonging to an A-IFS. For n elements belonging to an A-IFS we have

$$E = \frac{1}{n}\sum_{i=1}^{n} E(x_i) \tag{8}$$

A typical shape of a properly defined entropy measure is given in Figure 1. The same shape as in Figure 1 is obtained from (7) for each fixed value of a hesitation margin. In Figure 2 we have the values of entropy $E(x)$ (7) and its contour plot as a function of membership values and non-membership values. It is worth noticing (Figure 2) that for all the values of the hesitation margin, entropy reaches its maximum (equal to 1) for $\mu = \nu$. It is a proper feature of an entropy measure but a question arises if such an entropy measure conveys all knowledge important from the point of view of decision making.

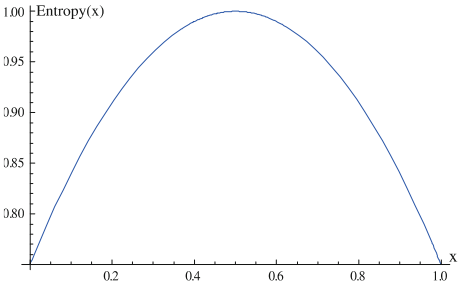


Fig. 1. A typical shape of a properly defined entropy measure

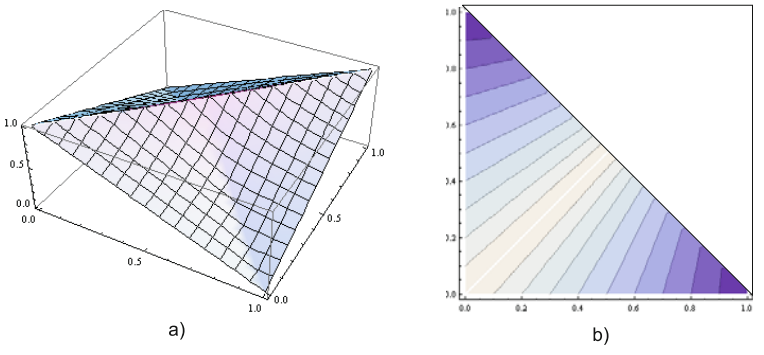
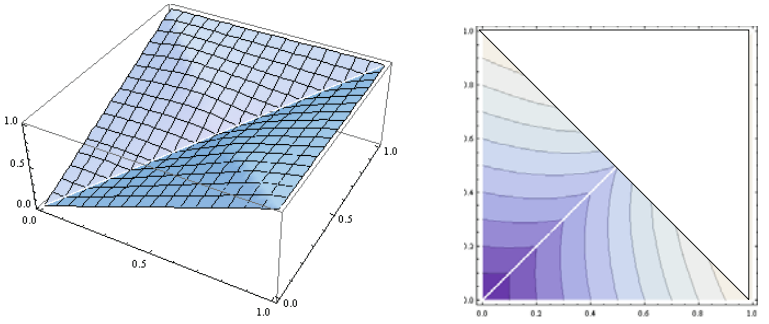


Fig. 2. a) – Entropy $E(x)$ (7); b) – its contour plot

3 Desirable Features of a Measure of Information, and Measure of Knowledge for the A-IFSs

Information concerning a separate element x belonging to an A-IFS is equal to $\mu(x) + \nu(x)$, or [cf. (4)]: $1 - \pi(x)$. But it is one aspect of information only. For each fixed π there are different possibilities of combination between μ and ν. The combination between them influences strongly the amount of knowledge from the point of view of decision making. The knowledge (for a fixed π) is different for the distant values between μ and ν, and for the close values between μ and ν. For example, if $\pi = 0.1$, the knowledge from a point of view of decision making for $\mu = 0.85$ and $\nu = 0.05$ is bigger than for the case: $\mu = 0.45$ and $\nu = 0.45$ (although in both cases $\mu + \nu = 0.9$). Entropy proposed by Szmidt and Kacprzyk ([15], [16]) takes this aspect into account, and is a good measure answering the question how fuzzy is an A-IFS (when considering entropy one is not interested in the reasons of fuzziness). But when decision making, one is also interested in making differences between the following situations (Szmidt and Kreinovich [18]):

– we have no information at all, and
– we have a large number of arguments in favor but an equally large number of arguments in favor of the opposite statement.



a) b)

Fig. 3. a) - measure $K(x)$; b) - its contourplot

In other words, we would like to have a measure making a difference between (0.5, 0.5, 0), and (0, 0, 1).

To distinguish between these two types of situations, we should take into account, beside entropy measure, also the hesitation margin π.

It seems that a good measure of the amount of knowledge (useful from the point of view of decision making) connected to a separate element $x \in A - IFS$ is:

$$K(x) = 1 - 0.5(E(x) + \pi(x)) \tag{9}$$

where $E(x)$ is an entropy measure given by (7) (Szmidt and Kacprzyk [15]), $\pi(x)$ is the hesitation margin.

Measure $K(x)$ (9) makes it possible to meaningfully represent what, in our context, is meant by the amount of knowledge, is simple – both conceptually and numerically, which will certainly be a big asset while solving complex real world problems which will be the subject of our next papers.

The properties of (9) are:

1. $0 \leq K(x) \leq 1$
2. $K(x) = K(x^C)$
3. For a fixed $E(x)$, $K(x)$ increases while π decreases,
4. For a fixed value of π, $K(x)$ behaves dually to an entropy measure (i.e., as $1 - E(x)$).

In Figure 3 we can see the shape of $K(x)$, and its contour plot. We may easily notice the desirable differences of the shapes of $K(x)$ and entropy $E(x)$ (Figure 2) from the point of view of decision making.

In a case of n elements, the total amount of knowledge K is:

$$K = \frac{1}{n}\sum_{i=1}^{n}(1 - 0.5(E(x_i) + \pi(x_i))) \tag{10}$$

Now we will evaluate the proposed measure of knowledge K on the problem formulated by Quinlan [8] – verifying if we obtain similar results.

Table 1. Evaluation of the attributes of the "Saturday Morning" data

Attribute	K	Entropy	$\overline{\pi}$
1	2	3	4
Outlook	0.49	0.56	0.453
Temperature	0.23	0.813	0.72
Humidity	0.47	0.535	0.535
Windy	0.26	0.744	0.735

The Quinlan's example, the so-called "Saturday Morning" example, considers classification with nominal data. The example is small enough and illustrative, yet is a challenge to many classification and machine learning methods. The main idea of solving the example by Quinlan was to select the best attribute to split the training set (Quinlan used a so called *Information Gain* which was a dual measure to Shannon's entropy). Quinlan obtained 100% accuracy, and the optimal solution (the minimal possible tree). In other words, the ranking of the attributes as pointed out by Quinlan [8] is the best as far as the amount of knowledge is concerned.

The limitation of space does not let us discuss the Quinlan example [8] and its A-IFS counterpart (see Szmidt and Kacprzyk [24]) in details. We present here only the final results, i.e. effects of evaluation of the attributes by the proposed measure of knowledge K (9) from the point of view of conveying the most knowledge. In Table 1 there are results for each attribute accessed from the point of view of the measure of the amount of knowledge K (9), entropy E, (7)–(8), and an average hesitation margin $\overline{\pi}$.

Let us remind that in the original solution given by Quinlan [8] (leading to the minimal tree), the order from the point of view of the most informative attributes was the following:

Outlook, Humidity, Windy, Temperature

If we order the attributes taking into account entropy (7)–(8), the most informative attributes are indicated by the smallest values in the 3rd column of Table 1:

Humidity, Outlook, Windy, Temperature

i.e., the order of the attributes is different (Humidity replaced Outlook – this order would not result in the smallest tree).

On the other hand, if we order the attributes taking into account the minimal average values of the hesitation margin only, the most informative attributes are (Table 1, 4th column):

Outlook, Humidity, Temperature, Windy

i.e., again, the order of the attributes is changed (Temperature replaced Windy).

But when we apply the knowledge measure (9), the results are the same as Quinlan's ones, (leading to the minimal tree, i.e., the most valuable from the point of view of decision making – 2nd column in Table 1), i.e.:

Outlook, Humidity, Windy, Temperature

.

4 Conclusions

A new measure of the amount of knowledge for the A-IFSs was proposed to be useful from the point of view of decision making. The new measure maintains advantages of the entropy measure (reflecting the relationship between the positive and negative information) and additionally emphasizes the influence of the lacking information (expressed by the hesitation margins).

Acknowledgment. Partially supported by the Ministry of Science and Higher Education Grant Nr N N519 384936.

References

1. Atanassov, K.: Intuitionistic Fuzzy Sets. VII ITKR Session. Sofia (Deposed in Centr. Sci.-Techn. Library of Bulg. Acad. of Sci., 1697/84) (1983) (in Bulgarian)
2. Atanassov, K.: Intuitionistic Fuzzy Sets: Theory and Applications. Springer, Heidelberg (1999)
3. Atanassov, K., Tasseva, V., Szmidt, E., Kacprzyk, J.: On the geometrical interpretations of the intuitionistic fuzzy sets. In: Atanassov, K., Kacprzyk, J., Krawczak, M., Szmidt, E. (eds.) Issues in the Representation and Processing of Uncertain and Imprecise Information. Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets, and Related Topics. EXIT, Warsaw (2005)
4. Bustince, H., Mohedano, V., Barrenechea, E., Pagola, M.: An algorithm for calculating the threshold of an image representing uncertainty through A-IFSs. In: IPMU 2006, pp. 2383–2390 (2006)
5. Bustince, H., Mohedano, V., Barrenechea, E., Pagola, M.: Image thresholding using intuitionistic fuzzy sets. In: Atanassov, K., Kacprzyk, J., Krawczak, M., Szmidt, E. (eds.) Issues in the Representation and Processing of Uncertain and Imprecise Information. Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets, and Related Topics. EXIT, Warsaw (2005)
6. De Luca, A., Termini, S.: A definition of a non-probabilistic entropy in the setting of fuzzy sets theory. Inform. And Control 20, 301–312 (1972)
7. Narukawa, Y., Torra, V.: Non-monotonic fuzzy measure and intuitionistic fuzzy set. In: Torra, V., Narukawa, Y., Valls, A., Domingo-Ferrer, J. (eds.) MDAI 2006. LNCS (LNAI), vol. 3885, pp. 150–160. Springer, Heidelberg (2006)
8. Quinlan, J.R.: Induction of decision trees. Machine Learning 1, 81–106 (1986)
9. Stewart, T.: Wealth of Knowledge, Doubleday, New York (2001)
10. Szmidt, E., Baldwin, J.: New similarity measure for intuitionistic fuzzy set theory and mass assignment theory. Notes on IFSs 9(3), 60–76 (2003)
11. Szmidt, E., Baldwin, J.: Entropy for intuitionistic fuzzy set theory and mass assignment theory. Notes on IFSs 10(3), 15–28 (2004)
12. Szmidt, E., Baldwin, J.: Intuitionistic Fuzzy Set Functions, Mass Assignment Theory, Possibility Theory and Histograms. IEEE World Congress on Computational Intelligence, 237–243 (2006)
13. Szmidt, E., Kacprzyk, J.: On measuring distances between intuitionistic fuzzy sets. Notes on IFS 3(4), 1–13 (1997)
14. Szmidt, E., Kacprzyk, J.: Distances between intuitionistic fuzzy sets. Fuzzy Sets and Systems 114(3), 505–518 (2000)
15. Szmidt, E., Kacprzyk, J.: Entropy for intuitionistic fuzzy sets. Fuzzy Sets and Systems 118(3), 467–477 (2001)

16. Szmidt, E., Kacprzyk, J.: Similarity of intuitionistic fuzzy sets and the Jaccard coefficient. In: IPMU 2004, pp. 1405–1412 (2004)
17. Szmidt, E., Kacprzyk, J.: Distances between intuitionistic fuzzy sets: straightforward approaches may not work. In: 3rd International IEEE Conference Intelligent Systems IS 2006, London, pp. 716–721 (2006)
18. Szmidt, E., Kreinovich, V.: Symmetry between true, false, and uncertain: An explanation. Notes on IFS 15(4), 1–8 (2009)
19. Szmidt, E., Kukier, M.: Classification of imbalanced and overlapping classes using intuitionistic fuzzy sets. In: 3rd International IEEE Conference Intelligent Systems IS 2006, London, pp. 722–727 (2006)
20. Szmidt, E., Kacprzyk, J.: Some problems with entropy measures for the Atanassov intuitionistic fuzzy sets. In: Masulli, F., Mitra, S., Pasi, G. (eds.) WILF 2007. LNCS (LNAI), vol. 4578, pp. 291–297. Springer, Heidelberg (2007)
21. Szmidt, E., Kacprzyk, J.: A New Similarity Measure for Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. In: IEEE Conf. on Fuzzy Systems, pp. 481–486 (2007a)
22. Szmidt, E., Kacprzyk, J.: A new approach to ranking alternatives expressed via intuitionistic fuzzy sets. In: Ruan, D., et al. (eds.) Computational Intelligence in Decision and Control, pp. 265–270. World Scientific, Singapore (2008)
23. Szmidt, E., Kacprzyk, J.: Amount of information and its reliability in the ranking of Atanassov's intuitionistic fuzzy alternatives. In: Rakus-Andersson, E., Yager, R., Ichalkaranje, N., Jain, L.C. (eds.) Recent Advances in Decision Making. SCI, vol. 222, pp. 7–19. Springer, Heidelberg (2009)
24. Szmidt, E., Kacprzyk, J.: Dealing with typical values via Atanassov's intuitionistic fuzzy sets. Int. J. of General Systems 39(5), 489–506 (2010)
25. Szmidt, E., Kukier, M.: A new approach to classification of imbalanced classes via Atanassov's intuitionistic fuzzy sets. In: Wang, H.-F. (ed.) Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery, pp. 65–102. Idea Group, USA (2008)
26. Szmidt, E., Kukier, M.: Atanassov's intuitionistic fuzzy sets in classification of imbalanced and overlapping classes. In: Chountas, P., Petrounias, I., Kacprzyk, J. (eds.) Intelligent Techniques and Tools for Novel System Architectures, pp. 455–471. Springer, Heidelberg (2008)
27. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
28. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Inform. Sciences 8, 199–249 (1975)

Distributivity of Implication Operations over t-Representable T-Norms Generated from Nilpotent T-Norms

Michał Baczyński

Institute of Mathematics, University of Silesia,
40-007 Katowice, ul. Bankowa 14, Poland
michal.baczynski@us.edu.pl

Abstract. Recently, in [3], we have discussed the distributive equation of implications $\mathcal{I}(x,\mathcal{T}_1(y,z)) = \mathcal{T}_2(\mathcal{I}(x,y),\mathcal{I}(x,z))$ over t-representable t-norms generated from strict t-norms in interval-valued fuzzy sets theory. In this work we continue these investigations, but for t-representable t-norms generated from nilpotent t-norms. As a byproduct result we show all solutions of some functional equation related to this case.

Keywords: Interval-valued fuzzy sets, Intuitionistic fuzzy sets, Fuzzy implication, Triangular norm, Distributivity, Functional equations.

1 Introduction

Distributivity of fuzzy implications over different fuzzy logic connectives has been studied in the recent past by many authors (see [1], [17], [5], [15], [16], [4], [2]). These equations have a very important role to play in efficient inferencing in approximate reasoning, especially fuzzy control systems (see [7]).

Recently, in [3], we have discussed the distributive equation of implications $\mathcal{I}(x,\mathcal{T}_1(y,z)) = \mathcal{T}_2(\mathcal{I}(x,y),\mathcal{I}(x,z))$, over t-representable t-norms generated from strict t-norms in interval-valued fuzzy sets theory. In this work we continue these investigations, but for t-norms generated from nilpotent t-norms. In [3], as a byproduct, we have obtained the solutions of the following functional equation:

$$f(u_1+v_1,u_2+v_2) = f(u_1,u_2)+f(v_1,v_2), \qquad (u_1,u_2),(v_1,v_2)\in L^{\infty}, \quad \text{(B1)}$$

where $L^{\infty} = \{(u_1,u_2)\in[0,\infty]^2 \mid u_1 \geq u_2\}$ and $f\colon L^{\infty}\to[0,\infty]$ is an unknown function. In this article we will present all solutions of the following equation:

$$f(\min(u_1+v_1,a),\min(u_2+v_2,a)) = \min(f(u_1,u_2)+f(v_1,v_2),b), \quad \text{(B2)}$$

where $a,b>0$ are fixed real numbers, $f\colon L^a\to[0,b]$ is an unknown function and $L^a = \{(u_1,u_2)\in[0,a]^2 \mid u_1\geq u_2\}$. This equation is related to the case with nilpotent t-norms. Such theoretical developments connected with solutions of different functional equations can be also useful in other topics like fuzzy mathematical morphology (see [8]) or similarity measures (cf. [6]).

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 25–32, 2011.

We assume that the reader is familiar with the notion of intuitionistic (by Atanassov) fuzzy sets theory and interval-valued fuzzy sets theory (in [9] it is shown that both theories are equivalent from the mathematical point of view). Since we are limited in number of pages, in this article we discuss main results in the language of interval-valued fuzzy sets, but they can be easily transformed to the intuitionistic fuzzy case. Let us define

$$L^I = \{(x_1, x_2) \in [0,1]^2 \ : \ x_1 \leq x_2\},$$
$$(x_1, x_2) \leq_{L^I} (y_1, y_2) \Longleftrightarrow x_1 \leq y_1 \wedge x_2 \leq y_2.$$

In the sequel, if $x \in L^I$, then we denote it by $x = [x_1, x_2]$. One can easily observe that $\mathcal{L}^I = (L^I, \leq_{L^I})$ is a complete lattice with units $0_{\mathcal{L}^I} = [0,0]$ and $1_{\mathcal{L}^I} = [1,1]$. An interval-valued fuzzy set on X is a mapping $A\colon X \to L^I$.

2 Basic Fuzzy Connectives

We assume that the reader is familiar with the classical results concerning basic fuzzy logic connectives, but we briefly mention some of the results employed in the rest of the work.

Definition 2.1. *Let $\mathcal{L} = (L, \leq_L)$ be a complete lattice. An associative, commutative operation $\mathcal{T}\colon L^2 \to L$ is called a t-norm if it is increasing and $1_{\mathcal{L}}$ is the neutral element of $\mathcal{T}$.*

Definition 2.2. *A t-norm T on $([0,1], \leq)$ is said to be* nilpotent*, if it is continuous and if each $x \in (0,1)$ is a nilpotent element of T, i.e., if there exists $n \in \mathbb{N}$ such that $x_T^{[n]} = 0$, where* $x_T^{[n]} := \begin{cases} x, & \text{if } n = 1, \\ T(x, x_T^{[n-1]}), & \text{if } n > 1. \end{cases}$

The following characterization of nilpotent t-norms is well-known.

Theorem 2.3 ([13]). *A function $T\colon [0,1]^2 \to [0,1]$ is a nilpotent t-norm if and only if there exists a continuous, strictly decreasing function $t\colon [0,1] \to [0,\infty]$ with $t(1) = 0$ and $t(0) < \infty$, which is uniquely determined up to a positive multiplicative constant, such that $T(x,y) = t^{-1}(\min(t(x) + t(y), t(0)))$, for all $x, y \in [0,1]$.*

In our article we shall consider the following special class of t-norms.

Definition 2.4 (see [10]). *A t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is called t-representable if there exist t-norms T_1 and T_2 on $([0,1], \leq)$ such that $T_1 \leq T_2$ and*

$$\mathcal{T}([x_1, x_2], [y_1, y_2]) = [T_1(x_1, y_1), T_2(x_2, y_2)], \qquad [x_1, x_2], [y_1, y_2] \in L^I.$$

It should be noted that not all t-norms on $\mathcal{L}^I$ are t-representable (see [10]).

One possible definition of an implication on $\mathcal{L}^I$ is based on the well-accepted notation introduced by Fodor and Roubens [12] (see also [11] and [14]).

Definition 2.5. *Let $\mathcal{L} = (L, \leq_L)$ be a complete lattice. A function $\mathcal{I}\colon L^2 \to L$ is called a fuzzy implication on $\mathcal{L}$ if it is decreasing with respect to the first variable, increasing with respect to the second variable and fulfills the following conditions: $\mathcal{I}(0_{\mathcal{L}}, 0_{\mathcal{L}}) = \mathcal{I}(1_{\mathcal{L}}, 1_{\mathcal{L}}) = \mathcal{I}(0_{\mathcal{L}}, 1_{\mathcal{L}}) = 1_{\mathcal{L}}$ and $\mathcal{I}(1_{\mathcal{L}}, 0_{\mathcal{L}}) = 0_{\mathcal{L}}$.*

3 Some New Results Pertaining to Functional Equations

In this section we show one new result related to functional equations, which will be crucial in obtaining main results.

Proposition 3.1 ([4, Proposition 3]). *Fix real $a, b > 0$. For a function $f\colon [0,a] \to [0,b]$ the following statements are equivalent:*

(i) f satisfies the functional equation

$$f(\min(x+y,a)) = \min(f(x)+f(y),b), \qquad x,y \in [0,a].$$

(ii) Either $f = 0$, or $f = b$, or $f(x) = \begin{cases} 0, & \text{if } x = 0, \\ b, & \text{if } x > 0, \end{cases}$ or there exists a unique constant $c \in \left[\frac{b}{a}, \infty\right)$ such that $f(x) = \min(cx, b)$, for all $x \in [0,a]$.

Proposition 3.2. *Fix real $a, b > 0$. Let $L^a = \{(u_1,u_2) \in [0,a]^2 \ : \ u_1 \geq u_2\}$. For a function $f\colon L^a \to [0,b]$ the following statements are equivalent:*

(i) f satisfies the functional equation (B2) for all $(u_1,u_2),(v_1,v_2) \in L^a$.

(ii) Either

$$f = 0, \tag{S1}$$

or

$$f = b, \tag{S2}$$

or

$$f(u_1,u_2) = \begin{cases} 0, & \text{if } u_2 = 0, \\ b, & \text{if } u_2 > 0, \end{cases} \tag{S3}$$

or

$$f(u_1,u_2) = \begin{cases} 0, & \text{if } u_1 = 0, \\ b, & \text{if } u_1 > 0, \end{cases} \tag{S4}$$

or there exists unique $c \in \left[\frac{b}{a}, \infty\right)$ such that

$$f(u_1,u_2) = \min(cu_2, b), \tag{S5}$$

or

$$f(u_1,u_2) = \begin{cases} \min(cu_1,b), & \text{if } u_1 = u_2, \\ b, & \text{if } u_1 > u_2, \end{cases} \tag{S6}$$

or

$$f(u_1,u_2) = \begin{cases} \min(cu_1,b), & \text{if } u_2 = 0, \\ b, & \text{if } u_2 > 0, \end{cases} \tag{S7}$$

or

$$f(u_1,u_2) = \min(cu_1, b), \tag{S8}$$

or there exist unique $c_1, c_2 \in \left[\frac{b}{a}, \infty\right)$, $c_1 \neq c_2$ such that

$$f(u_1,u_2) = \begin{cases} \min(c_1(u_1-u_2)+c_2u_2, b), & \text{if } u_2 < a, \\ b, & \text{if } u_2 = a, \end{cases} \tag{S9}$$

for all $(u_1,u_2) \in L^a$.

Proof. $(ii) \Longrightarrow (i)$ It is a direct calculation that the above functions satisfy (B2).

$(i) \Longrightarrow (ii)$ Let a function $f\colon L^a \to [0,b]$ satisfy equation (B2) for all (u_1,u_2), $(v_1,v_2) \in L^a$. Setting $u_1 = v_1 = a$ in (B2) we get

$$f(\min(a+a,a),\min(u_2+v_2,a)) = \min(f(a,u_2)+f(a,v_2),b), \qquad u_2,v_2 \in [0,a].$$

Let us denote $f_a(x) := f(a,x)$, for $x \in [0,a]$. Therefore, we get

$$f_a(\min(u_2+v_2,a)) = \min(f_a(u_2)+f_a(v_2),b), \qquad u_2,v_2 \in [0,a].$$

For this equation we can use solutions from Proposition 3.1. We have 4 possible cases for the function f_a.

1. If $f_a = 0$, then putting $u_1 = u_2 = a$ in (B2) we have

   $$f(\min(a+v_1,a),\min(a+v_2,a)) = \min(f(a,a)+f(v_1,v_2),b), \quad (v_1,v_2) \in L^a,$$

   so $f(a,a) = \min(f(a,a)+f(v_1,v_2),b)$, thus $0 = \min(f(v_1,v_2),b)$, hence $f(v_1,v_2) = 0$ for all $(v_1,v_2) \in L^a$ and we get first solution $f = 0$, i.e., (S1).
2. If $f_a(x) = \begin{cases} 0, & \text{if } x = 0 \\ b, & \text{if } x > 0 \end{cases}$, then putting $u_1 = a$ in (B2) we have

   $$f(a,\min(u_2+v_2,a)) = \min(f(a,u_2)+f(v_1,v_2),b), \qquad (v_1,v_2) \in L^a.$$

   If we take $u_2 = v_2 = 0$ above, then we get

   $$f(a,0) = \min(f(a,0)+f(v_1,0),b), \qquad v_1 \in [0,a],$$

   thus $0 = \min(f(v_1,0),b)$, i.e., $f(v_1,0) = 0$ for all $v_1 \in [0,a]$. If we take $u_2 = 0$ and $v_2 > 0$ above, then we get $f(a,v_2) = \min(f(v_1,v_2),b)$, for $v_1 \in [0,a]$, therefore $b = \min(f(v_1,v_2),b)$, i.e., $f(v_1,v_2) = b$. In summary, we get the solution (S3).
3. If $f_a(x) = \min(cx,b)$ with some real $c \in \left[\frac{b}{a},\infty\right)$, then putting $u_1 = a$ and $u_2 = 0$ in (B2) we have

   $$f_a(v_2) = \min(f_a(0)+f(v_1,v_2),b), \qquad (v_1,v_2) \in L^a,$$

   thus $f_a(v_2) = \min(f(v_1,v_2),b)$, so $\min(cv_2,b) = f(v_1,v_2)$ and we get the solution (S5) in this case.

Therefore, we need to solve our equation with the last possible assumption that $f_a = b$. Setting now $u_2 = v_2 = 0$ in (B2) we get

$$f(\min(u_1+v_1,a),0) = \min(f(u_1,0)+f(v_1,0),b), \qquad u_1,v_1 \in [0,a].$$

Let us denote $f^0(x) := f(x,0)$, for $x \in [0,a]$. Hence, we obtain

$$f^0(\min(u_1+v_1,a)) = \min(f^0(u_1)+f^0(v_1),b), \qquad u_1,v_1 \in [0,a].$$

For this equation we again can use solutions described in Proposition 3.1. We have 4 possible cases for the function f^0.

1. If $f^0 = 0$, then $f(a,0) = 0$, which contradicts our assumption $f_a = b$.
2. If $f^0 = b$, then putting $u_1 = u_2 = 0$ in (B2) we have

$$f(v_1, v_2) = \min(b + f(v_1, v_2), b), \qquad (v_1, v_2) \in L^a,$$

thus $f(v_1, v_2) = b$, hence we get next possible solution $f = b$, i.e., (S2).

3. If $f^0(x) = \begin{cases} 0, & \text{if } x = 0 \\ b, & \text{if } x > 0 \end{cases}$, then putting $u_2 = 0$ in (B2) we have

$$f(\min(u_1 + v_1, a), v_2) = \min(f(u_1, 0) + f(v_1, v_2), b).$$

Let us assume that $u_1 > 0$ and $v_1 = v_2$ above. Then we get

$$f(\min(u_1 + v_2, a), v_2) = \min(b + f(v_2, v_2), b), \quad u_1 \in (0, a], v_2 \in [0, a].$$

hence

$$f(\min(u_1 + v_2, a), v_2) = b, \qquad u_1 \in (0, a], v_2 \in [0, a].$$

Since $\min(u_1 + v_2, a) \in (v_2, a]$ and $v_2 \in [0, a]$, we have obtained the result that $f(x_1, x_2) = b$ for any $(x_1, x_2) \in L^a$ such that $x_1 > x_2$.
Let us take now $u_2 = u_1$ and $v_2 = v_1$ in (B2). Then we have

$$f(\min(u_1 + v_1, a), \min(u_1 + v_1, a)) = \min(f(u_1, u_1) + f(v_1, v_1), b),$$

for $u_1, v_1 \in [0, a]$. Let us denote $g(x) := f(x, x)$, for $x \in [0, a]$. Therefore,

$$g(\min(u_1 + v_1, a)) = \min(g(u_1) + g(v_1), b), \qquad u_1, v_1 \in [0, a].$$

For this equation we again can use solutions described in Proposition 3.1. We have 4 possible cases for the function g.

(a) If $g = 0$, then $f(a, a) = 0$, which contradicts our assumption $f_a = b$.
(b) If $g = b$, then $f(0, 0) = b$, which contradicts our assumption 3. on function f^0.
(c) If $g(x) = \begin{cases} 0, & \text{if } x = 0 \\ b, & \text{if } x > 0 \end{cases}$, then we get the solution (S4) in this case.
(d) If $g(x) = \min(cx, b)$ with some $c \in \left[\frac{b}{a}, \infty\right)$, then we get the solution (S6) in this case.

4. Let $f^0(x) = \min(c_1 x, b)$, with some real $c_1 \in \left[\frac{b}{a}, \infty\right)$ for $x \in [0, a]$. Putting $u_2 = u_1$ and $v_2 = v_1$ in (B2) we have

$$f(\min(u_1 + v_1, a), \min(u_1 + v_1, a)) = \min(f(u_1, u_1) + f(v_1, v_1), b),$$

for $u_1, v_1 \in [0, a]$. Let us denote $g(x) := f(x, x)$, for $x \in [0, a]$. Then we get

$$g(\min(u_1 + v_1, a)) = \min(g(u_1) + g(v_1), b), \qquad u_1, v_1 \in [0, a].$$

For this equation we again can use solutions described in Proposition 3.1. We have 4 possible cases for the function g.

(a) If $g = 0$, then $f(a, a) = 0$, which contradicts our assumption $f_a = b$.
(b) If $g = b$, then $f(0, 0) = b$, which contradicts assumption on function f^0.
(c) If $g(x) = \begin{cases} 0, & \text{if } x = 0 \\ b, & \text{if } x > 0 \end{cases}$, then putting $u_2 = 0$ and $v_1 = v_2$ in (B2) we get

$$f(\min(u_1 + v_2, a), v_2) = \min(f(u_1, 0) + f(v_2, v_2), b), \qquad u_1, v_2 \in [0, a],$$

thus

$$f(\min(u_1 + v_2, a), v_2) = \min\left(\min(c_1 u_1, b) + \begin{cases} 0, & \text{if } v_2 = 0, \\ b, & \text{if } v_2 > 0, \end{cases}, b\right) = \begin{cases} \min(c_1 u_1, b), & \text{if } v_2 = 0 \\ b, & \text{if } v_2 > 0 \end{cases},$$

for any $u_1, v_2 \in [0, a]$. Since $\min(u_1 + v_2, a) \in [v_2, a]$, this solution can be written as (S7).
(d) If $g(x) = \min(c_2 x, b)$, with some real $c_2 \in \left[\frac{b}{a}, \infty\right)$ for $x \in [0, a]$, then similarly as earlier, putting $u_2 = 0$ and $v_1 = v_2$ in (B2), we get

$$f(\min(u_1 + v_2, a), v_2) = \min(f(u_1, 0) + f(v_2, v_2), b), \qquad u_1, v_2 \in [0, a],$$

therefore

$$f(\min(u_1 + v_2, a), v_2) = \min(\min(c_1 u_1, b) + \min(c_2 v_2, b), b),$$

thus

$$f(\min(u_1 + v_2, a), v_2) = \min(c_1 u_1 + c_2 v_2, b).$$

If $c_1 = c_2$, then $f(\min(u_1 + v_2, a), v_2) = \min(c_1(u_1 + v_2), b)$, so this solution can be written as (S8). If $c_1 \neq c_2$, then we get the last solution (S9). □

4 Distributive Equation for t-Representable T-Norms

In this section we will show how we can use solutions presented in Proposition 3.2 to obtain all solutions, in particular fuzzy implications, of our main distributive equation

$$\mathcal{I}(x, \mathcal{T}_1(y, z)) = \mathcal{T}_2(\mathcal{I}(x, y), \mathcal{I}(x, z)), \qquad x, y, z \in L^I, \tag{D1}$$

where $\mathcal{I}$ is an unknown function and t-norms $\mathcal{T}_1$ and $\mathcal{T}_2$ on $\mathcal{L}^I$ are t-representable and generated from nilpotent t-norms T_1, T_2 and T_3, T_4, respectively.

Assume that projection mappings on $\mathcal{L}^I$ are defined as the following:

$$pr_1([x_1, x_2]) = x_1, \qquad pr_2([x_1, x_2]) = x_2, \qquad \text{for } [x_1, x_2] \in L^I.$$

In [3] we have shown that if $\mathcal{T}_1$ and $\mathcal{T}_2$ on $\mathcal{L}^I$ are t-representable, then

$$g^1_{[x_1,x_2]}([T_1(y_1, z_1), T_2(y_2, z_2)]) = T_3(g^1_{[x_1,x_2]}([y_1, y_2]), g^1_{[x_1,x_2]}([z_1, z_2])),$$
$$g^2_{[x_1,x_2]}([T_1(y_1, z_1), T_2(y_2, z_2)]) = T_4(g^2_{[x_1,x_2]}([y_1, y_2]), g^2_{[x_1,x_2]}([z_1, z_2])),$$

where $[x_1, x_2] \in L^I$ is arbitrarily fixed and functions $g^1_{[x_1,x_2]}, g^2_{[x_1,x_2]} \colon L^I \to L^I$ are defined by

$$g^1_{[x_1,x_2]}(\cdot) := pr_1 \circ \mathcal{I}([x_1, x_2], \cdot), \qquad g^2_{[x_1,x_2]}(\cdot) := pr_2 \circ \mathcal{I}([x_1, x_2], \cdot).$$

Let us assume that $T_1 = T_2$ and $T_3 = T_4$ are nilpotent t-norms generated from additive generators t_1 and t_3, respectively. Using the representation theorem of nilpotent t-norms (Theorem 2.3) we can transform our problem to the following equation (for a simplicity we deal only with g^1 now):

$$\begin{aligned} g^1_{[x_1,x_2]}([t_1^{-1}(\min(t_1(y_1) + t_1(z_1), t_1(0))), t_1^{-1}(\min(t_1(y_2) + t_1(z_2), t_1(0)))]) \\ = t_3^{-1}(\min(t_3(g^1_{[x_1,x_2]}([y_1, y_2])) + t_3(g^1_{[x_1,x_2]}([z_1, z_2])), t_3(0))). \end{aligned}$$

Hence

$$\begin{aligned} t_3 \circ g^1_{[x_1,x_2]}([t_1^{-1}(\min(t_1(y_1) + t_1(z_1), t_1(0))), t_1^{-1}(\min(t_1(y_2) + t_1(z_2), t_1(0)))]) \\ = \min(t_3 \circ g^1_{[x_1,x_2]}([y_1, y_2]) + t_3 \circ g^1_{[x_1,x_2]}([z_1, z_2]), t_3(0)). \end{aligned}$$

Let us put $t_1(y_1) = u_1$, $t_1(y_2) = u_2$, $t_1(z_1) = v_1$ and $t_1(z_2) = v_2$. Of course $u_1, u_2, v_1, v_2 \in [0, t_1(0)]$. Moreover $[y_1, y_2], [z_1, z_2] \in L^I$, thus $y_1 \le y_2$ and $z_1 \le z_2$. The generator t_1 is strictly decreasing, so $u_1 \ge u_2$ and $v_1 \ge v_2$. If we put

$$f_{[x_1,x_2]}(a, b) := t_3 \circ pr_1 \circ \mathcal{I}([x_1, x_2], [t_1^{-1}(a), t_1^{-1}(b)]), \qquad a, b \in [0, t_1(0)],\ a \ge b,$$

then we get the following functional equation

$$\begin{aligned} f_{[x_1,x_2]}(\min(u_1 + v_1, t_1(0)), \min(u_2 + v_2, t_1(0))) \\ = \min(f_{[x_1,x_2]}(u_1, u_2) + f_{[x_1,x_2]}(v_1, v_2), t_3(0)), \end{aligned} \tag{1}$$

where $(u_1, u_2), (v_1, v_2) \in L^{t_1(0)}$ and $f_{[x_1,x_2]} \colon L^{t_1(0)} \to [0, t_3(0)]$ is an unknown function. In a same way we can repeat all the above calculations, but for the function g^2, to obtain the following functional equation

$$\begin{aligned} f^{[x_1,x_2]}(\min(u_1 + v_1, t_1(0)), \min(u_2 + v_2, t_1(0))) \\ = \min(f^{[x_1,x_2]}(u_1, u_2) + f^{[x_1,x_2]}(v_1, v_2), t_3(0)), \end{aligned} \tag{2}$$

where

$$f^{[x_1,x_2]}(a, b) := t_3 \circ pr_2 \circ \mathcal{I}([x_1, x_2], [t_1^{-1}(a), t_1^{-1}(b)]).$$

Observe that (1) and (2) are exactly our functional equation (B2). Therefore, using solutions of Proposition 3.2, we are able to obtain the description of the vertical section $\mathcal{I}([x_1, x_2], \cdot)$ for a fixed $[x_1, x_2] \in L^I$. Since in this proposition we have 9 possible solutions, we should have 81 different solutions of (D1). Observe now that some of these solutions are not good, since the range of $\mathcal{I}$ is L^I. Finally, we need to notice that not all obtained vertical solutions in $\mathcal{L}^I$ can be used for obtaining fuzzy implication on $\mathcal{L}^I$ in the sense of Definition 2.5. We will describe solutions which are fuzzy implications in our future works.

Acknowledgment. This work has been supported by the Polish Ministry of Science and Higher Education Grant Nr N N519 384936.

References

1. Baczyński, M.: On a class of distributive fuzzy implications. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 9, 229–238 (2001)
2. Baczyński, M.: On the distributivity of fuzzy implications over continuous and Archimedean triangular conorms. Fuzzy Sets and Systems 161, 1406–1419 (2010)
3. Baczyński, M.: On the distributivity of implication operations over t-representable t-norms generated from strict t-norms in interval-valued fuzzy sets theory. In: Hüllermeier, E., Kruse, R., Hoffmann, F. (eds.) IPMU 2010. CCIS, vol. 80, pp. 637–646. Springer, Heidelberg (2010)
4. Baczyński, M., Jayaram, B.: On the distributivity of fuzzy implications over nilpotent or strict triangular conorms. IEEE Trans. Fuzzy Systems 17, 590–603 (2009)
5. Balasubramaniam, J., Rao, C.J.M.: On the distributivity of implication operators over T- and S-norms. IEEE Trans. Fuzzy Systems 12, 194–198 (2004)
6. Bustince, H., Pagola, M., Barrenechea, E.: Construction of fuzzy indices from fuzzy DI-subsethood measures: Application to the global comparison of images. Inform. Sci. 177, 906–929 (2007)
7. Combs, W.E., Andrews, J.E.: Combinatorial rule explosion eliminated by a fuzzy rule configuration. IEEE Trans. Fuzzy Systems 6, 1–11 (1998)
8. De Baets, B., Kerre, E., Gupta, M.: The fundamentals of fuzzy mathematical morphology, Part 1: Basic Concepts. Int. J. Gen. Syst. 23, 155–171 (1994)
9. Deschrijver, G., Kerre, E.E.: On the relationship between some extensions of fuzzy set theory. Fuzzy Sets and Systems 133, 227–235 (2003)
10. Deschrijver, G., Cornelis, C., Kerre, E.E.: On the representation of intuitionistic fuzzy t-norms and t-conorms. IEEE Trans. Fuzzy Systems 12, 45–61 (2004)
11. Deschrijver, G., Cornelis, C., Kerre, E.E.: Implication in intuitionistic and interval-valued fuzzy set theory: construction, classification and application. Internat. J. Approx. Reason. 35, 55–95 (2004)
12. Fodor, J., Roubens, M.: Fuzzy preference modelling and multicriteria decision support. Kluwer, Dordrecht (1994)
13. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms. Kluwer, Dordrecht (2000)
14. Mas, M., Monserrat, M., Torrens, J., Trillas, E.: A survey on fuzzy implication functions. IEEE Trans. Fuzzy Systems 15, 1107–1121 (2007)
15. Ruiz-Aguilera, D., Torrens, J.: Distributivity of strong implications over conjunctive and disjunctive uninorms. Kybernetika 42, 319–336 (2005)
16. Ruiz-Aguilera, D., Torrens, J.: Distributivity of residual implications over conjunctive and disjunctive uninorms. Fuzzy Sets and Systems 158, 23–37 (2007)
17. Trillas, E., Alsina, C.: On the law $[p \wedge q \to r] = [(p \to r) \vee (q \to r)]$ in fuzzy logic. IEEE Trans. Fuzzy Systems 10, 84–88 (2002)

Cuts of IF-sets Respecting Fuzzy Connectives*

Davide Martinetti[1], Vladimír Janiš[2], and Susana Montes[1]

[1] Department of Statistics and O.R., University of Oviedo, Spain
{martinettidavide.uo,montes}@uniovi.es
[2] Department of Mathematics, Faculty of Natural Sciences,
Matej Bel University, Slovak Republic
vladimir.janis@umb.sk

Abstract. Intuitionistic fuzzy sets (IF-sets) are a suitable tool to describe cases where it is useful to account not only the grade of membership to a collection, but also the grade of its non-membership. We consider the α-cuts of an IF-set A as crisp sets consisting of those elements x for which the truth value (in fuzzy logic) of the statement "x belongs to A and it is not true that x does not belong to A" is at least α. We describe properties of such cuts depending on the chosen type of conjunction and negation.

1 Introduction

There are cases where it is possible to estimate not only the grade of membership to a certain collection of objects but also the grade of non-membership. In case the sum of these grades on a normed scale is 1, then a fuzzy set is an appropriate tool to describe such a collection. However, there might be cases when the sum of these quantities is smaller, what means that there is ambiguity in the decision of membership or non-membership. For such cases the IF-sets (intuitionistic fuzzy sets) introduced by Atanassov in [1] and thoroughly studied in [3] are a suitable structure.

In various applications the IF-sets are used to model natural language statements and therefore it is very often impossible to avoid ambiguity. For example, a patient's temperature or blood pressure can be more or less exactly measured, but the symptoms like headache or insomnia can only be described in terms of a natural language. Operations with such statements are studied by the fuzzy logic.

In this work we focus our attention to the cuts (level sets) of IF-sets, which are crisp sets that can represent the chosen IF-set. We show that for IF-sets it is very natural to account the rules of the fuzzy logic even for the definition of

* The research in this paper is partly supported by the Internationalization Plan 2010 of the University of Oviedo, the Foundation for the promotion in Asturias of the scientific and technologic research grant BP10-090, by the Agency of the Slovak Ministry of Education for the Structural Funds of the EU, under project ITMS:26220120007 and the Spanish Ministry of Science and Innovation grant MTM2010-17844.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 33–40, 2011.

a cut. We study the relations between the chosen triangular norm as well as the negation and the resulting cuts.

2 Basic Concepts

By an IF-set we will understand an object described by a pair of real functions in the following sense:

Definition 1. *Let X be a universe of discourse. A pair $A = \langle \mu_A, \nu_A \rangle$ where $\mu_A, \nu_A : X \to [0,1]$ are functions fulfilling $\mu_A(x) + \nu_A(x) \leq 1$ for all $x \in X$ is called an* **intuitionistic fuzzy set** *(IF-set for short). The functions μ_A, ν_A are its membership and non-membership functions.*

Clearly each fuzzy set with a membership function μ can be understood as an IF-set $\langle \mu, 1 - \mu \rangle$. Hence IF-sets are a generalization of fuzzy sets.

The **support** of an IF-set A is the crisp set

$$supp(A) = \{x \in X | \mu_A(x) > 0 \text{ and } \nu_A(x) < 1\}.$$

By a **triangular norm** (t-norm) we understand a mapping $T : [0,1]^2 \to [0,1]$ that is associative, commutative, monotone in both variables and has 1 as the neutral element.

The most frequently used t-norms are:

1. Minimum t-norm: $T_M(x,y) = \min(x,y)$ for all $x, y \in [0,1]$.
2. Product t-norm: $T_P(x,y) = x \cdot y$, for all $x, y \in [0,1]$.
3. Łukasiewicz t-norm: $T_L(x,y) = \max(x + y - 1, 0)$, for all$x, y \in [0,1]$.

It is well-known that they are ordered in the following sense: $T_L \leq T_P \leq T_M$.

Recall that a *zero divisor* for a t-norm T is a number $x > 0$, for which there exists $y > 0$ such that $T(x,y) = 0$.

A **triangular conorm** (t-conorm) is a mapping $S : [0,1]^2 \to [0,1]$ that is associative, commutative, monotone, with neutral element 0.

By a **fuzzy negation** we understand a non-increasing mapping $N : [0,1] \to [0,1]$ such that $N(0) = 1, N(1) = 0$. If moreover N is continuous and decreasing, it is called **strict**. Fuzzy negations satisfying the involutive property, i.e. $N(N(x)) = x$, for all $x \in [0,1]$, are called **strong** fuzzy negations. Notice that each strong fuzzy negation is strict but the reverse is not true. The most classical example of fuzzy negation is the **standard negation** defined by $N_s(x) = 1 - x$, for all $x \in X$.

For every t-norm T and strong negation N, the operation S defined by $S(x,y) = N(T(N(x), N(y)))$, for all $(x,y) \in [0,1]^2$ is a t-conorm. Moreover, $T(x,y) = N(S(N(x), N(y)))$, for all $(x,y) \in [0,1]^2$. Then S is called the N-dual t-conorm to T. A triplet (T, S, N) is called a **De Morgan triplet** if and only if T is a t-norm, S is a t-conorm, N is a strong negation and $S(x,y) = N(T(N(x), N(y)))$, for all $(x,y) \in [0,1]^2$.

In order to generalize the definitions of set operations, Deschrijver and Kerre presented in 2002 [6] the following definitions: for every triplet (T, S, N) of t-norm, t-conorm and negation and for every pair of IF-sets A and B in $IFSs(X)$ (from now on, $IFSs(X)$ denotes the set of all intuitionistic fuzzy sets in a set X):

1. Intersection: $A \cap B = \langle \mu_{A\cap B}, \nu_{A\cap B} \rangle$ where $\mu_{A\cap B}(x) = T(\mu_A(x), \mu_B(x))$ and $\nu_{A\cap B}(x) = S(\nu_A(x), \nu_B(x))$, for all $x \in X$;
2. Union: $A \cup B = \langle \mu_{A\cup B}, \nu_{A\cup B} \rangle$ where $\mu_{A\cup B}(x) = S(\mu_A(x), \mu_B(x))$ and $\nu_{A\cup B}(x) = T(\nu_A(x), \nu_B(x))$, for all $x \in X$;
3. Complement: if $A = \langle \mu_A, \nu_A \rangle$, then $\overline{A} = \langle \nu_A, \mu_A \rangle$;
4. Inclusion: $A \subseteq B$ iff $\mu_A(x) \leq \mu_B(x)$ and $\nu_A(x) \geq \nu_B(x)$, for all $x \in X$.

3 Cuts of IF-sets

In applications it is useful for a given IF-set to deal with those sets of elements for which the grade of our assigning to a given collection is at least some prescribed value $\alpha \in [0, 1]$, i.e. α-cuts. For a fuzzy set B, its α-cut is the set $B_\alpha = \{x \in X; B(x) \geq \alpha\}$. For an IF-set the α-cut has been defined in [3] as the set (all the symbols have the same meaning as in Definition 1).

$$A_\alpha = \{x \in X; \mu_A(x) \geq \alpha, \nu_A(x) \leq 1 - \alpha\}.$$

The logic of this choice of is obvious. However, we can understand the statement on the elements belonging to A_α as a conjunction of the expressions: (x belongs to A) **and** (it is not true that x does not belong to A).

If we understand these sentences as fuzzy logic statements, it is natural to replace the conjunction by a triangular norm and the negation in the second one by a fuzzy negation. Some results (but only using the standard negation N_s) have been shown in [7]. Here we deal with the general form of the cuts. Moreover, we obtain a deeper study of the properties of the most possible general definition of cut of a IF-set. In particular they could be used in the case $N = N_s$.

Definition 2. *Let $A = \langle \mu_A, \nu_A \rangle$ be an IF-set on X, T and N a triangular norm and a fuzzy negation, respectively. Then, for all $\alpha \in [0, 1]$, define the α-cut of A as the crisp set*

$$(A)_{T,N,\alpha} = \{x \in X | T(\mu_A(x), N(\nu_A(x))) \geq \alpha\}.$$

Example 1. Let A be an IF-set on $[0, \infty)$ with the following membership and non-membership functions:

$$\mu_A(x) = 1 - x \text{ for } 0 \leq x \leq 1, \mu_A(x) = 0 \text{ for } x > 1,$$
$$\nu_A(x) = x/2 \text{ for } 0 \leq x \leq 2, \mu_A(x) = 1 \text{ for } x > 2.$$

Both membership and non-membership functions are represented in the left part of the Fig. 1, where it is immediate to see that A is an IF-set.

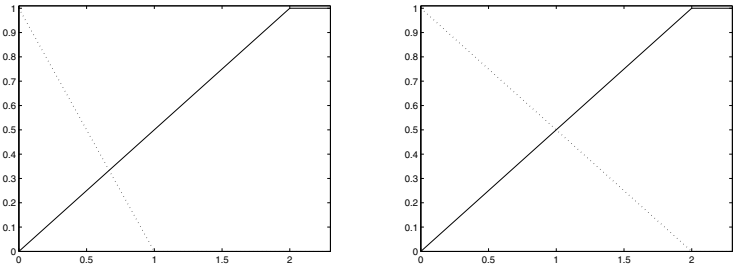


Fig. 1. Membership (dotted line) and non-membership (solid line) functions of A and B

We are going to consider the most frequently used t-norms, different values of α and we will obtain the corresponding α-cuts:

A

$N(x) = 1 - x$	T_L	T_P	T_M
$\alpha = 0.375$	$[0, 0.417]$	$[0, 0.5]$	$[0, 0.625]$
$\alpha = 0.5$	$[0, 0.333]$	$[0, 0.382]$	$[0, 0.5]$

$N^*(x) = 1 - x^2$	T_L	T_P	T_M
$\alpha = 0.375$	$[0, 0.550]$	$[0, 0.589]$	$[0, 0.625]$
$\alpha = 0.5$	$[0, 0.449]$	$[0, 0.471]$	$[0, 0.5]$

On the other hand, if we consider the IF-set B defined by $\mu_B(x) = 1 - \nu_A(x)$ and $\nu_B(x) = \nu_A(x)$, for all $x \in X$, it is clear that $A \subseteq B$ (see Fig. 1). Moreover, the corresponding cuts B are the following:

B

$N(x) = 1 - x$	T_L	T_P	T_M
$\alpha = 0.375$	$[0, 0.625]$	$[0, 0.775]$	$[0, 1.25]$
$\alpha = 0.5$	$[0, 0.5]$	$[0, 0.586]$	$[0, 1]$

$N^*(x) = 1 - x^2$	T_L	T_P	T_M
$\alpha = 0.375$	$[0, 0.871]$	$[0, 1]$	$[0, 1.25]$
$\alpha = 0.5$	$[0, 0.732]$	$[0, 0.806]$	$[0, 1]$

We can notice in the previous example the behavior of cuts in relation with the order between the t-norms ($T_L \leq T_P \leq T_M$), between the values of α ($0.375 \leq 0.5$), between the negations ($N \leq N^*$) and between the IF-sets ($A \subseteq B$). Although this is only an example, general results can be proven, as it is showed in the next proposition, that uses the new Definition 2 and some basic properties of t-norms and negation operators to generalize the results of the preceding example.

Proposition 1. *Let A and B be two IF-sets on X and let α be any element in the interval $(0,1]$. The following statements are true:*

1. *if $\alpha < \beta \leq 1$, then $(A)_{T,N,\alpha} \supseteq (A)_{T,N,\beta}$;*
2. *if $T_1 \leq T_2$, then $(A)_{T_1,N,\alpha} \subseteq (A)_{T_2,N,\alpha}$;*
3. *if $N_1 \leq N_2$, then $(A)_{T,N_1,\alpha} \subseteq (A)_{T,N_2,\alpha}$;*
4. *if $A \subseteq B$, then $(A)_{T,N,\alpha} \subseteq (B)_{T,N,\alpha}$.*

The cuts of an IF-set characterize the support of this set. Thus,

Proposition 2. *Let A be an IF-set on X, T a t-norm without zero divisors and N a negation function such that $N(x) > 0$ for every $x < 1$. Then $supp(A) = \cup_{\alpha>0}\{(A)_{T,N,\alpha}\}$.*

The condition imposed in the previous proposition for the fuzzy negation N is not too restrictive. In fact, let us notice that any strict negation N fulfills it ($N(x) > 0$ for every $x < 1$).

In general, for any fuzzy negation and a particular class of t-norms (left continuous t-norms), we can describe the behavior of the cuts w.r.t. the intersection of sets, by using just the first part of Proposition 1 and the left continuity of the t-norm.

Proposition 3. *Let A be an IF-set on X, let $\alpha \in (0,1]$ and let T be a left continuous t-norm. Then $(A)_{T,N,\alpha} = \cap_{\beta<\alpha}(A)_{T,N,\beta}$.*

This property enables us to formulate a kind of a representation theorem for IF-sets (see [9]). The equivalence of the next proposition is proved using the results of Proposition 3, while the necessary part involves the construction of a particular IF set A such that $\mu_A(x) = \sup\{\gamma \in (0,1] | x \in B_\gamma\}$ that it will be showed to satisfy the desired property.

Proposition 4. *Let $\{B_\alpha\}_{\alpha\in(0,1]}$ be a class of subsets in X. There is an IF-set A on X for which $(A)_{T,N,\alpha} = B_\alpha$ if and only if $B_\alpha = \cap_{\beta<\alpha} B_\beta$ for all $\alpha \in (0,1]$.*

Note that the correspondence between an IF-set A and the family of crisp sets $\{B_\alpha\}_{\alpha\in(0,1]}$ from the Proposition 4 is not one-to-one, as it is shown in the following example.

Example 2. Let $T = T_M$ and let N be an arbitrary negation. Let $A = (\mu_A, \nu_A)$ be an IF-set for which $N(\nu_A(x)) \geq \mu_A(x)$ (note that for the standard negation this inequality holds trivially), let ν_A be not identically zero.

Take $B = (\mu_B, \nu_B)$ where $\mu_B = \mu_A, \nu_B \leq \nu_A$ and for at least one $x \in X$ it holds $\nu_B(x) < \nu_A(x)$. As N is nonincreasing, we have $N(\nu_B(x)) \geq N(\nu_A(x))$ and so

$$T(\mu_A(x), N(\nu_A(x))) = T(\mu_B(x), N(\nu_B(x))) = \mu_A(x)$$

and hence A and B have the same cuts, although $A \neq B$.

4 Operations with IF-sets and Cuts

We will show some properties for different operations with IF-sets. In particular, we will work on the behaviour of the cuts w.r.t. the set operations. In all the cases we will be able to prove a first relationship (content) for an De Morgan triplet, but the equality will only be obtained in particular cases.

Theorem 1. *Let (T, S, N) be a De Morgan triplet and N a strong negation function. Then, for every pair of IF-sets A and B and for every $\alpha \in (0, 1]$, it holds that*

$$(A)_{T,N,\alpha} \cap (B)_{T,N,\alpha} \supseteq (A \cap_T B)_{T,N,\alpha}.$$

$(A)_{T,N,\alpha} \cap (B)_{T,N,\alpha} = (A \cap_T B)_{T,N,\alpha}$ *holds iff* $T(x, y) = \min(x, y)$.

The first part of Theorem 1 has been proved using our new Definition 2 and basic properties of the De Morgan triplet, while the second part is more complex and thus will not be presented here.

An analogous result is obtained for the union, although in this case the conditions for the equality are stronger, since it is not fulfilled for any fuzzy negation. As for the preceding Theorem 1, the first part of the next result can be easily proved using Definition 2 and De Morgan triplet's properties, while the second part is more extended and not reported here.

Theorem 2. *If (T, S, N) is a De Morgan triplet, then, for all $\alpha \in (0, 1]$, it holds that $(A)_{T,N,\alpha} \cup (B)_{T,N,\alpha} \subseteq (A \cup_S B)_{T,N,\alpha}$. The equality $(A)_{T,N,\alpha} \cup (B)_{T,N,\alpha} = (A \cup_S B)_{T,N,\alpha}$ holds iff $T = T_M$ and $N \geq N_s$.*

Once we have characterized the behaviour of cuts with respect to the union and intersection, we will continue with the complementary. In this case, it is not possible to obtain a general result for any $\alpha \in (0, 1]$.

Proposition 5. *If (T, S, N) is a De Morgan triplet, then, for all $\alpha \in (0, 1]$, it holds that $(\overline{A})_{T,N,\alpha} \subseteq \overline{(A_{T,N,\alpha})}$ for any $\alpha > e$, where e is the equilibrium point, that is, the value $e \in (0, 1)$ such that $N(e) = e$.*

A more general result $(\overline{A})_{T,N,\alpha} = \overline{(A_{T,N,\alpha})}$ does not hold.

Thus, let us consider the minimum t-norm, the standard negation ($N(x) = 1 - x$) and the IF-set A defined by $\mu_A(x) = 0.7$, $\nu_A(x) = 0.2$ for every $x \in X$.

For $\alpha = 0.1$, then $\min(\mu_A(x), 1 - \nu_A(x)) = \min(0.7, 0.8) = 0.7 \geq 0.1 = \alpha$ so $x \notin \overline{(A_{T,N,\alpha})}$, while $\min(\nu_A(x), 1 - \mu_A(x)) = \min(0.2, 0.3) = 0.2 \geq 0.1$ and so $x \in (\overline{A})_{T,N,\alpha}$. Hence, $(\overline{A})_{T,N,\alpha} \not\subseteq \overline{(A_{T,N,\alpha})}$. Thus, for $\alpha \leq e$ the inclusion proven in the previous proposition is not true in general.

For $\alpha = 0.9$, then $\min(\mu_A(x), 1 - \nu_A(x)) = 0.7 < 0.9 = \alpha$, so $x \in \overline{(A_{T,N,\alpha})}$, while $\min(\nu_A(x), 1 - \mu_A(x)) = 0.2 < 0.9$ and so $x \notin (\overline{A})_{T,N,\alpha}$. Hence, $\overline{(A_{T,N,\alpha})} \not\subseteq (\overline{A})_{T,N,\alpha}$. In this case $\alpha = 0.9 > e = 0.5$, so we have a counterexample that the equality is not fulfilled in general.

5 Cuts of Fuzzy Sets

It is well-known that the α-cut of an ordinary fuzzy set A with membership function μ_A is defined as the crisp set A_α such that $A_\alpha = \{x \in X | \mu_A(x) \geq \alpha\}$.

On the other hand, we know that any fuzzy set A can be seen as an IF-set, defined by $A = \langle \mu_A, 1 - \mu_A \rangle$ and therefore, we could obtain its cut as an IF-set, that is,

$$A_{T,N,\alpha} = \{x \in X | T(\mu_A(x), N(1 - \mu_A(x))) \geq \alpha\}.$$

Theorem 3. *Let (T, S, N) be a De Morgan triplet, let α be any element in the interval $(0, 1]$ and let A be a fuzzy set. If A_α represents its α-cut as fuzzy set and $(A)_{T,N,\alpha}$ its α-cut if we consider A as an IF-set, then it holds that*

$$(A)_{T,N,\alpha} \subseteq A_\alpha.$$

The equality $(A)_{T,N,\alpha} = A_\alpha$ holds iff $T = T_M$ and $N \geq N_s$.

Both the inclusion $(A)_{T,N,\alpha} \subseteq A_\alpha$ and the equality $(A)_{T,N,\alpha} = A_\alpha$ under conditions $T = T_M$ and $N \geq N_s$ can be easily proved using Definition 2 and De Morgan triplet's properties, while the necessity of conditions $T = T_M$ and $N \geq N_s$ for the equality needs more complex reasonings that are not reported here.

It is interesting to notice that the conditions obtained in this case for fuzzy sets are exactly the same that the ones obtained in Theorem 2 for the union. In both cases, let us notice that only the classical cut defined by Atanassov in [3] fulfills the equality. However, the conditions for the intersection (Theorem 1) are less restrictive and they can be fulfilled for a more general class of cuts, when the fuzzy negation is not necessary the standard one.

6 Conclusion

A new method for computing α-cuts has been proposed representing both statements *x is in A* and *is not true that x is not in A* through a unique expression that depends on a parameter α. The result is a more general definition of α-cut for IF-sets, that respects usual fuzzy logic connectives. The properties of this new definition with respect to usual sets operations has been investigated. It results that stronger properties are verified when the classical t-norm and negation are considered, while in the general case only some properties can be proved.

As in practical application the results of the statements evaluation need not always be comparable, it is reasonable to study similar questions also for lattice (or, more generally, poset) valued fuzzy sets and IF sets. As a starting point for this research direction it is possible to use the results for triangular norms on more general spaces (see [4] or [5]).

Acknowledgement. Grant 1/0297/11 provided by Slovak grant agency VEGA.

References

1. Atanassov, K.: Intuitionistic Fuzzy Sets, VII ITKR Session, Sofia (1983) (in Bulgarian)
2. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
3. Atanassov, K.: Intuitionistic Fuzzy Sets, Theory and Applications. Physica, Heidelberg (1999)
4. Callejas Bedregal, B., Salles Santos, H., Callejas Bedregal, R.: T-norms on bounded lattices: t-norm morphisms and operators. In: IEEE International Conference on Fuzzy Systems, Vancouver, pp. 22–28 (2006)
5. De Baets, B., Mesiar, R.: Triangular norms on product lattices. Fuzzy Sets and Systems 104, 61–75 (1999)
6. Deschrijver, G., Kerre, E.: A generalization of operators on intuitionistic fuzzy sets using triangular norms and conorms. Notes on IFS 8, 19–27 (2002)
7. Janiš, V.: t-Norm based cuts of intuitionistic fuzzy sets. Information Sciences 180, 1134–1137 (2010)
8. Klement, E.P., Mesiar, R., Pap, E.: Triangular norms. Kluwer Academic Publishers, Boston (2000)
9. Negoita, C.V., Ralescu, D.A.: Representation theorems for fuzzy concepts. Kybernetes 4, 169–174 (1975)

Generation of Interval-Valued Fuzzy Implications from K_α Operators

Renata Hax Sander Reiser[1] and Benjamín René Callejas Bedregal[2]

[1] Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas
Campus Capão do Leão, s/n 96010-000 Pelotas, Brazil
reiser@inf.ufpel.edu.br

[2] Depto. de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte
Campus Universitário s/n, 59072-970 Natal, Brazil
bedregal@dimap.ufrn.br

Abstract. This paper introduces the interval-valued fuzzy implications generated from fuzzy implications and from K_α operators showing that such construction generalizes the canonical representation of fuzzy implications. In addition, we also analyzed their conjugate construction preserving their main properties.

Keywords: Interval-valued fuzzy logic, interval-valued implications, interval automorphisms.

1 Introduction

Considering interval-valued fuzzy implications (IV-implications in short), this paper extends the work introduced by Bustince et al. [7], in which to construct IV-connectives (such as t-norms, t-conorms and fuzzy negations) it makes use of K-operators. Such structure enables an interval approach for the fuzzy connectives based on the concept of interval amplitudes. Starting with preliminar definitions of interval representations and fuzzy negations, we recall the $K_\alpha-$operator in Section 3, in order to study the conditions in which IV-implications satisfy the main properties of fuzzy implications. Section 4 considers the conjugate functions based on the action of an automorphism on IV-implications, followed by the Conclusion.

2 Preliminaries

2.1 Interval Representations

Consider the real unit interval $U = [0,1] \subseteq \Re$ and let $\mathbb{U}$ be the set of all subintervals of U, that is, $\mathbb{U} = \{[a,b] \mid 0 \leq a \leq b \leq 1\}$. The projection-functions $l, r : \mathbb{U} \to U$, defined as $l([a,b]) = a$ and $r([a,b]) = b$. For an arbitrary $X \in \mathbb{U}$, $l(X)$ and $r(X)$ will be denoted by $\underline{X}$ and $\overline{X}$, respectively. For each $x \in U$, the interval $[x,x]$ is called a degenerate interval and will be denoted by $\mathbf{x}_{\mathbb{U}}$. Among different relations on interval-valued fuzzy sets, the partial orders that are considered in this paper are the following:
(i) inclusion relation, defined by:

$$\forall X, Y \in \mathbb{U} : X \preceq_{\mathbb{U}} Y \Leftrightarrow \underline{X} \leq \underline{Y} \wedge \overline{Y} \leq \overline{X}. \quad (1)$$

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 41–49, 2011.

(ii) the component-wise *Kulisch-Miranker order* (also called *product order*), given by:

$$\forall X, Y \in \mathbb{U} : X \leq_{\mathbb{U}} Y \Leftrightarrow \underline{X} \leq \underline{Y} \wedge \overline{X} \leq \overline{Y}. \quad (2)$$

A function $F : \mathbb{U}^n \longrightarrow \mathbb{U}$ is an **interval representation** of a function $f : U^n \longrightarrow U$ if, for each $\boldsymbol{X} \in \mathbb{U}^n$ and $\boldsymbol{x} \in \boldsymbol{X}$, $f(\boldsymbol{x}) \in F(\boldsymbol{X})$ [22]. An interval function may be seen as a representation of a subset of real numbers. An interval function $F : \mathbb{U}^n \longrightarrow \mathbb{U}$ is a *better interval representation* of $f : U^n \longrightarrow U$ than $G : \mathbb{U}^n \longrightarrow \mathbb{U}$, denoted by $G \sqsubseteq F$, if, for each $\boldsymbol{X} \in \mathbb{U}^n$, $F(\boldsymbol{X}) \subseteq G(\boldsymbol{X})$. According to [22], the **canonical representation of a real function** $f : U^n \longrightarrow U$, is the interval function $\widehat{f} : \mathbb{U}^n \longrightarrow \mathbb{U}$ defined by

$$\widehat{f}(\boldsymbol{X}) = [\inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}\}, \sup\{f(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}\}]. \quad (3)$$

The interval function $\widehat{f}$ is well defined and for any other interval representation F of f, $F \sqsubseteq \widehat{f}$. It also returns a narrower interval than any other interval representation of f. Thus, $\widehat{f}$ has the *optimality property* of interval algorithms [13].

2.2 K_α Operators

Definition 1. *[7, Definition 2] Let $M_A : \mathcal{X} \to \mathbb{U}$ be a membership function on the universe $\mathcal{X} \neq \emptyset$. An interval-valued fuzzy set A (IVFS A in short) on $\mathcal{X}$ is defined by*

$$A = \{(x, M_A(x)) \, : \, x \in \mathcal{X}\}. \quad (4)$$

According to [7], an IVFS A described by Eq. (4) can be viewed as a fuzzy $L-$Goguen's sense. In addition, $(\mathbb{U}, \leq_{\mathbb{U}})$ is a complete lattice when, for all $X \in \mathbb{U}$, $\mathbf{0}_{\mathbb{U}} \leq_{\mathbb{U}} X \leq_{\mathbb{U}} \mathbf{1}_{\mathbb{U}}$.

Definition 2. *[7, Def. 3] For $\alpha \in U$, a function $K_\alpha : \mathbb{U} \longrightarrow U$ is a* $\mathbf{K}-$***operator*** *if:*

K1 $K_\alpha(\mathbf{x}_{\mathbb{U}}) = x$, *for all* $x \in U$;
K2 $K_0(X) = \underline{X}$, $K_1(X) = \overline{X}$, *for all* $X \in \mathbb{U}$;
K3 *If* $X \leq Y$ *then* $K_\alpha(X) \leq K_\alpha(Y)$, *for all* $X, Y \in \mathbb{U}$ *and* $\alpha \in U$;
K4 $\alpha \leq \beta$ *iff* $K_\alpha(X) \leq K_\beta(X)$, *for all* $X \in \mathbb{U}$.

Proposition 1. *Let $\alpha \in U$. Then $K_\alpha(K^\alpha) : \mathbb{U} \longrightarrow U$ is a $K-$operator defined by*

$$K_\alpha(X) = K_0(X) + \alpha W_X \quad (or \quad K^\alpha(X) = K_1(X) - \alpha W_X). \quad (5)$$

where W_X denotes the amplitude of the interval X, i.e. $W_X = \overline{X} - \underline{X}$.

Proof. Based on Def. 2, for all $X, Y \in \mathbb{U}$ and $\alpha \in U$, it follows that: (i) $K_0(\mathbf{x}_{\mathbb{U}}) = K_0(\mathbf{x}_{\mathbb{U}}) + \alpha \cdot 0 = K_0(\mathbf{x}_{\mathbb{U}}) = x$; (ii) firstly, if $\alpha = 0$ then $K_0(X) = K_0(X) + 0 \cdot W_X = \underline{X}$; and, if $\beta = 1$ it holds that $K_1(X) = K_0(X) + 1 \cdot W_X = \overline{X}$; (iii) if $X \leq Y$ then $K_\alpha(X) = K_0(X) + \alpha W_X \leq K_0(Y) + \alpha W_Y = K_\alpha(Y)$; (iv) $\alpha \leq \beta \leftrightarrow K_0(X) + \alpha W_X \leq K_0(X) + \beta W_X$, and so, $\alpha \leq \beta$ iff $K_\alpha(X) \leq K_\beta(X)$.

The role of W_X in Eq. (5) is to guarantee that $K_\alpha(X) \in X$ and $K^\alpha(X) \in X$.

Remark 1. When K_α and K^α are both $K-$operators, then $K_{N_S(\alpha)} = K^\alpha$. In this paper, we consider the generation of IV-implications from K_α operators, but it can also be done based on K^α operators.

Remark 2. Let $X \in \mathbb{U}$, $n \in \mathcal{N}^*$ and let K_α be the $K-$operator of equation (5):

(i) If $\underline{X} \leq \alpha = \frac{1}{n}$ and $\overline{X} = 1 - (n-1)\underline{X}$, then $K_\alpha(X) \leq \alpha$;
(ii) When $\overline{X} \leq 1 + (1 - \alpha^{-1})\underline{X}$ then $K_\alpha \leq \alpha$;
(iii) In addition, if $\alpha = 0.5$, $K_{0.5}(X)$ returns the middle-point of X.

2.3 Interval-Valued Fuzzy Negation

A function $N : U \to U$ is a *fuzzy negation* if

N1 : $N(0) = 1$ and $N(1) = 0$.
N2 : If $x \geq y$ then $N(x) \leq N(y)$, $\forall x, y \in I$.
In addition, fuzzy negations satisfying the involutive property are called strong fuzzy negations (SFN in short),see [15] and [6]:

N3 : $N(N(x)) = x$, $\forall x \in U$.

$N_S(x) = 1 - x$, called the standard negation, is an involutive function on U.

An interval function $\mathbb{N} : \mathbb{U} \longrightarrow \mathbb{U}$ is an **interval fuzzy negation** if, for any X, Y in $\mathbb{U}$, the following properties hold:

$\mathbb{N}1$: $\mathbb{N}([0,0]) = [1,1]$ and $\mathbb{N}([1,1]) = [0,0]$.
$\mathbb{N}2a$: If $X \geq Y$ then $\mathbb{N}(X) \leq \mathbb{N}(Y)$; and $\mathbb{N}2b$: If $X \subseteq Y$ then $\mathbb{N}(X) \supseteq \mathbb{N}(Y)$.

If $\mathbb{N}$ meets the involutive property, it is a strong interval fuzzy negation (SIFN in short):

$\mathbb{N}3$: $\mathbb{N}(\mathbb{N}(X)) = X$, $\forall X \in \mathbb{U}$.

$\mathbb{N}_S(X) = [1 - \overline{X}, 1 - \underline{X}]$ is a SIFN. Notice that $\mathbb{N}_S = \widehat{N_S}$.

Let $N : U \longrightarrow U$ be a negation. From Def. 2, a characterization of $\widehat{N}$ is given by:

$$\widehat{N}(X) = [N(\overline{X}), N(\underline{X})] = [N(K_1(X)), N(K_0(X))]. \tag{6}$$

Theorem 1. *[4, Theorem 4.2] Let $N : U \longrightarrow U$ be a (strong) fuzzy negation. Then $\widehat{N}$ is an (strong) interval fuzzy negation.*

Based on [7, Theorem 4,Corollary 2] when $N : U \to U$ is a strict fuzzy negation and $\mathbb{N}_1, \mathbb{N}_2 : \mathbb{U} \to \mathbb{U}$ are defined by,

$$\mathbb{N}_1(X) = [N(K_1(X)), N(K_0(X))] \text{ and } \mathbb{N}_2(X) = [N^{-1}(K_1(X)), N^{-1}(K_0(X))], \tag{7}$$

then, both, $\mathbb{N}_1$ and $\mathbb{N}_2$ are strict IV-negations. Moreover, $\mathbb{N}_1 = \mathbb{N}_2$ iff N is involutive.

Concluding this section, an $N-$dual $K_\alpha-$operator is introduced.

Proposition 2. *If $\mathbb{N} = \mathbb{N}_S$ then $W_X = W_{\mathbb{N}(X)}$.*

Proof. $W_{\mathbb{N}(X)} = N_S(K_0(X)) - N_S(K_1(X)) = (1 - K_0(X)) - (1 - K_1(X)) = W_X$.

3 Interval-Valued Fuzzy Implications

Several definitions for implications together with related properties have been given (see [1,2,6,10,14,17,19] and [23]). The agreement over these definitions is that the fuzzy implication should have the same behavior as the classical implication for the crisp case. Thus, $I : U^2 \longrightarrow U$ is a *fuzzy implication* if I meets the boundary conditions:

I1 : $I(1,1) = I(0,1) = I(0,0) = 1$ and $I(1,0) = 0$.

The properties of implications can be naturally extended from interval fuzzy degrees, when the respective degenerate intervals are considered. A function $\mathbb{I} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an **IV-implication**, if the following conditions hold:

$\mathbb{I}1$: $\mathbb{I}(1_{\mathbb{U}}, 1_{\mathbb{U}}) = \mathbb{I}(0_{\mathbb{U}}, 0_{\mathbb{U}}) = \mathbb{I}(0_{\mathbb{U}}, 1_{\mathbb{U}}) = 1_{\mathbb{U}}$ and $\mathbb{I}(1_{\mathbb{U}}, 0_{\mathbb{U}}) = 0_{\mathbb{U}}$.

Among several other properties which may be required for fuzzy implications, the following ones are considered in this paper:

I2: if $x \leq z$ then $I(x,y) \geq I(z,y)$;	$\mathbb{I}2$: if $X \leq_{\mathbb{U}} Z$ then $\mathbb{I}(X,Y) \geq_{\mathbb{U}} \mathbb{I}(Z,Y)$;
I3: if $y \leq t$ then $I(x,y) \geq I(x,t)$;	$\mathbb{I}3$: if $Y \leq_{\mathbb{U}} T$ then $\mathbb{I}(X,Y) \geq_{\mathbb{U}} \mathbb{I}(X,T)$;
I4: $I(0,y) = 1$;	$\mathbb{I}4$: $\mathbb{I}(\mathbf{0}_{\mathbb{U}}, Y) = \mathbf{1}_{\mathbb{U}}$;
I5: $I(x,1) = 1$;	$\mathbb{I}5$: $\mathbb{I}(X, \mathbf{1}_{\mathbb{U}}) = \mathbf{1}_{\mathbb{U}}$;
I6: $I(1,y) = y$;	$\mathbb{I}6$: $\mathbb{I}(\mathbf{1}, Y) = Y$;
I7: $I(x, I(y,z) = I(y, I(x,z))$;	$\mathbb{I}7$: $\mathbb{I}(X, \mathbb{I}(Y,Z) = \mathbb{I}(Y, \mathbb{I}(X,Z))$;
I8: $I(x,y) = 1$ iff $x \leq y$;	$\mathbb{I}8$: $\mathbb{I}(X,Y) = \mathbf{1}$ iff $X \leq_{\mathbb{U}} Y$;
I9: $I(x,y) \geq y$;	$\mathbb{I}9$: $\mathbb{I}(X,Y) \geq_{\mathbb{U}} Y$;
I10: $I(x,x) = 1$;	$\mathbb{I}10$: $\mathbb{I}(X,X) = \mathbf{1}_{\mathbb{U}}$;
I11: $I(x,y) = I(N(y), N(x))$;	$\mathbb{I}11$: $\mathbb{I}(X,Y) = \mathbb{I}(\mathbb{N}(Y), \mathbb{N}(X))$;
I12: $x > 0 \rightarrow I(x,0) < 1$;	$\mathbb{I}12$: $X >_{\mathbb{U}} \mathbf{0} \rightarrow \mathbb{I}(X, \mathbf{0}_{\mathbb{U}}) <_{\mathbb{U}} \mathbf{1}_{\mathbb{U}}$;
I13: $y < 1 \rightarrow I(1,y) < 1$.	$\mathbb{I}13$: $Y <_{\mathbb{U}} \mathbf{1}_{\mathbb{U}} \rightarrow \mathbb{I}(\mathbf{1}_{\mathbb{U}}, \mathbf{Y}) <_{\mathbb{U}} \mathbf{1}_{\mathbb{U}}$.

Proposition 3. *[5, Proposition 21] A fuzzy implication $I : U^2 \longrightarrow U$ satisfies* **I**2 *and* **I**3 *iff the IV-implication $\widehat{I}$, called* ***canonical representation*** *of I, can be expressed as*

$$\widehat{I}(X,Y) = [I(\overline{X}, \underline{Y}), I(\underline{X}, \overline{Y})]. \tag{8}$$

Proposition 4. *Let $K_\alpha, K_\beta : \mathbb{U} \longrightarrow U$ be $K-$operators such that $\alpha \leq \beta$ and $I_a, I_b : U^2 \rightarrow U$ be fuzzy implications satisfying* **I**2 *and* **I**3 *and such that $I_a \leq I_b$. The mapping $\mathbb{I}_{I_a,I_b} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an IV-implication given by:*

$$\mathbb{I}_{I_a,I_b}(X,Y) = [I_a(K_\beta(X), K_\alpha(Y)), I_b(K_\alpha(X), K_\beta(Y))]. \tag{9}$$

Proof. From $\mathbf{K}4$, **I**2 and **I**3 follows that $I_a(K_\beta(X), K_\alpha(Y)) \leq I_b(K_\alpha(X), K_\beta(Y))$ and so $\mathbb{I}_{I_a,I_b}$ is well defined. Following from Properties $\mathbf{K}1$ and $\mathbf{I}1$, $\mathbb{I}_{I_a,I_b}$ verifies $\mathbb{I}1$.

Remark 3. Connectives and Amplitudes: Let N_S and S_M be the standard negation and the maximum t-conorm. According to [7, Prop. 5 and Lemma 11], taking their corresponding N-representable IV negation $N_S(X) = [N(\overline{X}), N(\underline{X})]$ and S-representable IV t-conorm $S(X,Y) = [\max\{\overline{X}, \underline{Y}\}, \max\{\underline{X}, \overline{Y}\}]$, if $W_X = W_Y$ then $W_{\mathbb{N}} = W_X$ and $W_{\mathbb{S}} = W_X$. Now, taking the Klenne-Dienes implication obtained by $I_{S_M,N_S}(x,y) = S_M(N_S(x), y)$, which means $I_{S_M,N_S}(x,y) = \max\{N(x), y\}$, then it also holds that $W_{\mathbb{I}_{S_M,N_S}} = W_X$ if its corresponding I_{S_M,N_S}-representable IV-implication is given by:

$$\mathbb{I}_{S_M,N_S}(X,Y) = [max\{N(\overline{X}), \underline{Y}\}, max\{N(\underline{X}), \overline{Y}\}].$$

In this case, $a = b$, $\alpha = 0$ and $\beta = 1$, in Eq.(9).

Proposition 5. *Let $I : U^2 \rightarrow U$ be fuzzy implications satisfying* **I**2 *and* **I**3 *and $K_\alpha, K_\beta : \mathbb{U} \longrightarrow U$ be $K-$operators according to Def. 2. When $\alpha = 0$ and $\beta = 1$, then*

$$\mathbb{I}_{I_a,I_b} = \widehat{I}. \tag{10}$$

Proof. It follows from Eq. (9) in Proposition 5 and from Proposition 3.

Theorem 2. *Let $I_a, I_b : U^2 \rightarrow U$ be fuzzy implications satisfying* **I**2 *and* **I**3 *and such that $I_a \leq I_b$. Let $K_\alpha, K_\beta : \mathbb{U} \longrightarrow U$ be $K-$operators such that $\alpha \leq \beta$. The function $\mathbb{I}_{I_a,I_b}$ satisfies the property $\mathbb{I}k$ iff I_a, I_b satisfy the property $\mathbf{I}k$, for $k \in \{1, \ldots, 5, 12, 13\}$.*

Proof. $\mathbb{I}2$: Suppose that both I_a, I_b satisfy $\mathbf{I}2$. If $X \leq_{\mathbb{U}} Z$ then, by $\mathbf{K}3$ and $\mathbf{I}2$, for all $X, Z \in \mathbb{U}$ and $\beta, \alpha \in U$, it follows that:

$$(i) K_\beta(X) \leq K_\beta(Z) \Leftrightarrow I_a(K_\beta(X), K_\alpha(Y)) \geq I_a(K_\beta(Z), K_\alpha(Y))$$
$$(ii) K_\alpha(X) \leq K_\alpha(Z) \Leftrightarrow I_b(K_\alpha(X), K_\beta(Y)) \geq I_b(K_\alpha(Z), K_\beta(Y))$$

From (i) and (ii), it is immediate that $\mathbb{I}_{I_a,I_b}(X,Y) \geq_{\mathbb{U}} \mathbb{I}_{I_a,I_b}(X,Z)$.

$\mathbb{I}3$: Suppose that I_a, I_b satisfy $\mathbf{I}3$. If $Y \leq_{\mathbb{U}} Z$, then it follows that:

$$(i) K_\alpha(Y) \leq K_\alpha(Z) \Leftrightarrow I_a(K_\beta(X), K_\alpha(Y)) \leq I_a(K_\beta(X), K_\alpha(Z));$$
$$(ii) K_\beta(Y) \leq K_\beta(Z) \Leftrightarrow I_b(K_\alpha(X), K_\beta(Y)) \leq I_b(K_\alpha(X), K_\beta(Z)).$$

Therefore, from (i) and (ii), $\mathbb{I}_{I_a,I_b}(X,Y) \leq_{\mathbb{U}} \mathbb{I}_{I_a,I_b}(X,Z)$.

$\mathbb{I}4$: Suppose that I_a, I_b satisfy $\mathbf{I}4$. Then, by Eq.(9) and $\mathbf{K}1$, for all Y, one can obtain:

$$\begin{aligned}\mathbb{I}_{I_a,I_b}(\mathbf{0}_{\mathbb{U}}, Y) &= [I_a(K_\beta(\mathbf{0}_{\mathbb{U}}), K_\alpha(Y)), I_b(K_\alpha(\mathbf{0}_{\mathbb{U}}), K_\beta(Y))] \\ &= [I_a(0, K_\alpha(Y)), I_b(0, K_\beta(Y))] = [1,1] = \mathbf{1}_{\mathbb{U}}\end{aligned}$$

$\mathbb{I}5$: If I_a, I_b satisfy $\mathbf{I}5$, by Eq.(9) and $\mathbf{K}1$, for all X, one can obtain:

$$\begin{aligned}\mathbb{I}_{I_a,I_b}(X, \mathbf{1}_{\mathbb{U}}) &= [I_a(K_\beta(X), K_\alpha(\mathbf{1}_{\mathbb{U}})), I_b(K_\alpha(X), K_\beta(\mathbf{1}_{\mathbb{U}}))] \\ &= [I_a(K_\beta(X), 1), I_b(K_\alpha(X), 1)] = [1,1] = \mathbf{1}_{\mathbb{U}}.\end{aligned}$$

$\mathbb{I}12$: Suppose I_a, I_b satisfy $\mathbf{I}12$ and for $X \in \mathbb{U}$, $X > 0$, which means, by $\mathbf{K}3$, $K_\alpha(X) \geq 0$ and $K_\beta(X) > 0$. Then, by Eq.(9) and $\mathbf{K}1$, one can obtain:

$$\begin{aligned}\mathbb{I}_{I_a,I_b}(X, \mathbf{0}_{\mathbb{U}}) &= [I_a(K_\beta(X), K_\alpha(\mathbf{0}_{\mathbb{U}})), I_b(K_\alpha(X), K_\beta(\mathbf{0}_{\mathbb{U}}))] \\ &= [I_a(K_\beta(X), 0), I_b(K_\alpha(X), 0)] < \mathbf{1}_{\mathbb{U}}.\end{aligned}$$

$\mathbb{I}13$: Suppose I_a, I_b satisfy $\mathbf{I}13$ and for $Y \in \mathbb{U}$, $Y < 1$, which means, by $\mathbf{K}3$, $K_\alpha(Y) < 1$ and $K_\beta(Y) < 1$. Then, by Eq.(9) and $\mathbf{K}1$, one can also obtain:

$$\begin{aligned}\mathbb{I}_{I_a,I_b}(\mathbf{1}_{\mathbb{U}}, Y) &= [I_a(K_\beta(\mathbf{1}_{\mathbb{U}}), K_\alpha(Y)), I_b(K_\alpha(\mathbf{1}_{\mathbb{U}}), K_\beta(Y))] \\ &= [I_a(1, K_\alpha(Y)), I_b(1, K_\beta(Y))] < \mathbf{1}_{\mathbb{U}}.\end{aligned}$$

Hereinafter, in order to analise other properties, from the two next propositions, necessary conditions are also presented, in addition to the constrain stated in Proposition 5.

Proposition 6. *Let I_a and I_b be implications verifying* $\mathbf{I}6$ *(*$\mathbf{I}7$,$\mathbf{I}9$, $\mathbf{I}10$*). If $\beta = 1$ and $\alpha = 0$ then function $\mathbb{I}_{I_a,I_b}$ satisfies the property* $\mathbb{I}6$ *(*$\mathbb{I}7$, $\mathbf{I}9$, $\mathbb{I}10$*), respectively.*

Proof. $\mathbb{I}6$: When I_a and I_b are implications verifying $\mathbf{I}6$ then:

$$\mathbb{I}_{I_a,I_b}(\mathbf{1}_{\mathbb{U}}, Y) = [I_a(K_1(\mathbf{1}_{\mathbb{U}}), K_0(Y)), I_b(K_0(\mathbf{1}_{\mathbb{U}}), K_1(Y)] = [K_0(Y), K_1(Y)] = Y.$$

$\mathbb{I}7$: Let I_a and I_b be implications verifying $\mathbf{I}7$. then:

$$\begin{aligned}&\mathbb{I}_{I_a,I_b}(X, \mathbb{I}_{I_a,I_b}(Y,Z)) \\ &= [I_a(K_1(X), I_a(K_1(Y), K_0(Z)), I_b(K_0(X), I_b(K_0(Y), K_1(Z)))] \text{by Eq (9)} \\ &= [I_a(K_1(Y), I_a(K_1(X), K_0(Z)), I_b(K_0(Y), I_b(K_0(X), K_1(Z)))] \text{by } \mathbf{I}7 \\ &= \mathbb{I}_{I_a,I_b}(Y, \mathbb{I}_{I_a,I_b}(X,Z))\end{aligned}$$

$\mathbb{I}9$: When I_a and I_b are implications verifying **I**9 then:

$$\mathbb{I}_{I_a,I_b}(X,Y)=[I_a(K_1(X),K_0(Y)),I_b(K_0(X),K_1(Y)]\leq_{\mathbb{U}}[K_0(Y),K_1(Y)]=Y.$$

So, $\mathbb{I}_{I_a,I_b}$, for $\alpha=0,\beta=1$, satisfies $\mathbb{I}6$, $\mathbb{I}7$ and $\mathbb{I}9$ if I_a, I_b satisfy **I**6, **I**7 and **I**9.

Proposition 7. *Let I_a and I_b be implications verifying* **I**8 *(***I**10*) . If $\beta=\alpha$ and $K_\alpha(X)\leq K_\alpha(Y)$ implies that $X\leq_{\mathbb{U}}Y$ for all $X,Y\in\mathbb{U}$, then $\mathbb{I}_{I_a,I_b}$ satisfies* $\mathbb{I}8$ *($\mathbb{I}$10).*

Proof. ($\Rightarrow$) If $\beta=\alpha$ then for all $X,Y\in\mathbb{U}$, whenever $X\leq_{\mathbb{U}}Y$, by **K**3, $K_\alpha(X)\leq K_\beta(Y)$ and $K_\beta(X)\leq K_\alpha(Y)$. So, by **I**8, it holds that:

$$\mathbb{I}_{I_a,I_b}(X,Y)=[I_a(K_\beta(X),K_\alpha(Y)),I_b(K_\alpha(X),K_\beta(Y))]=\mathbf{1}_{\mathbb{U}}.$$

($\Leftarrow$) If $\mathbb{I}_{I_a,I_b}(X,Y)=\mathbf{1}_{\mathbb{U}}$, then $I_a(K_\beta(X),K_\alpha(Y))=1$ and $I_b(K_\alpha(X),K_\beta(Y))=1$. Again, since both I_a and I_b satisfy **I**8, it follows that $K_\beta(X)\leq K_\alpha(Y)$ and $K_\alpha(X)\leq K_\beta(Y)$. Since $\alpha=\beta$ then $K_\alpha(X)\leq K_\alpha(Y)$. So, by hypothesis, a $K-$operator verifies **K4** $X\geq Y$.

For the case of **I**10, when $\alpha=\beta$:

$$\mathbb{I}_{I_a,I_b}(X,X)=[I_a(K_\alpha(X),K_\alpha(X)),I_b(K_\alpha(X),K_\alpha(X))]=\mathbf{1}_{\mathbb{U}}.$$

4 Interval Automorphism

A mapping $\rho:U\longrightarrow U$ is an **automorphism** if it is bijective and monotonic: $x\leq y\ \Rightarrow\ \rho(x)\leq\rho(y)$ [16]. An equivalent definition is given in [6], where $\rho:U\longrightarrow U$ is an automorphism if it is a continuous and strictly increasing function such that $\rho(0)=0$ and $\rho(1)=1$. The set of all automorphisms on U is denoted by $Aut(U)$. Automorphisms are closed under composition and inversion operators: if ρ and ρ' are automorphisms then $\rho\circ\rho'(x)=\rho\circ\rho'(x))$ an $\rho^{-1}\circ\rho(x)=x$.

The **action of** ρ **on** I, denoted by I^ρ, is also a fuzzy implication defined as follows

$$I^\rho(x,y)=\rho^{-1}(I(\rho(x),\rho(y))) \tag{11}$$

A mapping $\varrho:\mathbb{U}\longrightarrow\mathbb{U}$ is an **interval automorphism** if it is bijective and satisfies the equivalence: $X\leq_{\mathbb{U}}Y\Leftrightarrow\varrho(X)\leq_{\mathbb{U}}\varrho(Y)$ [11,12].

The set of all interval automorphisms $\varrho:\mathbb{U}\longrightarrow\mathbb{U}$ is denoted by $Aut(\mathbb{U})$. In what follows, Eq.(12) provides a canonical construction of interval automorphisms from automorphisms and, therefore a bijection between the sets $Aut(U)$ and $Aut(\mathbb{U})$.

Theorem 3. *[11, Theorem 3] Let $\varrho:\mathbb{U}\longrightarrow\mathbb{U}$ be an interval automorphism. Then there exists an automorphism $\rho:U\longrightarrow U$ such that*

$$\varrho(X)=[\rho(\underline{X}),\rho(\overline{X})]. \tag{12}$$

Interval automorphisms can be generated from the point of view of an automorphism representation. Based on automorphism representation theorem proved in[3], $\widehat{\rho}:\mathbb{U}\to\mathbb{U}$ is an interval automorphism characterized as:

$$\widehat{\rho}(X)=[\rho(\underline{X}),\rho(\overline{X})]. \tag{13}$$

In the following, an interval automorphism ϱ acts on an IV-implication $\mathbb{I}$ generating new IV-implication $\mathbb{I}^\varrho$ named conjugate function of $\mathbb{I}$.

Let $\varrho : \mathbb{U} \longrightarrow \mathbb{U}$ be an interval automorphism and $\mathbb{I} : \mathbb{U}^2 \longrightarrow \mathbb{U}$ be an IV-implication. Then the mapping $\mathbb{I}^\varrho : \mathbb{U}^2 \longrightarrow \mathbb{U}$ is an IV-implication given by

$$\mathbb{I}^\varrho(X,Y) = \varrho^{-1}(\mathbb{I}(\varrho(X), \varrho(Y))) \tag{14}$$

The results in the last subsection are concerned with the generation of new IV-implications and K_α-operators based on the action of interval automorphisms.

Proposition 8. *The action of an automorphism ρ on a $K-$operator K_α is also a $K-$operator, defined as follows:*

$$K_\alpha^\rho(X) = \rho^{-1}(K_\alpha(\widehat{\rho}(X))). \tag{15}$$

Proof. Based on Def. 2, it follows the next four properties:

K1 By Eq.(11), $K_\alpha^\rho(\mathbf{x}_\mathbb{U}) = \rho^{-1}(K_\alpha([\rho(x), \rho(x)])) = x$.

K2 By Eq. (12), it follows: (i) $K_0^\rho(X) = \rho^{-1}(K_0([\rho(\underline{X}), \rho(\overline{X})])) = \rho^{-1}(\rho(\underline{X})) = \underline{X}$; and (ii) $K_1^\rho(X) = \rho^{-1}(K_1([\rho(\underline{X}), \rho(\overline{X})]) = \rho^{-1}(\rho(\overline{X}))$. Then, $K_\alpha^\rho(X) = \overline{X}$.

K3 If $X \leq Y$, since K_α, ρ^{-1} and $\widehat{\rho}$ are nondecreasing functions, it holds that: $K_\alpha^\rho(X) = \rho^{-1}(K_\alpha(\widehat{\rho}(X)) \leq \rho^{-1}(K_\alpha(\widehat{\rho}(Y)) =\leq K_\alpha^\rho(Y)$.

K4 Firstly, if $\alpha \leq \beta$, then by **K4**, $K_\alpha(\widehat{\rho}(X)) \leq K_\beta(\widehat{\rho}(Y))$. So, for all $X \in \mathbb{U}$ it holds that $K_\alpha^\rho(X) = \rho^{-1}(K_\alpha(\widehat{\rho}(X))) \leq \rho^{-1}(K_\beta(\widehat{\rho}(Y))) = K_\beta^\rho(Y)$. And, if $K_\alpha^\rho(X) \leq K_\beta^\rho(X)$ then, by **K**4, $\widehat{\rho}(X) \leq_\mathbb{U} \widehat{\rho}(Y)$ which means $X \leq_\mathbb{U} Y$. Thus, the converse also is held.

Therefore, one can conclude that K_α^ρ is a $K-$operator, in the sense of Def. 2.

Theorem 4. *Let $I_a, I_b : U^2 \to U$ be fuzzy implications, $K_\alpha, K_\beta : \mathbb{U} \longrightarrow U$ be $K-$operators according to Def. 2 and $\varrho : \mathbb{U} \to \mathbb{U}$ be an interval automorphism. Then*

$$\mathbb{I}^\varrho_{I_a,I_b}(X,Y) = \mathbb{I}_{I_a^\rho,I_b^\rho}(X,Y) \tag{16}$$

Proof. Firstly, $\mathbb{I}^\varrho_{I_a,I_b}$ is well defined. By Theorem 3, there exists an automorphism $\rho : U \longrightarrow U$ such that $\varrho(X) = [\rho(\underline{X}), \rho(\overline{X})]$. By Theorem 8, it follows that :

$$\begin{aligned}
\mathbb{I}^\varrho_{I_a,I_b}(X,Y) &= \mathbb{I}_{I_a,I_b}(\varrho(X), \varrho(Y)) \text{By Eq.(14)} \\
&= \varrho^{-1}[I_a(K_\beta(\varrho(X)), K_\alpha(\varrho(Y)), I_b(K_\alpha(\varrho(X)), K_\beta(\varrho(Y)))] \text{By Eq.(15)} \\
&= \varrho^{-1}[I_a(\rho(K_\beta^\rho(X), \rho(K_\alpha^\rho(X)), I_b(\rho(K_\alpha^\rho(X), \rho(K_\beta^\rho(Y))] \text{By Eq.(11)} \\
&= [I_a^\rho(K_\beta^\rho(X), K_\alpha^\rho(Y), I_b^\rho(K_\alpha^\rho(X), K_\beta^\rho(Y)] \text{By Eq.(11)} \\
&= \mathbb{I}_{I_a^\rho,I_b^\rho}(X,Y), \text{By Eq.(9)}.
\end{aligned}$$

Theorem 5. *The conjugate function $\mathbb{I}^\varrho_{I_a,I_b}$ satisfies the property $\mathbb{I}k$ iff $\mathbb{I}_{I_a,I_b}$ satisfies the property $\mathbb{I}k$, for $k \in \{1, 2, \ldots, 5, 13, \ldots 15\}$.*

Proof. It can be obtained based on Theorem 4, analogously to Theorem 5.

5 Conclusion and Final Remarks

We mainly discussed under which conditions the generation of IV-implications from fuzzy implications and from K_α operators preserve the main properties of implications. The conjugate of an IV-implication $\mathbb{I}_{I_a,I_b}$ is studied, based on the action of an automorphism ρ on their underlying implications I_a and I_b and on K-operators. Ongoing work is focussed on the extension of such approach in order to study the S-, QL- and R-implication interval classes including interval-valued t-norms and t-conorms [8]. IV-implications could be used to extend the use of fuzzy implications in, for example, neuro-fuzzy systems [20], fuzzy systems control [18] and making-decision [21].

Acknowledgments. Thanks for the anonymous referees for their valuable comments and suggestions.

References

1. Baczynsk, M.: Residual Implications Revisited, Notes on the Smets-Magrez. Fuzzy Sets and Systems 145(2), 267–277 (2004)
2. Balasubramaniam, J.: Yager's New Class of Implications J_f and Some Classical Tautologies. Inf. Sci. 177(3), 930–946 (2007)
3. Bedregal, B., Takahashi, A.: The Best Interval Representation of T-Norms and Automorphisms. Fuzzy Sets and Systems 157(24), 3220–3230 (2006)
4. Bedregal, B., Takahashi, A.: Interval-Valued Versions of T-conorms, Fuzzy Negations and Fuzzy Implications. In: IEEE Proc. of the Intl. Conf. on Fuzzy Systems, pp. 1981–1987. IEEE Press, Vancouver (2006)
5. Bedregal, B., Dimuro, G., Santiago, R., Reiser, R.: On Interval Fuzzy S-implications. Inf. Sci. 180(8), 1373–1389 (2010)
6. Bustince, H., Burillo, P., Soria, F.: Automorphism, Negations and Implication Operators. Fuzzy Sets and Systems 134(2), 209–229 (2003)
7. Bustince, H., Barrenechea, E., Pagola, M.: Generation of Interval-valued Fuzzy and Atanossov's Intuitionistic Fuzzy Connectives from Fuzzy Connectives and from K_α Operators: Law of Conjuntions and Disjuntions, Amplitute. Int. J. Intell. Syst. 23(6), 680–714 (2008)
8. Dimuro, G., Bedregal, B., Santiago, R., Reiser, R.: Interval Additive Generators of Interval T-Norms and Interval T-Conorms. Information Science 181, 3898–3916 (2011)
9. Fodor, J.C.: Contrapositive Symmetry of Fuzzy Implications. Fuzzy Sets and Systems 69(2), 141–156 (1995)
10. Fodor, J.C., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Publisher, Dordrecht (1994)
11. Gehrke, M., Walker, C., Walker, E.: Some Comments on Interval Valued Fuzzy Sets. Int. J. Intell. Syst. 11, 751–759 (1996)
12. Gehrke, M., Walker, C., Walker, E.: Algebraic Aspects of Fuzzy Sets and Fuzzy Logics. In: Workshop on Current Trends and Development in Fuzzy Logic, Thessaloniki-Greece, pp. 101–170 (1998)
13. Hickey, T., Ju, Q., Emdem, M.: Interval Arithmetic: from Trinciples to Implementation. J. of the ACM 48(5), 1038–1068 (2001)
14. Horcik, R., Navara, M.: Validation Sets in Fuzzy Logics. Kybernetika 38(2), 319–326 (2002)
15. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer, Dordrecht (2000)

16. Klement, E.P., Navara, M.: A survey on different triangular norm-based fuzzy logics. Fuzzy Sets and Systems 101(2), 241–251 (1999)
17. Leski, J.M.: ε-Insensitive Learning Techniques for Approximate Reasoning System. Int. J. of Computational Cognition 1(1), 21–77 (2003)
18. Piskunov, A.: Fuzzy Implication in Fuzzy Systems Control. Fuzzy Sets and Systems 45, 25–35 (1995)
19. Ruan, D., Kerre, E.E.: Fuzzy Implication Operators and Generalized Fuzzy Methods of Cases. Fuzzy Sets and Systems 54, 23–37 (1993)
20. Rutkowska, D., Nowicki, R., Hayashi, Y.: Parallel Processing by Implication-based Neuro-fuzzy Systems. In: Wyrzykowski, R., Dongarra, J., Paprzycki, M., Waśniewski, J. (eds.) PPAM 2001. LNCS, vol. 2328, pp. 539–542. Springer, Heidelberg (2002)
21. Sandri, S., Sibertin-Blanc, C., Torra, V.: Multicriteria Fuzzy System Using Residuated Implication Operatorss and Fuzzy Arithmetic. In: Torra, V., Narukawa, Y., Yoshida, Y. (eds.) MDAI 2007. LNCS (LNAI), vol. 4617, pp. 57–67. Springer, Heidelberg (2007)
22. Santiago, R., Bedregal, B., Acióly, B.: Formal Aspects of Correctness and Optimality in Interval Computations. Formal Asp. of Comput. 18(2), 231–243 (2006)
23. Yager, R.R.: On Some New Classes of Implication Operators and Their Role in Approximate Reasoning. Inf. Sci. 167(1–4), 193–216 (2004)

An Approach to General Quantification Using Representation by Levels*

Daniel Sánchez[1],[2], Miguel Delgado[2], and María-Amparo Vila[2]

[1] European Centre for Soft Computing
Edificio Científico-Tecnológico, Mieres, Asturias, Spain
daniel.sanchezf@softcomputing.es
[2] 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
mdelgado@ugr.es,vila@decsai.ugr.es

Abstract. In this paper we propose an extension of generalized quantification to the fuzzy case using a recently proposed level representation of fuzziness. The level representation allows the extension of crisp quantification to the fuzzy case in a simple way, keeping all its properties. The expressive power of this extension to the theory of generalized quantifiers goes far beyond the usual fuzzy quantification framework based on absolute and relative fuzzy quantifiers. The proposal offer many potentially interesting possibilities for developing applications inspired in the Computing with Words and Perceptions paradigm, remarkably linguistic summarization of data.

Keywords: Fuzzy quantification, generalized quantifiers, representation by levels.

1 Introduction

In the framework of Computing with Words and Perceptions [22], the linguistic summarization of data requires the evaluation of the accomplishment degree of linguistic expressions about data, where linguistic terms are represented as fuzzy subsets of data domains [19,11]. Linguistic expressions involving quantifiers, called quantified sentences, are among the most employed for this purpose. Zadeh's framework for quantification [21], considers expressions 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. There are many different approaches for evaluating the accomplishment of this type of sentences, see among others [21,6,18,5,4,9,20,3,13,16].

* This work has been partially supported by the Spanish Government under project TIN2009-08296, and by the Andalusian Government (Junta de Andalucía, Consejería de Innovación, Ciencia y Empresa) under project P07-TIC-03175 "Representación y Manipulación de Objetos Imperfectos en Problemas de Integración de Datos: Una Aplicación a los Almacenes de Objetos de Aprendizaje".

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 50–57, 2011.

Linguistic quantifiers in Zadeh's framework are normal, convex fuzzy subsets of $\mathbb{Z}$ (absolute quantifiers) or $[0,1]$ (relative quantifiers), that generalize the ordinary quantifiers $\exists$ and $\forall$ of first order logic [21]. However, some very common types of quantified statements in natural language cannot be modeled using this kind of quantifiers. In order to fill this gap, Glöckner [7] proposed an approach to generalize fuzzy quantification to any quantifier in the *Theory of Generalized Quantifiers* (TGQ) [12]. Generalized quantifiers include both absolute and relative quantifiers as particular cases, providing a way to represent many other quantifiers with richer semantics and expressive power. A number of important proposals along this line can be found in the literature [7,1,2,8,4,9,10].

In this paper we propose an alternative way for extending the TGQ to the fuzzy case along this line. Our approach is based on the representation of fuzziness by means of levels proposed in [15,17]. This representation fits perfectly the requirements of this problem, since concepts and schemes can be extended to the fuzzy case in a direct and straightforward way, keeping all the properties of the crisp case. We already showed the suitability of this model in modeling quantified sentences in Zadeh's framework [16].

2 Extending Generalized Quantifiers to the Fuzzy Case

In this section we briefly introduce the notion of generalized quantifier and Glöckner's approach for extending it to the fuzzy case.

2.1 Generalized Quantifiers

In TGQ, a generalized quantifier is a second-order predicate whose arguments are taken from the power set of a reference set X. For instance, the usual quantifiers $\exists$ and $\forall$ can be defined by two binary predicates $P_\exists : \mathcal{P}(X) \times \mathcal{P}(X) \to \{0,1\}$ and $P_\forall : \mathcal{P}(X) \times \mathcal{P}(X) \to \{0,1\}$ as follows:

$$P_\exists(X_1, X_2) = 1 \text{ iff } X_1 \cap X_2 \neq \emptyset \tag{1}$$

$$P_\forall(X_1, X_2) = 1 \text{ iff } X_1 \subseteq X_2 \tag{2}$$

Eq. (1) corresponds to the quantified sentence *At least one* X_1 *are* X_2 whilst Eq. (2) corresponds to *All* X_1 *are* X_2. Another example is the quantifier *at least 4* and *less than 80%*, that can be defined respectively as:

$$P_{\geq 4}(X_1, X_2) = 1 \text{ iff } |X_1 \cap X_2| \geq 4 \tag{3}$$

$$P_{<80\%}(X_1, X_2) = 1 \text{ iff } \frac{|X_1 \cap X_2|}{|X_1|} < 0.8 \tag{4}$$

However, generalized quantifiers are much more general. For example, quantifiers of arity greater than 2 are possible, like for instance *More* X_1 *than* X_2 *are* X_3. This is a ternary quantifier P that can be defined as:

$$P_{more}(X_1, X_2, X_3) = 1 \text{ iff } |X_1 \cap X_3| > |X_2 \cap X_3| \tag{5}$$

2.2 Fuzzy Extension: Semi-fuzzy Quantifiers

Generalized quantifiers as introduced in the previous section are crisp predicates. However, in practice it is necessary to evaluate the accomplishment degree of statements involving fuzzy sets. For the classical quantifiers and following Eqs. (1) and (2), it is enough to consider fuzzy degrees of intersection and inclusion of fuzzy sets, respectively, a topic that has been extensively studied in the literature. However, the situation is not so simple with other quantifiers.

As a first step towards generalizing quantifiers to the fuzzy case, Glöckner introduces the notion of *semi-fuzzy quantifier* [7]. A semi-fuzzy quantifier is a *fuzzy* second-order predicate that takes as arguments *crisp* subsets of X. Hence, a binary semi-fuzzy quantifier has the form $P : \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow [0,1]$. Absolute and relative fuzzy quantifiers are particular cases of quantifiers that can be represented as semi-fuzzy quantifiers. For instance, given an absolute fuzzy quantifier Q, it can be represented by the semi-fuzzy quantifier P_Q defined as $P_Q(X_1, X_2) = Q(|X_1 \cap X_2|)$. Similarly, the relative fuzzy quantifier Q' can be represented by $P_{Q'}$ defined as $P_{Q'}(X_1, X_2) = Q'(|X_1 \cap X_2|/|X_1|)$.

In order to be able to evaluate quantified statements involving fuzzy subsets of X, it is necessary to provide what Glöckner calls *quantifier fuzzification mechanisms* (QFMs). Different QFMs have been proposed in [7,1,2,8,4,9,10].

3 Representation by Levels

The representation by levels (called restriction-level representation in previous works) of a fuzzy concept is an assignment from $[0,1]$ to $\mathcal{P}(X)$. As a particular case, a fuzzy set is a representation by levels following the usual representation as a collection of nested alpha-cuts. However, there are two main differences between representation by levels and fuzzy set theory:

- Contrary to the alpha-cut representation, the sets are not necessarily nested.
- All the operations, definitions, and properties are applied in each level independently.

3.1 Representation

In [15,17] we assume that for each concept there is a finite set of relevant levels $\Lambda = \{\alpha_1, \ldots, \alpha_m\}$ verifying that $1 = \alpha_1 > \alpha_2 > \cdots > \alpha_m > \alpha_{m+1} = 0$, $m \geq 1$. 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.

Definition 1. *A level representation is a pair (Λ, ρ) where Λ is a set of levels and ρ is a function*

$$\rho : \Lambda \rightarrow \mathcal{P}(X) \tag{6}$$

The function ρ indicates the crisp realization that represents the fuzzy concept in each level. Given a fuzzy concept F represented by (Λ_F, ρ_F), we define the set of crisp representatives of F, Ω_F, as

$$\Omega_F = \{\rho_F(\alpha) \mid \alpha \in \Lambda_F\} \tag{7}$$

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

$$\rho(\alpha) = \rho(\alpha_i) \tag{8}$$

Obviously, (Λ, ρ) on X is a crisp set $A \subseteq X$ iff $\forall \alpha \in \Lambda$, $\rho(\alpha) = A$.

3.2 Random-Set View

Given (Λ_A, ρ_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

$$Pos(\rho_A(\alpha_i)) = \alpha_i. \tag{9}$$

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

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

$$m(Y) = \sum_{\alpha_i \mid Y = \rho(\alpha_i)} \alpha_i - \alpha_{i+1}. \tag{10}$$

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 level representation can be seen as a random set *plus a structure indicating dependencies between the possible representations of different properties.*

3.3 Advantages

The main advantages of this representation are:

- The extension of crisp definitions and operations to the fuzzy case is straightforward. In a sense, a fuzzy concept is transformed into a collection of crisp concepts that are operated independently (levelwise) using the available crisp operations.
- All the properties of the crisp case are kept.

4 Our Approach to Fuzzy Generalized Quantification

The aforementioned advantages of the level representation make it highly suitable for the purposes of providing a QFM. Since semi-fuzzy quantifiers operate on crisp sets, they can be applied to each level in the representation of fuzzy concepts in order to obtain a level representation of the accomplishment degree, as follows:

4.1 Levelwise Evaluation

Definition 3. *Let P be a n-ary semi-fuzzy quantifier on X. Let $A_1, \ldots, A_n$ be fuzzy concepts represented by levels as $(\Lambda_{A_1}, \rho_{A_1}), \ldots, (\Lambda_{A_n}, \rho_{A_n})$. The evaluation of $E \equiv P(A_1, \ldots, A_n)$ is a level representation (Λ_E, ρ_E), where*

$$\Lambda_E = \bigcup_{i=1}^{n} \Lambda_{A_i} \tag{11}$$

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

$$\rho_E(\alpha) = P(\rho_{A_1}(\alpha), \ldots, \rho_{A_n}(\alpha)) \tag{12}$$

As an example, consider the atomic concepts A_1, A_2, A_3 defined by the following fuzzy sets:

$$A_1 = 1/x_1 + 0.8/x_2 + 0.5/x_3 + 0.4/x_5$$

$$A_2 = 0.9/x_1 + 0.6/x_3 + 0.5/x_4$$

$$A_3 = 0.4/x_1 + 1/x_3 + 1/x_4$$

Then we have $\Lambda_{A_1} = \Lambda_{\neg A_1} = \{1, 0.8, 0.5, 0.4\}$, $\Lambda_{A_2} = \Lambda_{\neg A_2} = \{1, 0.9, 0.6, 0.5\}$, and $\Lambda_{A_3} = \Lambda_{\neg A_3} = \{1, 0.4\}$. In Table 1 we can see several properties derived from these 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 level representation in general). Notice that since level operations 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 concepts derived from the A_i, using the ternary generalized quantifier P_{more} defined in Eq. (5) and the

Table 1. Level representation of several concepts

α	$\rho_{A_1}(\alpha)$	$\rho_{\neg A_1}(\alpha)$	$\rho_{A_2}(\alpha)$	$\rho_{\neg A_2}(\alpha)$	$\rho_{A_1 \wedge \neg A_2}(\alpha)$	$\rho_{A_3}(\alpha)$
1	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\emptyset$	X	$\{x_1\}$	$\{x_3, x_4\}$
0.9	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\emptyset$	$\{x_3, x_4\}$
0.8	$\{x_1, x_2\}$	$\{x_3, x_4, x_5\}$	$\{x_1\}$	$\{x_2, x_3, x_4, x_5\}$	$\{x_2\}$	$\{x_3, x_4\}$
0.6	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_1, x_3\}$	$\{x_2, x_4, x_5\}$	$\{x_2\}$	$\{x_3, x_4\}$
0.5	$\{x_1, x_2, x_3\}$	$\{x_4, x_5\}$	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$	$\{x_2\}$	$\{x_3, x_4\}$
0.4	$\{x_1, x_2, x_3, x_5\}$	$\{x_4\}$	$\{x_1, x_3, x_4\}$	$\{x_2, x_5\}$	$\{x_2, x_5\}$	$\{x_1, x_3, x_4\}$

Table 2. Some quantified sentences involving A_i

E_1	More A_1 than A_2 are A_3	E_3	Q_M of A_2 are A_3	E_5	$\exists$ of A are $\neg$ A
E_2	More A_2 than A_3 are A_1	E_4	Q_M of $\neg A_1$ are $\neg A_2$	E_6	$\forall$ of $(A_1 \wedge \neg A_2)$ are $\neg A_3$

Table 3. Evaluation of sentences in Table 2

α	E_1	E_2	E_3	E_4	E_5	E_6
1	0	0	0	1	0	1
0.9	0	1	0	1	0	0
0.8	0	1	0	1	0	1
0.6	0	1	1/2	1	0	1
0.5	0	1	2/3	1/2	0	1
0.4	0	0	1	0	0	1

quantifiers $Q_M(x) = x$, $\exists$, and $\forall$. Table 3 shows the corresponding evaluation of the sentences in table 2.

All the properties of evaluation with semi-fuzzy quantifiers are kept since we are calculating levelwise. This is important since a collection of properties that any suitable method should satisfy have been proposed [7,5,9], but it is not easy to show that a certain QFM satisfy them. The existing QFM are in general sound but sophisticated proposals in which a hard work has been devoted to study the fulfilment of these properties.

Another problem of some existing QFM that is solved with our approach is that computational complexity is quadratic or even more. In our case, 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 the A_i being the alpha-cut representation of fuzzy sets, the complexity is $O(nlogn)$, as shown in [5].

Our proposal has as a particular case the evaluation method proposed in [16], in which absolute and relative quantifiers only are considered. In the same work we showed that properties involving the negation are verified by our scheme. 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 sentences of the kind $\exists\ A$ are $\neg A$ is definitely 0. This goes beyond the development of QFMs, as it is related to the very nature of the fuzziness representation. This property comes from the fact that in level representation, the representation of $A \wedge \neg A = \emptyset$.

4.2 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 representation by levels, we would proceed operating in each level independently. In [16] we proposed to summarize the information as follows:

Definition 4. *The numerical summary $S(E)$ of an evaluation (Λ_E, ρ_E) is given by*

$$S(E) = \sum_{\beta \in \Omega_E} m_E(\beta) \cdot \beta \tag{13}$$

Table 4 shows the evaluation of E_1 - E_6 following definition 4.

Table 4. Numerical evaluation of sentences in Table 2

$S(E_1)$	$S(E_2)$	$S(E_3)$	$S(E_4)$	$S(E_5)$	$S(E_6)$
0	0.5	7/60	0.35	0	0.9

Let us remark that since this is a summary of the evaluation by levels, again the required properties for the evaluation are kept. In particular, as the intuition suggests, the final evaluation of sentence E_5 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 level representations, we obtain a coherent result whilst representing properly the fuzziness of the properties.

In the particular case when A,D are fuzzy sets, we have again that the evaluation of $E =$ "Q of D are A" is $S(E) = GD_Q(D/A)$, where $GD_Q(D/A)$ is the evaluation of E as given by the method GD proposed in [5]. This is the case in sentences E_3 to E_6.

5 Conclusions

The representation by levels and the corresponding operations are, by their properties, specially well suited for the extension to the fuzzy case of generalized quantification. As future work, we shall study the relation of our approach to those proposed in [7,1,2,8,4,9,10]. Furthermore, the representation by levels offer a direct way to extend syllogisms and reasoning with generalized quantifiers to the fuzzy case with many potential applications [14], that we shall explore. In addition, we will apply these results in linguistic summarization of data, data mining, and quantification in fuzzy description logics.

References

1. Barro, S., Bugarín, A., Cariñena, P., Díaz-Hermida, F.: A framework for fuzzy quantification models analysis. IEEE Transactions on Fuzzy Systems 11(1), 89–99 (2003)
2. 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)
3. Cui, L., Li, Y.: Linguistic quantifiers based on choquet integrals. Int. Journal of Approximate Reasoning 48, 559–582 (2008)

4. Díaz-Hermida, F., Losada, D.E., Bugarín, A., Barro, S.: A probabilistic quantifier fuzzification mechanism: The model and its evaluation for information retrieval. IEEE Transactions on Fuzzy Systems 13(5), 688–700 (2005)
5. Delgado, M., Sánchez, D., Vila, M.A.: Fuzzy cardinality based evaluation of quantified sentences. International Journal of Approximate Reasoning 23, 23–66 (2000)
6. Dubois, D., Prade, H.: Fuzzy cardinality and the modeling of imprecise quantification. Fuzzy Sets and Systems 16, 199–230 (1985)
7. Glöckner, I.: DFS - an axiomatic approach to fuzzy quantification. Technical Report TR97-06, Technical Faculty, University Bielefeld, 33501 Bielefeld, Germany (1997)
8. Glöckner, I.: Evaluation of quantified propositions in generalized models of fuzzy quantification. Int. Journal of Approximate Reasoning 37, 93–126 (2004)
9. Glöckner, I.: Fuzzy Quantifiers: A Computational Theory. Springer, Heidelberg (2006)
10. Glöckner, I.: Generalized fuzzy quantifiers and the modeling of fuzzy branching quantification. Int. Journal of Intelligent Systems 24, 624–648 (2009)
11. Kacprzyk, J., Zadrozny, S.: Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. Information Sciences 173, 281–304 (2005)
12. Keenan, E.L., Westerståhl, D.: Generalized quantifiers in linguistics and logic. In: Benthem, J.V., Meulen, A.T. (eds.) Handbook of Logic and Language. ch.15, pp. 837–893. Elsevier, Amsterdam (1997)
13. 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)
14. Liu, Y., Kerre, E.: An overview of fuzzy quantifiers part I (interpretations) and II (reasoning and applications). Fuzzy Sets and Systems 95, 1–21, 135–146 (1998)
15. Sánchez, D., Delgado, M., Vila, M.A.: A restriction level approach to the representation of imprecise properties. In: IPMU 2008, pp. 153–159 (2008)
16. Sánchez, D., Delgado, M., Vila, M.A.: Fuzzy quantification using restriction levels. In: Di Gesù, V., Pal, S.K., Petrosino, A. (eds.) WILF 2009. LNCS, vol. 5571, pp. 28–35. Springer, Heidelberg (2009)
17. Sánchez, D., Delgado, M., Vila, M.A., Chamorro-Martínez, J.: On a non-nested level-based representation of fuzziness. Fuzzy Sets and Systems (In press)
18. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on System, Man and Cybernetics 18(1), 183–190 (1988)
19. Yager, R.R.: On linguistic summaries of data. In: Piatetsky-Shapiro, G., Frawley, W.J. (eds.) Knowledge Discovery in Databases, pp. 347–363. AAAI/MIT Press (1991)
20. Ying, M.: Linguistic quantifiers modeled by sugeno integrals. Artificial Intelligence 170, 581–606 (2006)
21. Zadeh, L.A.: A computational approach to fuzzy quantifiers in natural languages. Computing and Mathematics with Applications 9(1), 149–184 (1983)
22. Zadeh, L.A.: A new direction in AI. Toward a computational theory of perceptions. AI Magazine 22(1), 73–84 (2001)

Towards Learning Fuzzy DL Inclusion Axioms

Francesca A. Lisi[1] and Umberto Straccia[2]

[1] Dipartimento di Informatica, Università degli Studi di Bari "Aldo Moro", Italy
lisi@di.uniba.it
[2] ISTI - CNR, Pisa, Italy
straccia@isti.cnr.it

Abstract. Fuzzy Description Logics (DLs) are logics that allow to deal with vague structured knowledge. Although a relatively important amount of work has been carried out in the last years concerning the use of fuzzy DLs as ontology languages, the problem of automatically managing fuzzy ontologies has received no attention so far. We report here our preliminary investigation on this issue by describing a method for inducing inclusion axioms in a fuzzy DL-Lite like DL.

1 Introduction

Description Logics (DLs) [1] play a key role in the design of *ontologies*. An ontology consists of a hierarchical description of important concepts in a particular domain, along with the description of the properties (of the instances) of each concept. In this context, DLs are important as they are essentially the theoretical counterpart of the *Web Ontology Language OWL 2*[1], the current standard language to represent ontologies, and its profiles.[2] E.g., DL-Lite [2] is the DL behind the *OWL 2 QL* profile.

It is well-known that "classical" ontology languages are not appropriate to deal with *vague knowledge*, which is inherent to several real world domains [13]. So far, several fuzzy extensions of DLs can be found in the literature (see the survey in [8]). Although a relatively important amount of work has been carried out in the last years concerning the use of fuzzy DLs as ontology languages, the problem of automatically managing fuzzy ontologies has received no attention so far. In this work, we report our preliminary investigation on this issue by describing a method for inducing inclusion axioms in a fuzzy DL-Lite like DL. The method follows the machine learning approach known as *Inductive Logic Programming* (ILP) by adapting known results in ILP concerning crisp rules to the novel case of fuzzy DL inclusion axioms.

The paper is structured as follows. Section 2 is devoted to preliminaries on ILP and fuzzy DLs. Section 3 describes our preliminary contribution to the problem in hand, also by means of an illustrative example. Section 4 concludes the paper with final remarks and comparison with related work.

[1] http://www.w3.org/TR/2009/REC-owl2-overview-20091027/
[2] http://www.w3.org/TR/owl2-profiles/

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 58–66, 2011.

2 Preliminaries

2.1 Learning Rules with ILP

Inductive Logic Programming (ILP) was born at the intersection between Concept Learning and Logic Programming [9].

From Logic Programming it has borrowed the Knowledge Representation (KR) framework, i.e. the possibility of expressing facts and rules in the form of Horn clauses. In the following, rules are denoted by $A(\mathbf{x}) \rightarrow H(\mathbf{x})$ where $\mathbf{x}$ is the vector of the n variables that appear in the rule, $A(\mathbf{x}) = A_0(\mathbf{x}) \wedge \ldots \wedge A_q(\mathbf{x})$ represents the antecedent (called the *body*) of the rule, and $H(\mathbf{x})$ is the consequent (called *head*) of the rule. The predicate H pertains to the concept to be learnt (called *target*). Given an attribute domain $\mathcal{D}$ and a vector $\mathbf{t} \in \mathcal{D}_n$ of n values of the domain, we denote the ground substitution of the variable $\mathbf{x}$ with $\mathbf{t}$ by $H(\mathbf{t}) = \sigma[\mathbf{x}/\mathbf{t}]H(\mathbf{x})$. Then $H(\mathbf{t})$ is true or false in a given interpretation.

From Concept Learning it has inherited the inferential mechanisms for induction, the most prominent of which is *generalisation*. A distinguishing feature of ILP is the use of prior domain knowledge during the induction process. The classical ILP problem is described by means of two logic programs: (i) the *background theory* $\mathcal{B}$ which is a set of ground facts and rules; (ii) the *training set* $\mathcal{E}$ which is a set of ground facts, called *examples*, pertaining to the target predicate. It is often split in $\mathcal{E}^+$ and $\mathcal{E}^-$, which correspond respectively to positive and negative examples. If only $\mathcal{E}^+$ is given, $\mathcal{E}^-$ can be deduced by using the Closed World Assumption (CWA)[3]. The task of induction is to find, given $\mathcal{B}$ and $\mathcal{E}$, a set of rules $\mathcal{H}$ such that: (i) $\forall e \in \mathcal{E}^+, \mathcal{B} \cup \mathcal{H} \models e$ and (ii) $\forall e \in \mathcal{E}^-, \mathcal{B} \cup \mathcal{H} \not\models e$. Two further restrictions hold naturally. One is that $\mathcal{B} \not\models \mathcal{E}^+$ since, in such a case, $\mathcal{H}$ would not be necessary to explain $\mathcal{E}^+$. The other is $\mathcal{B} \cup \mathcal{H} \not\models \bot$, which means that $\mathcal{B} \cup \mathcal{H}$ is a consistent theory. Usually, rule induction fits with the idea of providing a compression of the information contained in $\mathcal{E}$. A rule r *covers* an example $e \in \mathcal{E}$ wrt $\mathcal{B}$ iff $\mathcal{B} \cup \{r\} \models e$.

A popular ILP algorithm for learning sets of rules from relational data is FOIL [10]. It performs a greedy search in order to maximise a gain function. The rules are induced until all positive examples are covered or no more rules are found that overcome the threshold. When a rule is induced, the positive examples covered by the rule are removed from $\mathcal{E}$. The function FOIL-Learn-One-Rule reported in Figure 1 starts with the most general clause $(\top \rightarrow H(\mathbf{x}))$ and specialises it step by step by adding literals in the antecedent. The rule r is accepted when its confidence degree $cf(r)$ (see later on) overcomes a fixed threshold θ and it does not cover any negative example. The Gain function is computed by the formula:

$$\textsc{Gain}(cf(r_1), cf(r_2)) = p * (log_2(cf(r_1)) - log_2(cf(r_2))) \ , \tag{1}$$

where p is the number of distinct examples covered by the rule r_1, *i.e.* $p = |\{e \mid \mathcal{B} \cup \{r_1\} \models e\}|$. Thus, the gain is positive iff the new rule is more informative

[3] Anything that is not stated as true is assumed to be false.

function FOIL-Learn-One-Rule(h, $\mathcal{E}^+$, $\mathcal{E}^-$, $\mathcal{B}$): r
begin
1. $A(\mathbf{x}) \leftarrow \top$;
2. $r \leftarrow \{A(\mathbf{x}) \rightarrow H(\mathbf{x})\}$;
3. $\mathcal{E}^-_r \leftarrow \mathcal{E}^-$;
4. **while** $cf(r) < \theta$ **and** $\mathcal{E}^-_r \neq \emptyset$ **do**
5. $\quad A_{best}(\mathbf{x}) \leftarrow A(\mathbf{x})$;
6. $\quad maxgain \leftarrow 0$;
7. $\quad$ **foreach** $l \in \mathcal{B}$ **do**
8. $\quad\quad gain \leftarrow$ Gain($cf(A(\mathbf{x}) \wedge l(\mathbf{x}) \rightarrow H(\mathbf{x})), cf(A(\mathbf{x}) \rightarrow H(\mathbf{x}))$;
9. $\quad\quad$ **if** $gain \geqslant maxgain$ **then**
10. $\quad\quad\quad maxgain \leftarrow gain$;
11. $\quad\quad\quad A_{best}(\mathbf{x}) \leftarrow A(\mathbf{x}) \wedge l(\mathbf{x})$;
12. $\quad\quad$ **endif**
13. $\quad$ **endforeach**
14. $\quad r \leftarrow \{A_{best}(\mathbf{x}) \rightarrow H(\mathbf{x})\}$;
15. $\quad \mathcal{E}^-_r \leftarrow \mathcal{E}^-_r \setminus \{e \in \mathcal{E}^- | \mathcal{B} \cup r \models e\}$;
16. **endwhile**
17. **return** r
end

Fig. 1. Algorithm for learning one rule in FOIL

in the sense of Shannon's information theory (i.e. iff the confidence degree increases). If there are some literals to add which increase the confidence degree, the gain tends to favor the literals that offer the best compromise between the confidence degree and the number of examples covered.

Given a Horn clause $A(\mathbf{x}) \rightarrow H(\mathbf{x})$, its confidence degree is given by: $cf(A(\mathbf{x}) \rightarrow H(\mathbf{x})) = P(A(\mathbf{x}) \wedge H(\mathbf{x}))/P(A(\mathbf{x}))$. Confidence degrees are computed in the spirit of domain probabilities. Input data in ILP problems are supposed to describe one interpretation under CWA. We call $\mathcal{I}_{ILP}$ this interpretation. So, given a fact f, we define

$$\mathcal{I}_{ILP} \models f \text{ iff } \mathcal{B} \cup \mathcal{E} \models f \ . \tag{2}$$

The domain $\mathcal{D}$ is the Herbrand domain described by $\mathcal{B}$ and $\mathcal{E}$. We take P as a uniform probability on $\mathcal{D}$. So the confidence degree in a clause $A(\mathbf{x}) \rightarrow H(\mathbf{x})$ is:

$$cf(A(\mathbf{x}) \rightarrow H(\mathbf{x})) = \frac{|\mathbf{t} \in \mathcal{D}^n \mid \mathcal{I}_{ILP} \models (A(\mathbf{t}) \wedge H(\mathbf{t})) \text{ and } H(\mathbf{t}) \in \mathcal{E}^+|}{|\mathbf{t} \in \mathcal{D}^n \mid \mathcal{I}_{ILP} \models A(\mathbf{t}) \text{ and } H(\mathbf{t}) \in \mathcal{E}|} \tag{3}$$

where $|\cdot|$ denotes set cardinality. Testing all possible $\mathbf{t} \in \mathcal{D}^n$ is not tractable in practice. However, we can equivalently restrict the computation to the substitutions that map variables to constants in their specific domains. In fact, this computation is equivalent to a database query and thus, we can also use some optimization strategy such as indexing or query ordering. This makes the computation tractable although it remains costly.

2.2 A DL-Lite Like Description Logic and Its Fuzzy Extensions

The logic we adopt is based on a fuzzy extension of the DL-Lite [2] DL without negation, and is supported by the SoftFacts system [15][4]. DL-Lite supports unary predicates (called *concepts*) and binary predicates (called *roles*).

[4] See, http://www.straccia.info/software/SoftFacts/SoftFacts.html

A *knowledge base* $\mathcal{K} = \langle \mathcal{F}, \mathcal{O}, \mathcal{A} \rangle$ consists of a *facts component* $\mathcal{F}$, an *ontology component* $\mathcal{O}$ and an *abstraction component* $\mathcal{A}$, which are defined as follows (for a detailed account of the semantics, see [14]).

Facts Component. $\mathcal{F}$ is a finite set of expressions of the form

$$R(c_1, \ldots, c_n)[s] \ , \tag{4}$$

where R is an n-ary relation, every c_i is a constant, and s is a degree of truth (or *score*) in $[0, 1]$ indicating to which extent the tuple $\langle c_1, \ldots, c_n \rangle$ is an instance of relation R.[5] Facts are stored in a relational database. We may omit the score component and in such case the value 1 is assumed.

Ontology Component. The ontology component is used to define the relevant abstract concepts and relations of the application domain by means of inclusion axioms. Specifically, $\mathcal{O}$ is a finite set of *inclusion axioms* having the form

$$Rl_1 \sqcap \ldots \sqcap Rl_m \sqsubseteq Rr \ , \tag{5}$$

where $m \geqslant 1$, all Rl_i and Rr have the same arity and each Rl_i is a so-called *left hand relation* and Rr is a *right hand relation*. We assume that relations occurring in $\mathcal{F}$ do not occur in inclusion axioms (so, we do not allow that database relation names occur in $\mathcal{O}$). The intuitive semantics is that if a tuple $\mathbf{c}$ is instance of each relation Rl_i to degree s_i then $\mathbf{c}$ is instance of Rr to degree $\min(s_1, \ldots, s_m)$.

The exact syntax of the relations appearing on the left-hand and right hand side of inclusion axioms is specified below:

$$\begin{array}{l} Rl \longrightarrow A \mid R[i_1, i_2] \\ Rr \longrightarrow A \mid R[i_1, i_2] \mid \exists R.A \end{array} \tag{6}$$

where A is an *atomic concept* and R is a role with $1 \leqslant i_1, i_2 \leqslant 2$. Here $R[i_1, i_2]$ is the projection of the relation R on the columns i_1, i_2 (the order of the indexes matters). Hence, $R[i_1, i_2]$ has arity 2. Additionally, $\exists R.A$ is a so-called qualified existential quantification on roles which corresponds to the FOL formula $\exists y.R(x, y) \wedge A(y)$ where $\wedge$ is interpreted as the min t-norm.

Abstraction Component. $\mathcal{A}$ is a set of statements that allow to connect atomic concepts and relations to physical relational tables. Essentially, it is used as a wrapper to the underlying database and, thus, prevents that relational table names occur in the ontology. Formally, an *abstraction statement* is of the form

$$R \mapsto (c_1, \ldots, c_n)[c_{score}].sql \ , \tag{7}$$

where sql is a SQL statement returning n-ary tuples $\langle c_1, \ldots, c_n \rangle$ ($n \leqslant 2$) with score determined by the c_{score} column. The tuples have to be ranked in decreasing order of score and, as for the fact component, we assume that there cannot be two records $\langle \mathbf{c}, s_1 \rangle$ and $\langle \mathbf{c}, s_2 \rangle$ in the result set of sql with $s_1 \neq s_2$ (if there are, then we remove the one with the lower score). The score c_{score} may be omitted

[5] The score s may have been computed by some external tool, such as a classifier, etc.

and in that case the score 1 is assumed for the tuples. We assume that R occurs in $\mathcal{O}$, while all of the relational tables occurring in the SQL statement occur in $\mathcal{F}$. Finally, we assume that there is at most one abstraction statement for each abstract relational symbol R.

Query Language. A *query* consists of a "conjunctive query", with a scoring function to rank the answers. A *ranking query* [7] is of the form

$$q(\mathbf{x})[s] \leftarrow \exists \mathbf{y}\ R_1(\mathbf{z}_1)[s_1], \ldots, R_l(\mathbf{z}_l)[s_l], \\ \mathsf{OrderBy}(s = f(s_1, \ldots, s_l, p_1(\mathbf{z}'_1), \ldots, p_h(\mathbf{z}'_h)) \tag{8}$$

where

1. q is an n-ary relation, every R_i is a binary relation. $R_i(\mathbf{z_i})$ may also be of the form $(z \leqslant v), (z < v), (z \geqslant v), (z > v), (z = v), (z \neq v)$, where z is a variable, v is a value of the appropriate concrete domain;
2. $\mathbf{x}$ are the *distinguished variables*.
3. $\mathbf{y}$ are existentially quantified variables called the *non-distinguished variables*. We omit to write $\exists \mathbf{y}$ when $\mathbf{y}$ is clear from the context;
4. $\mathbf{z_i}, \mathbf{z'_j}$ are tuples of constants or variables in $\mathbf{x}$ or $\mathbf{y}$;
5. $s, s_1, \ldots, s_l$ are distinct variables and different from those in $\mathbf{x}$ and $\mathbf{y}$;
6. p_j is an n_j-ary *fuzzy predicate* assigning to each n_j-ary tuple $\mathbf{c}_j$ a *score* $p_j(\mathbf{c}_j) \in [0, 1]$. We require that an n-ary fuzzy predicate p is *safe*, that is, there is not an m-ary fuzzy predicate p' such that $m < n$ and $p = p'$. Informally, all parameters are needed in the definition of p.
7. f is a *scoring* function $f\colon ([0, 1])^{l+h} \rightarrow [0, 1]$, which combines the scores of the l relations $R_i(\mathbf{c}'_i)$ and the n fuzzy predicates $p_j(\mathbf{c}''_j)$ into an overall *score* to be assigned to the rule head $R(\mathbf{c})$. We assume that f is *monotone*, that is, for each $\mathbf{v}, \mathbf{v}' \in ([0, 1])^{l+h}$ such that $\mathbf{v} \leqslant \mathbf{v}'$, it holds $f(\mathbf{v}) \leqslant f(\mathbf{v}')$, where $(v_1, \ldots, v_{l+h}) \leqslant (v'_1, \ldots, v'_{l+h})$ iff $v_i \leqslant v'_i$ for all i. We also assume that the computational cost of f and all fuzzy predicates p_i is bounded by a constant.

We call $q(\mathbf{x})[s]$ its *head*, $\exists \mathbf{y}.R_1(\mathbf{z}_1)[s_1], \ldots, R_l(\mathbf{z}_l)[s_l]$ its *body* and $\mathsf{OrderBy}(s = f(s_1, \ldots, s_l, p_1(\mathbf{z}'_1), \ldots, p_h(\mathbf{z}'_h))$ the *scoring atom*. We also allow the scores $[s]$, $[s_1]$, $\ldots$, $[s_l]$ and the scoring atom to be omitted. In this case we assume the value 1 for s_i and s instead. The informal meaning of such a query is: if $\mathbf{z}_i$ is an instance of R_i to degree at least or equal to s_i, then $\mathbf{x}$ is an instance of q to degree at least or equal to s, where s has been determined by the scoring atom. The *answer set* $ans_{\mathcal{K}}(q)$ over $\mathcal{K}$ of a query q is the set of tuples $\langle \mathbf{t}, s \rangle$ such that $\mathcal{K} \models q(\mathbf{t})[s]$ with $s > 0$ (informally, $\mathbf{t}$ satisfies the query to non-zero degree s).

3 Towards Fuzzy DL-Lite Like Inclusion Axioms Learning

We now show how we may learn DL-Lite like inclusion axioms. We consider a learning problem where:

- the background theory $\mathcal{B}$ is a DL-Lite like knowledge base $\langle \mathcal{F}, \mathcal{O}, \mathcal{A} \rangle$;
- the training set $\mathcal{E}$ is a collection of fuzzy DL-Lite like facts of the form Eq. 4 and labeled as either positive or negative examples for the target atomic concept H. We assume that $\mathcal{F} \cap \mathcal{E} = \emptyset$;

– the target theory $\mathcal{H}$ is a set of inclusion axioms of the form

$$C_1 \sqcap \ldots \sqcap C_n \sqsubseteq H \tag{9}$$

where H is an atomic concept and each C_i has syntax

$$C \longrightarrow A \mid \exists R.A \mid \exists R.\top \ . \tag{10}$$

Now, we adapt Eq. 2 to our case and define

$$\mathcal{I}_{ILP} \models C(t) \text{ iff } \mathcal{B} \cup \mathcal{E} \models C(t)[s] \text{ and } s > 0 \ . \tag{11}$$

That is, we write $\mathcal{I}_{ILP} \models C(t)$ iff t can be inferred being instance of concept C to a non-zero degree.

Now, in order to account for multiple fuzzy instantiations of fuzzy predicates occurring in inclusion axioms, we customise Eq. 3 into the following formula for computing the confidence degree:

$$cf(RL \sqsubseteq H) = \frac{\sum_{t \in P} RL(t) \Rightarrow H(t)}{|D|} \ , \text{ where} \tag{12}$$

– $P = \{t \mid \mathcal{I}_{ILP} \models C_i(t), \mathcal{I}_{ILP} \models H(t) \text{ and } H(t)[s] \in \mathcal{E}^+\}$;
– $D = \{t \mid \mathcal{I}_{ILP} \models C_i(t) \text{ and } H(t)[s] \in \mathcal{E}\}$;
– $RL = C_1 \sqcap \ldots \sqcap C_n$;
– $\Rightarrow$ is an implication function (see, *e.g.* [6]);
– $RL(t) = \min(s_1, \ldots s_n)$, with $\mathcal{B} \cup \mathcal{E} \models C_i(t)[s_i]$;
– $H(t) = s$ with $H(t)[s] \in \mathcal{E}$.

Essentially, the numerator sums over all positive instances making the left hand and right hand side true, where $RL(t) \Rightarrow H(t)$ denotes the degree to which the implication holds. Clearly, the more positive instances supporting the inclusion axiom there are, the higher is the confidence degree of it. Note that the confidence score can be determined easily by submitting appropriate queries via our query language. From an algorithm point of view, it suffices to change the FOIL-Learn-One-Rule at step 7., where now l may be either an atomic concept A, or of the form $\exists R.A$ or $\exists R.\top$.

We illustrate our proposal via an example. For illustrative purposes consider the following case involving hotels. So, assume we have a background theory $\mathcal{B}$ with a relational database $\mathcal{F}$ storing facts such as

HotelTable

id	rank	noRooms	hasRoomID
h1	3	21	r1
h2	5	123	r2
h3	4	95	r3

RoomTable

id	price	roomType
r1	60	single
r1	90	double
r2	80	single
r2	120	double
r3	70	single
r3	90	double

Tower

id
t1

Park

id
p1
p2

DistanceTable

id	from	to	time
d1	h1	t1	10
d2	h2	p1	15
d3	h3	p2	5

an ontology $\mathcal{O}$[6] encompassing the following inclusion axioms

$$Park \sqsubseteq Attraction \ , \ Tower \sqsubseteq Attraction \ , \ Attraction \sqsubseteq Site$$

[6] http://donghee.info/research/SHSS/ObjectiveConceptsOntology(OCO).html

Fig. 2. (a) Trapezoidal function $trz(x; a, b, c, d)$, (b) triangular function $tri(x; a, b, c)$, (c) left shoulder function $ls(x; a, b)$, and (d) right shoulder function $rs(x; a, b)$

and abstraction statements:

$Hotel \mapsto (h.id)$. SELECT h.id
FROM HotelTable h

$hasRank \mapsto (h.id, h.rank)$. SELECT h.id, h.rank
FROM HotelTable h

$cheapPrice \mapsto (h.id, r.price)[score]$. SELECT h.id, r.price, $cheap$(r.price) AS score
FROM HotelTable h, RoomTable r
WHERE h.hasRoomID = r.id
ORDER BY score

$closeTo \mapsto (from, to)[score]$. SELECT d.from, d.to $closedistance$(d.time) AS score
FROM DistanceTable d
ORDER BY score

where $cheap(p)$ is a function determining how cheap a hotel room is given its price, modelled as *e.g.* a so-called left-shoulder function $cheap(p) = ls(p; 50, 100)$, while $closedistance(d) = ls(d; 5, 25)$. (see Figure 2 for typical fuzzy membership functions). Assume now that our target concept H is $GoodHotel$. As illustrative example, we compute the confidence score, according to Eq. 12, of

$$r : Hotel \sqcap \exists cheapPrice.\top \sqcap \exists closeTo.Attraction \sqsubseteq GoodHotel$$

i.e., a good hotel is one having a cheap price and close to an attraction. We assume that

- the implication function is Gödel (*i.e.*, it returns 1 if $x \leqslant y$, y if $x > y$);
- $\mathcal{E}^+ = \{GoodHotel^+(h1)[0.6], GoodHotel^+(h2)[0.8]\}$, while $\mathcal{E}^- = \{GoodHotel^-(h3)[0.4]\}$;
- $GoodHotel^+ \sqsubseteq GoodHotel$ and $GoodHotel^- \sqsubseteq GoodHotel$ occur in $\mathcal{B}$.

Now, it can be verified that for $\mathcal{T} = \mathcal{B} \cup \mathcal{E}$

1. The query

   $$q(h)[s] \leftarrow GoodHotel^+(h), cheapPrice(h, p)[s_1], closeTo(h, a)[s_2], Attraction(a), s{=}\min(s_1, s_2)$$

   has answer set over $\mathcal{T}$, $ans_{\mathcal{T}} = \{(h1, 0.75), (h2, 0.4)\}$;
2. The query

   $$q(h)[s] \leftarrow GoodHotel(h), cheapPrice(h, p)[s_1], closeTo(h, a)[s_2], Attraction(a), s = \min(s_1, s_2)$$

   has answer set over $\mathcal{T}$, $ans_{\mathcal{T}} = \{(h1, 0.75), (h2, 0.4), (h3, 0.6)\}$;
3. Therefore, according to Eq. 12, $P = \{h1, h2\}$, while $D = \{h1, h2, h3\}$;
4. As a consequence,

$$cf(r) = \frac{0.75 \Rightarrow 0.6 + 0.4 \Rightarrow 0.8}{3} = \frac{0.6 + 1.0}{3} = 0.5333 \ .$$

4 Final Remarks

In this paper we have proposed a method for inducing ontology inclusion axioms within the KR framework of a fuzzy DL-Lite like DL. The method extends FOIL, a popular ILP algorithm for learning sets of crisp rules, in a twofold direction: from crisp to fuzzy and from rules to inclusion axioms. Indeed, related FOIL-like algorithms are reported in the literature [12,3,11] but they can only learn fuzzy rules. Another relevant work is the formal study of fuzzy ILP contributed by [5]. Yet, it is less promising than our proposal from the practical side. Last, close to our application domain, [4] faces the problem of inducing equivalence axioms in a fragment of OWL corresponding to the $\mathcal{ALC}$ DL.

For the future we intend to define appropriate specialization operators for the fuzzy DL being considered. Also we would like to investigate in depth the impact of Open World Assumption (holding in DLs) on the proposed ILP setting, and implement and experiment our method. Finally, it can be interesting to analyze the effect of the different implication functions on the learning process.

References

1. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, Cambridge (2003)
2. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Data complexity of query answering in description logics. In: Proc. of the 10th Int. Conf. on Principles of Knowledge Representation and Reasoning, pp. 260–270 (2006)
3. Drobics, M., Bodenhofer, U., Klement, E.-P.: FS-FOIL: an inductive learning method for extracting interpretable fuzzy descriptions. Int. J. Approximate Reasoning 32(2-3), 131–152 (2003)
4. Hellmann, S., Lehmann, J., Auer, S.: Learning of OWL Class Descriptions on Very Large Knowledge Bases. Int. J. on Semantic Web and Information Systems 5(2), 25–48 (2009)
5. Horváth, T., Vojtás, P.: Induction of fuzzy and annotated logic programs. In: Muggleton, S.H., Otero, R., Tamaddoni-Nezhad, A. (eds.) ILP 2006. LNCS (LNAI), vol. 4455, pp. 260–274. Springer, Heidelberg (2007)
6. Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic: theory and applications. Prentice-Hall, Inc., Upper Saddle River (1995)
7. Lukasiewicz, T., Straccia, U.: Top-k retrieval in description logic programs under vagueness for the semantic web. In: Prade, H., Subrahmanian, V.S. (eds.) SUM 2007. LNCS (LNAI), vol. 4772, pp. 16–30. Springer, Heidelberg (2007)
8. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the semantic web. Journal of Web Semantics 6, 291–308 (2008)
9. Nienhuys-Cheng, S.-H., de Wolf, R.: Foundations of Inductive Logic Programming. LNCS(LNAI), vol. 1228. Springer, Heidelberg (1997)
10. Quinlan, J.R.: Learning logical definitions from relations. Machine Learning 5, 239–266 (1990)
11. Serrurier, M., Prade, H.: Improving expressivity of inductive logic programming by learning different kinds of fuzzy rules. Soft Computing 11(5), 459–466 (2007)

12. Shibata, D., Inuzuka, N., Kato, S., Matsui, T., Itoh, H.: An induction algorithm based on fuzzy logic programming. In: Zhong, N., Zhou, L. (eds.) PAKDD 1999. LNCS (LNAI), vol. 1574, pp. 268–274. Springer, Heidelberg (1999)
13. Straccia, U.: Reasoning within fuzzy description logics. Journal of Artificial Intelligence Research 14, 137–166 (2001)
14. Straccia, U.: SoftFacts: a top-k retrieval engine for a tractable description logic accessing relational databases. Technical report (2009)
15. Straccia, U.: SoftFacts: A top-k retrieval engine for ontology mediated access to relational databases. In: Proc. of the 2010 IEEE Int. Conf. on Systems, Man and Cybernetics, pp. 4115–4122. IEEE Press, Los Alamitos (2010)

A Rough Set Approach to Spatio-temporal Outlier Detection

Alessia Albanese[1], Sankar K. Pal[2], and Alfredo Petrosino[1]

[1] University of Naples Parthenope - 80143 Naples, Italy
{alessia.albanese,alfredo.petrosino}@uniparthenope.it
http://cvprlab.uniparthenope.it
[2] Indian Statistical Institute, Kolkata 700 108, India
sankar@isical.ac.in
http://www.isical.ac.in/

Abstract. Detecting outliers which are grossly different from or inconsistent with the remaining spatio-temporal dataset is a major challenge in real-world knowledge discovery and data mining applications. In this paper, we deal with the outlier detection problem in spatio-temporal data and we describe a rough set approach that finds the top outliers in an unlabeled spatio-temporal dataset. The proposed method, called Rough Outlier Set Extraction (ROSE), relies on a rough set theoretic representation of the outlier set using the rough set approximations, i.e. lower and upper approximations. It is also introduced a new set, called Kernel set, a representative subset of the original dataset, significative to outlier detection. Experimental results on real world datasets demonstrate its superiority over results obtained by various clustering algorithms. It is also shown that the kernel set is able to detect the same outliers set but with such less computational time.

1 Introduction

Spatio-temporal data mining is a growing research area dedicated to the disclosure of hidden knowledge in large spatio-temporal databases, mainly through detecting periodic patterns and outliers detection. This paper addresses the problem of outlier detection in spatio-temporal data using rough set theory, proposed by Pawlak [6]. Only a few methods for outlier detection, in general and in spatio-temporal context, exploit rough set theory in order to define degrees of outlierness based on rough set concepts. Nguyen in [3] discusses a method for the detection/evaluation of outliers, as well as how to elicit background domain knowledge from outliers using multi-level approximate reasoning schemes. Y. Chen, D. Miao, and R. Wang in [4] demonstrate the application of granular computing model using information tables for outlier detection. F. Jiang, Y. Sui and C. Cao in [5] propose a new definition of outliers that exploits the rough membership function. In contrast to those approaches that interpret the rough set theory from the operator-oriented point of view [2], our method exploits the set-oriented view of rough set theory in order to define the concept of outlier in

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 67–74, 2011.

terms of its lower and upper approximations, keeping into account those objects that can neither be ruled in nor ruled out as members of the target concept.

We also introduce a new set, named Kernel Set. This is a selected subset of elements able to describe the original dataset both in terms of data structure and obtained results. We have shown the advantages of considering the Kernel Set in term of computation time by comparing the Rough Outlier Sets extracted by the original dataset with those extracted by the Kernel Set.

At this aim, the paper is organized as follows. Section 2 defines the problem. Section 3 reports the new rough set approach ROSE (Rough Outlier Set Extraction) to detect Spatio-Temporal (ST) Rough Outlier Set. Section 4 defines the Kernel Set and explains its significance to outlier detection. Sections 5 and 6 present executed tests on a real world dataset and the performance evaluation of the algorithm. Finally, conclusion remarks are given in Section 7.

2 Spatio-temporal Outlier Detection Problem

Let $S =< U, A >$ be an information system with U a normalized dataset and A its set of attributes. U can be written as follows:

$$U = \{p_i \equiv (z_{i1}, z_{i2}, ..., z_{im}) \in [0,1]^m, \quad i = 1, ..., N\}$$

where $p_i,\ i = 1, ..., N$ is a m-dimensional feature vector and A=$\{a_1, a_2, a_3, ..., a_m\}$ is its attribute set.

The proposed definition of the **Outlier Detection Problem** is as following:

Definition 1. *Given U, an integer $n > 0$ and a measure $d_{p_i}(U)$, defined over every $p_i \in U$, the Outlier Detection Problem consists of finding $\overline{n} \geq n$ objects $p_1, p_2, ..., p_n, p_{n+1}, ..., p_{\overline{n}} \in U$ such that*

$$d_{p_1}(U) \geq d_{p_2}(U) \geq ... \geq d_{p_n}(U) = d_{p_{n+1}}(U) ... = d_{p_{\overline{n}}}(U) > d_{p_j}(U), \\ \forall j = \overline{n} + 1, ..., N$$

The concept of measure is used to determine the degree of dissimilarity of each object with respect to others. Then, the *n-Outlier Set* can be formally defined:

Definition 2. *A n-Outlier Set $O \subseteq U$ is the set of $\overline{n} \geq n$ objects:*

$$O = \{p_1, ..., p_n, p_{n+1}, ..., p_{\overline{n}} \in U \ / \ \ d_{p_1}(U) \geq ... \geq d_{p_{n(U)}} = d_{p_{n+1}}(U) ... = \\ d_{p_{\overline{n}}}(U) > d_{p_j}(U) \ \ \forall j = \overline{n} + 1, ..., N\}$$

where $d_{p_i}(U),\ \forall i = 1, ..., N$ is a measure defined and computed on U.

From definition 2 it follows that $\tau = d_{p_n}(U)$ is the **outlierness threshold**, i.e.

$$\tau = inf\{max_1(d_p(U), d_q(U)), ..., max_n(d_p(U), d_q(U))\}, \forall p, q \in U \qquad (1)$$

Starting from the definition of spatial and temporal outlier due to Birant and Alp [8], we propose the following definitions applied to ST data. In this case U is a ST normalized dataset in which, at least, three attributes must be present, i.e. : the two spatial attributes and the temporal one.

Definition 3. *Given U, an integer $n > 0$ and a measure on spatial and temporal components $d^{s,t}_{p_i}(U)$, defined over every $p_i \in U$, an object $p \in U$ is a ST-***Outlier** *iff $d^{s,t}_p(U) \geq \tau$ where τ is defined as in (1).*

In a ST–context, a feasible measure to be associated to each object is based on the distances from its spatial and its temporal k-nearest neighbors [10]. Precisely:

$$d^{s,t}_p(U) = \alpha \cdot d^s_p(U) + \beta \cdot d^t_p(U) \tag{2}$$

where:

$$d^s_p(U) = \sum_{j=1}^{k} d^s(p, N^s(p, p_j)) \;\; and \;\; d^t_p(U) = \sum_{j=1}^{k} d^t(p, N^t(p, p_j)), \quad \forall p \in U \tag{3}$$

$k > 0$ is nearest neighbors number, $N^s(p, p_j)$ and $N^t(p, p_j)$ are the j-*th* spatial and temporal nearest neighbor of p, respectively and α, β are such that $\alpha+\beta = 1$. Definition 1 defines **ST-Outlier Detection Problem**, selecting a measure as in (2).

3 Rough Outlier Set Extraction (ROSE)

The goal of the proposed approach is to exploit the rough set theory to define the *Outlier Set* such as a *Rough Outlier Set (ROS)*. Let $S =< U, A >$ be an information system with U a ST normalized dataset and A its attribute set.

Given $n > 0$ (outlier number), we want to describe $O \subseteq U$ *(n–Outlier Set)* as

$$< \underline{B}(O), \overline{B}(O) > (Rough\ n - Outlier\ Set) \tag{4}$$

where $\underline{B}(O)$ is the *B–Lower approximation* and $\overline{B}(O)$ is the*B–Upper approximation* of *n–Outlier Set* with respect to an attribute subset $B \subseteq A$.

The *B–Lower approximation* $\underline{B}(O)$ is defined as the set of objects that can be certainly classified as members of the set O on the basis of the knowledge in B, while the objects in the *B–Upper approximation* $\overline{B}(O)$ as possible members of O on the basis of the knowledge in B.

At this aim, let I_B be the *B–indiscernibility relation* on the universe U:

$$I_B = \{(p_i, p_j) \in U \times U : a(p_i) = a(p_j),\ \forall a \in B\}$$

The equivalence classes $[p_j]_B$ or granules G_j of the partition induced by I_B on U are such that:

$$U = \bigcup_{j=1}^{N} G_j \;\; and \;\; G_j \cap G_j = \emptyset, \;\; i \neq j.$$

The measure in (2) is used as a spatio-temporal weight $\overline{\omega}_{G_j}(s, t, i)$, to be assigned to every granule G_j, depending on space, indicated by s, and/or on time, by t, and on iteration, by i and then the considered attribute subsets B are spatio-temporal attributes, only spatial and only temporal attribute. In our framework, the *B–Lower* and *B–Upper approximation*, at iteration i, can be defined as follows:

Definition 4. *The B–Lower approximation* $\underline{B}_i(O)$ *of n-Outlier Set O, at iteration i, is:* $\underline{B}_i(O) = \{G_j \subseteq U : \overline{\omega}_{G_j} > \tau_i\}$

$$where\ \tau_i = inf\ \{max_1^i\ (\overline{\omega}_{G_j}, \overline{\omega}_{G_k}), ..., max_n^i\ (\overline{\omega}_{G_j}, \overline{\omega}_{G_k})\},\ \forall\ G_j, G_k \subseteq U \quad (5)$$

Definition 5. *The B–Upper approximation* $\overline{B}_i(O)$ *of n–Outlier Set O, at iteration i, is:*

$$\overline{B}_i(O) = \{G_j \subseteq U : \overline{\omega}_{G_j} > \overline{\tau}_i\} \quad where: \quad \overline{\tau}_i = \tau_{i-1},\ \forall\ \ i >= 2 \qquad (6)$$

The threshold τ_1 is computed as the minimum value among the n higher values of weights assigned to the granules at first iteration.

The iterative procedure will stop when the thresholds does not vary anymore then the best Lower and Upper approximations in (4) have been reached.

ROSE Algorithm. The *Rough Outlier Set Extraction Algorithm* is designed to receive as input the universe U, k the nearest neighbors number and n the number of outliers to detect. The output of the procedure is the *ROS* (*Upper*, *Lower*, *Boundary* and *Negative Region*). At each iteration, the procedure randomly selects a subset of objects and computes their weights considering spatial and/or temporal components depending on the attribute subset B, with respect to, the *ROS* has been calculating. *UpdateUpperApprox* and *UpdateLowerApprox* functions compute the *lower* and *upper approximation* of *ROS*, using the current τ and previous τ_prev thresholds as defined in (5) and (6) respectively. A pruning strategy identifies objects from U having their weight under the threshold in order to build the *Negative Region*.

4 Kernel Set and Relevance to Outlier Detection

Let us now define *Kernel Set* $K \subseteq U$ that is a representative subset of the universe U that characterizes the overall dataset. Intuitively, this subset of U is able to maintain the structure of the universe U.

Definition 6. *Given* U *and two integers* $n > 0$, $k > 0$ *(number of nearest neighbors),* $d(U)$ *a measure defined on* U*, the Kernel Set* K *is built by adding each object* $p \in U$ *such that one of the following properties holds:*

1. $d_p(U) \geq \tau$
2. *if* $d_p(U) < \tau$*, then* $\exists q \in U$ *such that* $p \in NN^k(q)$ *and* $d_q(U) < \tau$ *and* $d_q(K - \{p\}) \geq \tau$

where $NN^k(q)$ *is the set of* k*-nearest neighbors of* q *and* $d(K)$ *is the restriction of* $d(U)$ *on* $K \subseteq U$*.*

The Definition 6 states that the objects that belong to the **Kernel Set** are:

1. object p for which $d_p(U) \geq \tau$ and hence belongs to *n–Outlier Set*.
2. object p that, even if $d_p(U) < \tau$, is one of the nearest neighbors of an object q for which hold $d_q(U) < \tau$ and $d_q(K - \{p\}) \geq \tau$.

The second property states that, once these objects p have been added to K, the measure of the object q become less than τ also in K as it is in U. Otherwise, the global structure of the dataset should be altered. The procedure allows to build also the *Kernel Set*, following the definition ([6]).
Some properties of Kernel Set have been proved:

1. a Kernel Set K contains the *n–Outlier Set*: $K \supseteq O$.
2. The Outlier Set O_K, computed from Kernel Set K is a superset of O computed from U: $O_K \supseteq O$.

The motivation of *Kernel Set* is that it is significative to outlier detection.

Indeed, outlier detection is a time consuming task, the use of *Kernel Set*, instead of U, as input of the *ROSE* procedure, have two major advantages:

– same results in terms of *rough outlier set* is obtained
– computational time is reduced due to the lower cardinality of *Kernel Set* respect to U.

5 Experimental Results and Discussion

The validation results are reported on the real-world dataset, named School Buses [7], consisting of 145 trajectories of two school buses collecting and delivering students around Athens metropolitan area in Greece for 108 distinct days.

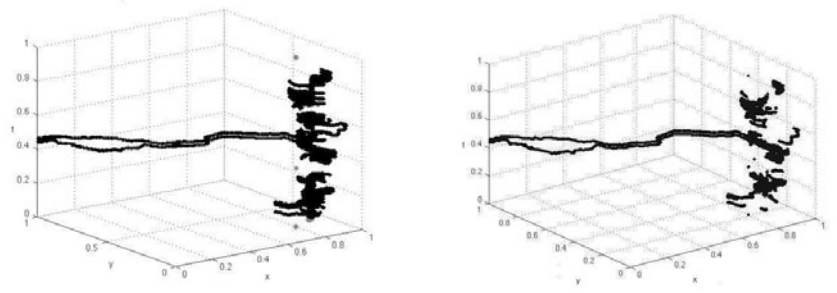
Let $< U, A >$ be the information system. U is the ST School Buses dataset, normalized and with some injected only temporal outliers (Figure 1(a)):

$$\mathrm{U} = \{p_i \equiv (z_{i,1}, z_{i,2}, z_{i,3}) \in [0,1]^3,\ i = 1, ..., N\}$$

where $(z_{i,1}, z_{i,2})$ are cartesian coordinates of the *i-th* object, $z_{i,3}$ its time-stamp. In this case, $A = \{x, y, t\}$ is the attribute set.

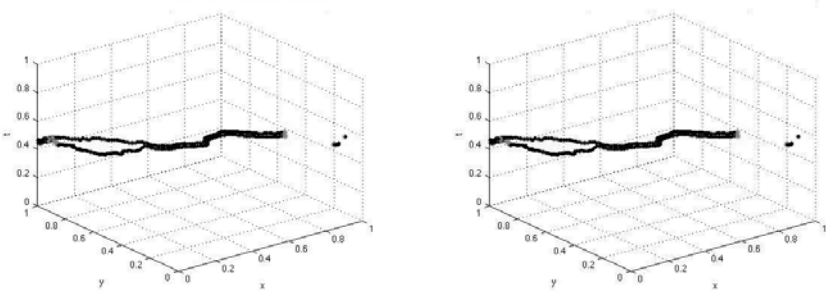
Spatial Rough Outlier Set Extraction from U
We want to describe $O \subseteq U$ as: $< \underline{B}(O), \overline{B}(O) >$ where, in this case, $B = \{x, y\} \subseteq A$, constituted by the spatial attributes. Specifically, the lower approximation and boundary at last step of *spatial-ROS* are represented and shown in Figure 2(a), where boundaries are reported in gray color. Many interesting objects should be missed without keeping into account the boundary.

Spatio-temporal Rough Outlier Set Extraction from U
We want to describe $O \subseteq U$ as: $< \underline{B}(O), \overline{B}(O) >$ where, in this case, $B = \{x, y, t\} = A$, so we are describing the *ST-ROS*. The ST-outliers will be the more relevant spatial and temporal outliers (see injected temporal outliers marked as gray stars in Figure 1(a)). The lower approximation includes the most part of the spatial and temporal outliers, while the upper includes the remaining part of the temporal outliers and some other spatial outliers. Figure 3(a) shows the lower approximation, while Figure 3(b) shows the lower approximation with boundaries in gray color.



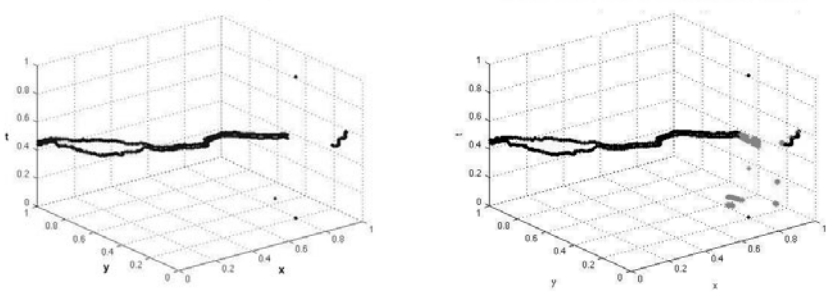
(a) (b)

Fig. 1. School Buses Testing Subset: (a) Injected Temporal Outliers (b) Kernel Set



(a) (b)

Fig. 2. Last Step of S-Rough Outlier Set: (a) Lower Approximation U Boundary from Dataset (b) Lower Approximation U Boundary from Kernel Set



(a) (b)

Fig. 3. Last Step of ST-Rough Outlier Set: (a) Lower Approximation (b) Lower Approximation U Boundary

Spatial Rough Outlier Set Extraction from Kernel Set
Figure 1(b) shows the *Kernel Set* of School Buses testing subset. Now, we want to show benefits of considering this set, comparing *S-ROS* extracted by the universe U with one extracted by the *Kernel Set*.

Starting from the *Kernel Set* and selecting only spatial components, the *ROS* is built by our approach *ROSE*. At this aim, let $< U, A >$ be the information system, with $U = K$. Then $B = \{x, y\}$. Figure 2(b) shows the lower approximation with boundaries in gray color at the last iteration. Comparing these results with the last iteration of *ROSE* for the extraction of the *Spatial Rough Outlier Set* from the entire universe U, shown in Figure 2(a), we can appreciate that the results are the same with a considerable computational time benefit.

Table 1. (a) Spatial and (b) Spatio-Temporal Outlier Detection - Quantitative Evaluation of Algorithms - Chosen Initial Centroids

Methods	α **Index**	ρ **Index**	γ **Index**	DB **Index**
ROSE	0.9836	0.0164	0.9987	N.A.
RFCM	0.5448	0.4551	0.9250	0.0736
RPCM	0.4725	0.5274	0.7919	1.1077
RFPCM	0.5645	0.4354	0.9007	0.8983

Methods	α **Index**	ρ **Index**	γ **Index**	DB **Index**
ROSE	0.8941	0.1059	0.9514	N.A.
RFCM	0.3549	0.6450	0.6444	1.8066
RPCM	0.3283	0.6716	0.5914	1.1077
RFPCM	0.3651	0.6348	0.6618	1.3299

6 Quantitative Measures and Indices

In this section, we use performance indices as introduced by Maji and Pal in [9] such as α, ρ and γ indices, as well as the *DaviesBouldin (DB)* measure as introduced in [1], to evaluate the performance of ROSE compared with some other *rough–fuzzy* clustering algorithms [9], i.e.: RFCM - Rough Fuzzy C-Means, RPCM - Rough Possibilistic C-Means, RFPCM - Rough Fuzzy Possibilistic C-Means. Parameter setting: $c = 2$ (cluster number, i.e. *inlier cluster* and *outlier cluster*), ω and $\tilde{\omega}$ (importance of lower and boundary) both equal to 0.5. We report only the final prototypes of the best solution, obtained for a particular choice of initial centroids. Table 1(a) and Table 1(b) report the best results obtained for RFCM, RPCM and RFPCM. Table 1(a) and Table 1(b) compare the performance of these algorithms with respect to α, ρ, γ and $DBindex$ in Spatial and ST-Outlier Detection respectively. Although the hybridization versions of c-means algorithm were not designed as outlier detectors, generate good prototypes for $c = 2$. In Spatial Outlier Detection, the RFPCM provides the best results and the results of other two are quite similar to that of the RFPCM; while in ST-Outlier Detection, the RPCM outperform them. The proposed ROSE algorithm performs better than RFCM, RPCM and RFPCM algorithms, both in terms of qualitative measures and of outliers detected, as shown in figures 3(b) and 2(a).

7 Conclusions

The paper reports a new rough set based outlier detection method, called *ROSE*, that has been theoretically grounded based on a definition of Outlier Set as Rough Set. The results of the proposed method have been shown and have been also compared with some other *rough–fuzzy* clustering algorithms, incorporating the concepts of rough sets, producing reasonable results both from quantitative and qualitative standpoints. A definition of a new set, called *Kernel Set*, has been also provided. The *Kernel Set* is a subset of U, significative to outlier detection. It has been shown that this set is able to detect the same outliers with less computational time.

References

1. Bezdek, J.C., Pal, N.R.: Some new indexes for cluster validity. IEEE Trans. Syst., Man, Cybern. B, Cybern. 28(3), 301–315 (1988)
2. Yao, Y.Y.: Two views of the theory of rough sets in finite universes. International Journal of Approximate Reasoning 15, 291–317 (1996)
3. Nguyen, T.T.: Outlier Detection: An Approximate Reasoning Approach. Springer, Heidelberg (2007)
4. Chen, Y., Miao, D., Wang, R.: Outlier Detection Based on Granular Computing. Springer, Heidelberg (2008)
5. Jiang, F., Sui, Y., Cunge: Outlier Detection Based on Rough Membership Function. Springer, Heidelberg (2006)
6. Pawlak, Z.: Rough Sets, Theoretical Aspects of Reasoning about data. Kluwer, Dordrecht (1991)
7. Frentzos, E., Gratsias, K., Pelekis, N., Theodoridis, Y.: Nearest Neighbor Search on Moving Object Trajectories. In: Anshelevich, E., Egenhofer, M.J., Hwang, J. (eds.) SSTD 2005. LNCS, vol. 3633, pp. 328–345. Springer, Heidelberg (2005)
8. Birant, D., Kut, A.: Spatio-Temporal Outlier Detection in Large Databases. Journal of Computing and Information Technology 14(4), 291–297 (2006)
9. Maji, P., Pal, S.K.: Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices. IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics 37(6) (December 2007)
10. Albanese, A., Petrosino, A.: A Non Parametric Approach to the Outlier Detection in Spatio-Temporal Data Analysis. In: D'Atri, et al. (eds.) Springer book Information Technology and Innovation Trends in Organizations (2011)

From Fuzzy Models to Granular Fuzzy Models

Witold Pedrycz

Department of Electrical & Computer Engineering
University of Alberta, Edmonton Canada
and
Systems Research Institute, Polish Academy of Sciences
Warsaw, Poland
pedrycz@ee.ualberta.ca

Abstract. Fuzzy models occupy one of the dominant positions on the research agenda of fuzzy sets exhibiting a wealth of conceptual developments and algorithmic pursuits as well as a plethora of applications. Granular fuzzy modeling dwelling on the principles of fuzzy modeling opens new horizons of investigations and augments the existing design methodology exploited in fuzzy modeling. In a nutshell, granular fuzzy models are constructs built upon fuzzy models or a family of fuzzy models. We elaborate on a number of compelling reasons behind the emergence of granular fuzzy modelling, and granular modeling, in general. Information granularity present in such models plays an important role. Given a fuzzy model M, the associated granular model incorporates granular information to quantify a performance of the original model, facilitate collaborative pursuits of knowledge management and knowledge transfer. We discuss several main categories of granular fuzzy models where such categories depend upon the formalism of information granularity giving rise to interval-valued fuzzy models, fuzzy fuzzy model (fuzzy^2 models, for short), and rough -fuzzy models. The design of granular fuzzy models builds upon two fundamental concepts of Granular Computing: the principle of justifiable granularity and an optimal allocation (distribution) of information granularity. The first one supports a construction of information granules of a granular fuzzy model. The second one emphasizes the role of information granularity being treated as an important design asset. The underlying performance indexes guiding the design of granular fuzzy models are discussed and a multiobjective nature of the construction of these models is stressed.

Keywords: information granularity, granular fuzzy models, principle of justifiable granularity, knowledge management, granularity allocation.

1 Introduction

In system modeling, knowledge management comes vividly into the picture when dealing with a collection of individual models. Such models being considered as sources of knowledge, are engaged in some collective pursuits of a collaborative development to arrive at modeling outcomes of a global nature. The result comes

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 75–82, 2011.

in the form of a so-called granular fuzzy model, which directly reflects upon and quantifies the diversity of the available sources of knowledge (local models) involved in knowledge management.

Let us consider a system for which constructed is a family of models. The system can be perceived from different points of view, observed over some time periods and analyzed at different levels of detail. Subsequently, the resulting models are built with different objectives in mind. They offer some particular, albeit useful views at the system. We are interested in forming a *holistic* model of the system by taking advantage of the individual sources of knowledge - models, which have been constructed so far. When doing this, we are obviously aware that the sources of knowledge exhibit diversity and hence this diversity has to be taken into consideration and carefully quantified. No matter what the local models look like, it is legitimate to anticipate that the global model (say, the one at the higher level of hierarchy) is more general, abstract. Granularity of information [1] [7] becomes of paramount importance, both from the conceptual as well as algorithmic perspective, in the realization of granular fuzzy models. Subsequently, processing realized at the level of information granules gives rise to the discipline of Granular Computing [2] [3]. From the algorithmic perspective, fuzzy clustering [4] and clustering are regarded as fundamental development frameworks in which information granules are constructed.

The objective of this study is to introduce a concept of granular fuzzy models, offer several convincing arguments behind their emergence and provide with taxonomy of categories of tasks in which information granularity arises as an important design asset facilitating constructive pursuits in knowledge management. We start with a discussion of the two conceptual and algorithmic underpinnings of Granular Computing, namely a principle of justifiable granularity and an optimal allocation of information granularity (Section 2 and 3). These two ideas are essential to the realization of granular fuzzy models. Four fundamental modes of knowledge management are identified (Section 4). Conclusions are covered in Section 5.

2 The Principle of Justifiable Granularity

Here we are concerned with the formation of a single information granule based on some experimental evidence being a set of a single-dimensional numeric data $D = \{x_1, x_2, ..., x_N\}$. The principle of justifiable granularity [5] is concerned with a formation of a meaningful information granule. Such construct has to adhere to the two intuitively requirements:

(i) the numeric evidence accumulated within the bounds of Ω has to be as high as possible. By doing so, we anticipate that the existence of the information granule is well motivated (justified) as being reflective of the existing experimental data.
(ii) at the same time, the information granule should be as specific as possible meaning that it comes with a well-defined semantics. In other words, we would like to have Ω as detailed (specific) as possible.

While these two requirements are appealing, they have to be translated into some operational framework where the formation of the information granule can be realized. This framework depends upon the formalism of information granulation, viz. a way in which information granules are described as sets, fuzzy sets, shadowed sets, rough sets, probabilistic granules and alike. To start with a simple and convincing constructs, let us treat Ω as an interval to be constructed. The first requirement is quantified by counting the number of data falling within the bounds of Ω. In the simplest scenario, we can look at the cardinality of Ω, namely $\text{card}\{x_k \in \Omega\}$. More generally, we can consider an increasing functional of the cardinality, say $f_1(\text{card}\{x_k \in \Omega\})$. The simplest case concerns an identity functional, $f_1(u) = u$. The specificity of the information granule can be quantified by looking at its size. The length of the interval Ω or a decreasing functional of the length, f_2, can serve as a sound measure of specificity. The lower the value of $f_2(\text{length}(\Omega))$, the higher the specificity is. It is apparent that the two requirements discussed above are in conflict.

Let us proceed with the detailed construct of interval information granules. We start with a determination of the numeric representative of the set of data D. A sound representative is its median, $\text{med}(D)$ as it is robust estimator of the sample and is one of the elements of D. An information granule Ω is formed by forming its lower and upper bound, denoted by a and b, respectively; refer also to Figure 1.

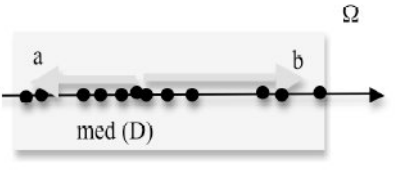


Fig. 1. Optimization of interval information granule Ω

The determination of the bounds is realized independently. In this sense, we can concentrate on the optimization of the upper bound (b). The calculations of the lower bound (a) are carried out in an analogous fashion. The length of Ω, which quantifies the specificity of the information granule is given now as $|\text{med}(D) - b|$. More generally, we have $f_2(|\text{med}(D) - b|)$. The cardinality of the information granule takes into account the elements of D positioned to the right from the median, $\text{card}\{x_k \in \Omega, x_k > \text{med}(D)\}$. Again in general, we compute $f_1(\text{card}\{x_k \in \Omega, x_k > \text{med}(D)\})$. As the requirements of experimental evidence and specificity are in conflict, we can either resort ourselves to multiobjective optimization or consider a maximization of the product $V = f_1 * f_2$ whose optimization is to be realized with respect to the upper bound of the information granule, that is $V(b_{opt}) = \max_{b>\text{med}(D)} V(b)$. One among possible design alternatives, we can consider the functionals f_1 and f_2 assuming the following form

$$f_1(u) = u \tag{1}$$

and

$$f_2(u) = \exp(-\alpha u) \tag{2}$$

where α is a positive parameter offering some flexibility in the produced information granule A. Its role is to calibrate an impact of the specificity criterion on the constructed information granule.

3 Optimal Allocation of Information Granularity

Information granularity is an important design asset. Information granularity allocated to the original numeric construct elevates a level of abstraction (generalizes) of the original construct developed at the numeric level. A way in which such an asset is going to be distributed throughout the construct or a collection of constructs to make the abstraction more efficient, is a subject to optimization.

Consider a certain mapping $y = f(\boldsymbol{x}, \boldsymbol{a})$ with $\boldsymbol{a}$ being a vector of parameters. The mapping can be sought in a general way. One may think of a fuzzy model, neural network, polynomial, differential equation, linear regression, etc. The granulation mechanism $\boldsymbol{G}$ is applied to a giving rise to its granular counterpart, $\boldsymbol{A}= \boldsymbol{G}(\boldsymbol{a})$ and subsequently producing a granular mapping, $Y = \boldsymbol{G}(f(\boldsymbol{x}, \boldsymbol{a})) = f(\boldsymbol{x}, \boldsymbol{G}(\boldsymbol{a})) = f(\boldsymbol{x}, \boldsymbol{A})$. Given the diversity of the underlying constructs as well as a variety of ways information granules can be formalized, we arrive at a suite of interesting constructs: granular neural networks, say interval neural networks, fuzzy neural networks, probabilistic neural networks, etc.

There are a number of well-justified and convincing arguments behind elevating the level of abstraction of the existing constructs. Those include: an ability to realize various mechanisms of collaboration, quantification of variability of sources of knowledge considered, better modelling rapport with systems when dealing with nonstationary environments. In what follows, we will elaborate on the general categories of problems in which information granularity plays a pivotal role.

Information granularity provided to form a granular construct is a design asset whose allocation throughout the mapping can be guided by certain optimization criteria. Let us discuss the underlying optimization problem in more detail. In addition to the mapping itself, we are provided with some experimental evidence in the form of input-output pairs $(\boldsymbol{x}_k, t_k)$, $k = 1, 2, , M$. Given is a level of information granularity ϵ , $\epsilon \in [0, 1]$. We allocate the available level ϵ to the parameters of the mapping, $\dim(\boldsymbol{a}) = h$, so that the some optimization criteria are satisfied while the allocation of granularity satisfies the following balance $\epsilon = \sum_{i=1}^{h} \epsilon_i$ where ϵ_i is a level of information granularity associated with the i-th parameter of the mapping. All of the individual allocations are organized in a vector format $[\epsilon_1 \epsilon_2 ... \epsilon_h]^T$.

There are two optimization criteria to be considered in the optimization. The first one is concerned with the coverage of data t_k. For $\boldsymbol{x}_k$ we compute $Y_k = f(\boldsymbol{x}_k, \boldsymbol{G}(\mathbf{a}))$ and determine a degree of inclusion of t_k in information granule Y_k, $\text{incl}(t_k, Y_k) = t_k \subset Y_k$. Then we compute an average sum of the degrees of inclusion taken over all data, that is $\frac{1}{M} \sum_{k=1}^{M} \text{incl}(t_k, Y_k)$. Depending upon the formalism of information granulation, the inclusion returns a Boolean value in case of intervals (sets) or a certain degree of inclusion in case of fuzzy sets.

The second criterion is focused on the specificity of Y_k- we want it to be as high as possible. The specificity could be viewed as a decreasing function of the length of the interval in case of set -based information granulation. For instance, one can consider the inverse of the length of Y_k, say $1/\text{length}(Y_k)$, $\exp(-\text{length}(Y_k))$, etc. In case of fuzzy sets, one consider the specificity involving the membership grades. The length of the fuzzy set Y_k is computed by integrating the lengths of the β-cuts, $\int_0^1 lenght(Y_k^\beta)\beta d\beta$.

More formally, the two-objective optimization problem is formulated as follows. Distribute (allocate) a given level of information granularity ϵ so that the following two criteria are maximized.

$$\begin{aligned} &Max_{\epsilon_1,\epsilon_2,\ldots,\epsilon_h} Q_1, \quad Q_1 = \frac{1}{M}\sum_{k=1}^{M} \text{incl}(t_k, Y_k) \\ &Max_{\epsilon_1,\epsilon_2,\ldots,\epsilon_h} Q_2, \quad Q_2 = g(\text{length}(Y_k)) \\ &\text{(where } g \text{ is a decreasing function of its argument)} \\ &\text{subject to } \epsilon = \sum_{i=1}^{h} \epsilon_i \end{aligned} \tag{3}$$

A simpler, single-objective optimization scenario involves a coverage criterion regarded as a single most essential criterion considered in the problem

$$Max_{\epsilon_1,\epsilon_2,\ldots,\epsilon_h} Q, \quad Q = \frac{1}{M}\sum_{k=1}^{M} \text{incl}(t_k, Y_k) \text{ subject to } \epsilon = \sum_{i=1}^{h} \epsilon_i \tag{4}$$

We can arrive at some global view at the relationship that is independent from a specific value of ϵ by taking an area under curve (AUC) computed as $\text{AUC}=\int_0^1 Q(\epsilon)d\epsilon$. The higher the value of the AUC, the higher the performance of the granular version of the mapping.

Information granularity can be realized in the setting of a certain information allocation protocol. Several main categories of such protocols can be envisioned:

- $\mathtt{P}_1$: uniform allocation of information granularity. This process is the simplest one and in essence does not call for any optimization mechanism. The numeric parameter a of the mapping is replaced by the information granule $\boldsymbol{G}(a)$, which is the same in terms of the size and the distribution around a. If the formal setup of $\boldsymbol{G}$ concerns intervals then the numeric parameters of the mapping are replaced by intervals of the same length (ϵ) and distributed symmetrically around the parameters of the mapping.
- $\mathtt{P}_2$: uniform allocation of information granularity with asymmetric position of intervals.
- $\mathtt{P}_3$: non-uniform allocation of information granularity with symmetrically distributed intervals of information granules.
- $\mathtt{P}_4$: non-uniform allocation of information granularity with asymmetrically distributed intervals of information granules.

P_5: An interesting point of reference, which is helpful in assessing a relative performance of the above methods, is to consider a random allocation of granularity. By doing this, one can quantify how the optimized and carefully thought out process of granularity allocation is superior over a purely random allocation process.

Depending upon the formalism of information granularity, the protocols can be made more specific. For instance, Figure 2 illustrates a collection of scenarios where information granules are represented as intervals.

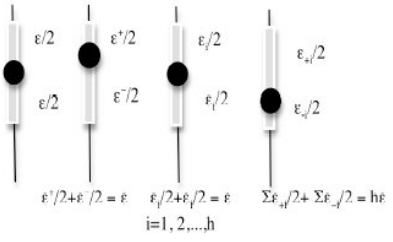


Fig. 2. Protocols $P_1 - P_4$ of information granularity allocation - a case of interval information granulation

4 Fundamental Modes of Knowledge Management

Fuzzy models are sought as sources of knowledge. They interact in a variety of ways. It is advantageous to introduce a certain taxonomy, in which we identify four modes of interaction as shown in Figure 3:

Aggregation of sources of knowledge. The focal point concerns the formation of a global view about the system based upon the results (outcomes) produced by the individual models.

Granular compression. A source of knowledge is compressed by forming a more compact model based on the original model. The compression brings about a concept of granularity- the compressed counterpart is more compact but inherently granular.

Building consensus. The sources of knowledge (models) are actively engaged in the realization of a holistic view however in contrast to the previous category, there is an active involvement of the individual sources in the overall process in the sense they can constructively react to the global view created based on the individual and make some adjustments to themselves to increase a level of consensus.

Knowledge transfer. A source of knowledge has been formed on some experimental evidence present in the past. New limited data originating from the similar process (which could evolve over time) stipulate that the existing source of knowledge could be used effectively by generalizing originally available model by

making its granular. The granular model takes into account an effect of a partial relevance of the knowledge (model) acquired so far and through the allocation of information granules makes provisions with this regard.

In all these situations, the concept of information granules plays a pivotal role in several meaningful ways. First this concept facilitates or becomes a necessary prerequisite to the realization of knowledge management. Second, which is equally important, it quantifies the quality of the resulting constructs in the language of information granularity meaning that they become inherently made more abstract by incorporating information granules. Thus the specificity of the results is fully reflective of the diversity of the sources of knowledge and this helps assess the feasibility of interaction processes as well as undertake some necessary steps in cases the lack of specificity becomes too substantial (which speaks to the high level of diversity among sources knowledge being overly detrimental to the interaction there). Furthermore Figure 3 identifies a location of information granules as they emerge in the processes described above (the granular constructs are depicted by the shaded rectangular shapes).

In the first case, Fig.3 (a), granularity of information is reflective of the existing diversity of the sources of knowledge. In granular compression, the factor of granularity of information quantifies a trade-off between achieved level of compression and the associated level of abstraction of the resulting construct. In the two other cases, Fig.3 (c) - (d) information granularity is regarded as an important design asset which helps establish the underlying processes of consensus formation or quantifying knowledge transfer. The level of information granularity is introduced into the process externally and this design asset can be further

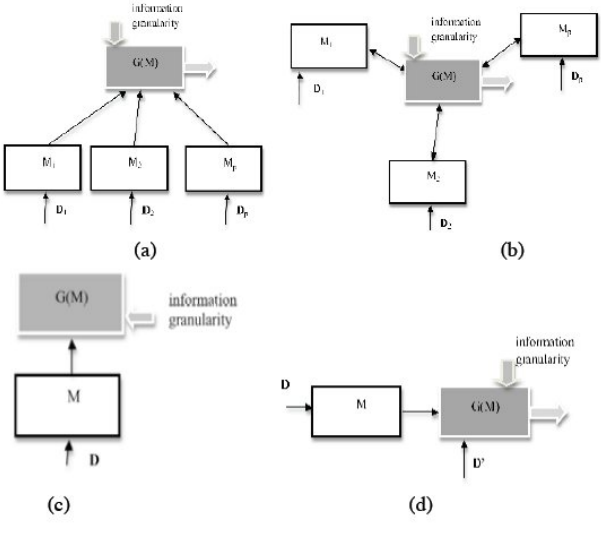


Fig. 3. Four modes of knowledge management: (a) aggregation, (b) consensus building, (c) granular compression, and (d) knowledge transfer

optimized in terms of its allocation to the corresponding models involved in the scheme.

The granular models obtained as a result the interaction establish themselves at the higher level of abstraction in comparison with the available sources of knowledge. The term of granular fuzzy model is conceptually quite general as the realization depends upon a way in which information granules are expressed. Depending on this, we may talk about fuzzy fuzzy (fuzzy^2) models, interval fuzzy models, rough fuzzy models, etc.

5 Concluding Comments

Granular fuzzy models arise as a new conceptually sound alternative of system modeling. It has been demonstrated that knowledge management involving a number of sources of knowledge is directly associated with the design of information granules. It needs to be stressed that the level of information granularity is essential to the realization of knowledge management either supporting a formation of information granules or facilitating collaboration through endowing the individual sources of knowledge by some additional flexibility associated with the constructed information granules. The study offers a general taxonomy of the key scenarios. It has to be noted, however, that only several representative design problems were sketched that deserve detailed algorithmic investigations.

References

1. Bargiela, A., Pedrycz, W.: Granular Computing: An Introduction. Kluwer Academic Publishers, Dordrecht (2003)
2. Bargiela, A., Pedrycz, W. (eds.): Human-Centric Information Processing Through Granular Modelling. Springer, Heidelberg (2009)
3. Bargiela, A., Pedrycz, W.: Toward a theory of Granular Computing for human-centered information processing. IEEE Transactions on Fuzzy Systems 16(2), 320–330 (2008)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, N. York (1981)
5. Pedrycz, W., Gomide, F.: Fuzzy Systems Engineering: Toward Human-Centric Computing. John Wiley, Hoboken (2007)
6. Pedrycz, W., Song, M.: Analytic Hierarchy Process (AHP) in group decision making and its optimization with an allocation of information granularity. IEEE Trans. on Fuzzy Systems (to appear 2011)
7. Zadeh, L.A.: Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90, 111–117 (1997)

Multi-objective Evolutionary Fuzzy Systems

Pietro Ducange and Francesco Marcelloni

Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, Largo Lucio Lazzarino 1 - 56122 Pisa, Italy
{p.ducange,f.marcelloni}@iet.unipi.it

Abstract. Several methods have been proposed to automatically generate fuzzy rule-based systems (FRBSs) from data. At the beginning, the unique objective of these methods was to maximize the accuracy with the result of often neglecting the most distinctive feature of the FRBSs, namely their interpretability. Thus, in the last years, the automatic generation of FRBSs from data has been handled as a multi-objective optimization problem, with accuracy and interpretability as objectives. Multi-objective evolutionary algorithms (MOEAs) have been so often used in this context that the FRBSs generated by exploiting MOEAs have been denoted as multi-objective evolutionary fuzzy systems. In this paper, we introduce a taxonomy of the different approaches which have been proposed in this framework. For each node of the taxonomy, we describe the relevant works pointing out the most interesting features. Finally, we highlight current trends and future directions.

Keywords: Multi-objective evolutionary fuzzy systems, Fuzzy rule-based systems, Interpretability in fuzzy rule-based systems, Multi-objective evolutionary algorithms.

1 Introduction

Fuzzy rule-based systems (FRBSs) have been successfully applied to different engineering fields such as control, pattern recognition, system identification and signal analysis. FRBSs consist of a linguistic rule base (RB), a data base (DB) containing the fuzzy sets associated with the linguistic terms used in the RB and a fuzzy logic inference engine. The most natural approach to FRBS design is to elicit the knowledge from a human expert and to codify this knowledge in the RB and DB of the FRBS. In some application domains, this knowledge can be limited, due, for instance, to the complexity of the domain, thus making this natural approach not practicable. To overcome this problem, methods have been proposed in the literature to generate the RB and DB by extracting this knowledge from available information (typically, input-output samples). At the beginning, such generation was generally performed with the unique objective of maximizing the accuracy, but soon the researchers realized that these accuracy-driven approaches typically produce FRBSs characterized by a high number of rules and by linguistic fuzzy partitions with a low level of comprehensibility, thus loosing that feature which has made FRBSs preferable to other approaches

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 83–90, 2011.

in real applications, namely interpretability [1]. To overcome this problem, in the last decade, new methods have been proposed to generate FRBSs taking not only accuracy, but also interpretability of rule and data bases into consideration. Hence, the generation of FRBSs requires to solve a multi-objective optimization problem.

Multi-Objective Evolutionary Algorithms (MOEAs) have been so extensively used in this framework that the term "Multi-Objective Evolutionary Fuzzy Systems" (MOEFSs) has been coined to identify the hybridization of FRBSs with MOEAs [2]. MOEAs are employed to generate FRBSs with different trade-offs between accuracy and interpretability by learning the overall RB or by selecting subsets of rules from heuristically determined initial RBs, and by learning the overall DB or by tuning a DB proposed by the experts. Since accuracy and interpretability are conflicting objectives, MOEFSs do not generate a unique optimal FRBS, but a set of FRBSs, denoted Pareto optimal set, characterized by different optimal trade-offs among the objectives.

In the paper, first of all, we will discuss how interpretability of an FRBS has been evaluated in the MOEFSs proposed so far. Second, we will introduce a taxonomy of MOEFSs based on how RB and DB are processed during the evolutionary process. For each node of the taxonomy, we will describe the most relevant features of the approaches proposed in the literature. Finally, we will discuss hot topics and new challenges.

2 Interpretability in Fuzzy Rule-Based Systems

Interpretability of FRBSs has been widely discussed in the last years, especially in the framework of MOEFSs. FRBSs can be considered intrinsically "transparent" because their behavior can be explained in terms of their components and their relations [3]. On the other hand, the concept of "transparency" is quite far from system interpretability. Among the different types of FRBSs, Mamdani FRBSs have had a predominant role in MOEFSs, thanks to their feature of being completely defined in linguistic terms and therefore particularly comprehensible to the users. In the following, we will focus on this type of FRBSs.

Since interpretability is a subjective concept, it is hard to propose a world-wide agreed definition and consequently a universal measure of interpretability. Thus, researchers have focused their attention on discussing some factors which characterize interpretability and on proposing some constraints which have to be satisfied for these factors. Considerations on the main factors that influence interpretability can be found in [1] [4]. A homogeneous description of semantic and syntactic interpretability issues regarding both the RB and the DB has been recently published in a survey on interpretability constraints [3].

In [5], a taxonomy of fuzzy model interpretability has been proposed in terms of both low-level and high-level interpretability. Low-level interpretability is related to the semantic constraints that ensure fuzzy partition interpretability while high-level interpretability is associated with a set of criteria defined on the RB. Furthermore, a conceptual framework for characterizing interpretability

of fuzzy systems has been introduced in [6]: this framework includes a global description of the FRBS structure, on the basis of the taxonomy and constraints discussed in [3] and [5], respectively, and a local explanation for understanding the system behavior. This local explanation considers a number of factors such as inference mechanisms, aggregation, conjunction and disjunction operators, defuzzification and rule type, which affect the system behavior.

Recently, the most relevant measures and strategies exploited to design interpretable FRBSs have been reviewed in [7]. Here, a taxonomy of the interpretability measures has been proposed by considering two different dimensions, namely semantic and complexity, at RB and DB levels. As shown in Fig. 1, this taxonomy is therefore organized into four quadrants. In the following sections, we will discuss how most of these measures, especially the ones in the complexity-RB and semantic-DB quadrants, have been used in MOEFSs.

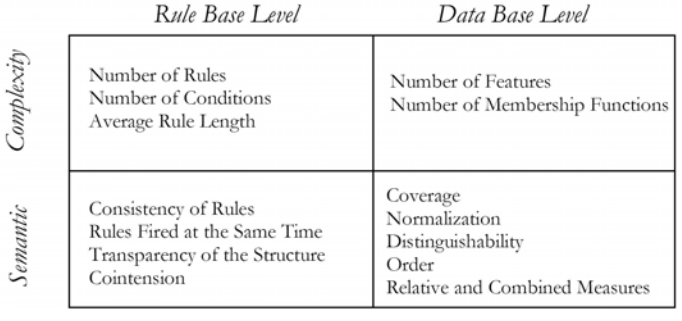


Fig. 1. Taxonomy of interpretability measures proposed in the literature

3 A Taxonomy of Multi-objective Evolutionary Fuzzy Systems

In [2], F. Herrera introduced a taxonomy of genetic fuzzy systems based on how FRBS components are processed during the evolution. In Fig. 2, we adapt this taxonomy to MOEFSs. At the first level, MOEFSs can be divided into two macro-classes, namely multi-objective evolutionary (MOE) tuning and learning, according to the type of process used to design the FRBS components. Although it is difficult to make a clear distinction between tuning and learning, we can adopt the following definition proposed in [2]: in the former, we tune the components, typically the membership functions, of a pre-existing KB so as to increase the accuracy of the FRBSs, whereas in the latter, we learn the FRBS components possibly without any pre-existing KB or at most with a partially defined KB. The MOE learning approaches are, in their turns, split into four classes according to the specific components which are learned during the evolutionary process. MOE tuning and learning have been also combined in some recent approaches. In the following, we will briefly describe the classes of approaches by referring to the most important techniques proposed per each class.

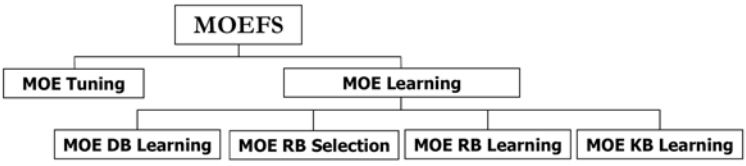


Fig. 2. MOEFS taxonomy

3.1 MOE Tuning

Given a KB, optimization is performed by adjusting the DB definition, without changing the existing RB. Typically, the RB is provided by an expert or is generated by exploiting some heuristic approach, commonly Wang and Mendel [8], and the initial DB is obtained by uniformly partitioning the universes of the linguistic variables used in the rules. MOE tuning exploits chromosomes that define the operations used to tune the membership functions.

In [9], for each linguistic variable, the chromosome codifies the parameters of 5 operators, namely one non-linear scaling function and four linguistic modifiers, used to adapt the DB: the first five bits, one for each operator, control whether the corresponding operator is applied, and the remaining 72 bits are organized in sub-strings of 8 bits, where each sub-string determines the value of a different parameter, via Gray decoding and quantization. The well-known NSGA-II [10] is used to generate an approximated Pareto front of DBs by optimizing accuracy and a novel index based on fuzzy ordering relations.

Some recent tuning approaches are performed together with RB selection and will be discussed in Section 3.3.

3.2 MOE DB Learning

In MOE DB learning, the DB generation process wraps an RB heuristic-based learning process. Chromosomes consist generally of two parts: an integer part which codifies the granularity of the partitions and a real part which defines the parameters of the membership functions. At each iteration of the MOEA, an RB is generated for each DB by using some heuristic approach such as Wang and Mendel [8], and then the accuracy and the interpretability of the overall KB are computed. One of the most recent approaches of DB learning has been proposed in [11], where granularity is chosen in [1..7] and [2..7] for, respectively, the input and output variables, and a real gene in [-0.1,0.1] for each linguistic variable determines the lateral displacement of each partition.

3.3 MOE RB Selection

In MOE RB selection, the evolutionary process generates RBs by selecting rules from an initial RB produced by some heuristic. Binary chromosomes of length equal to the number of rules in the initial RB are generally used where each bit

determines whether the corresponding rule of the initial RB is selected or not. Accuracy and complexity are computed by using a pre-defined DB, typically obtained by uniformly partitioning the input and output linguistic variables. The pioneer work [12] in MOEFSs belongs to this class. Here, accuracy and interpretability are measured in terms of number of correctly classified training patterns and number of rules, respectively. This work has been enriched in [13] by proposing a novel heuristic to generate the initial RB, by introducing a multi-objective genetic local search algorithm where a local search procedure adjusts the selection of each candidate rule and by combining this algorithm with a learning algorithm of rule weights.

In [14], the RB selection has been extended with a second level of selection performed on the antecedent conditions of each selected rule. The chromosome is composed by a vector of pairs $p_m = (k_m, v_m)$, where k_m is an integer which identifies a rule in the initial RB and v_m is a vector of bits. Each bit in v_m determines whether the corresponding condition of rule k_m has to be included or not in the RB.

MOE RB selection is often performed together with DB tuning. In these cases, both RB complexity and DB integrity are taken into account during the evolutionary process. In [15] [16] [17], chromosomes consist of two parts, which adopt binary and real coding, respectively. The former selects rules from the initial RB and the latter tunes the membership functions parameters. Triangular membership functions are identified by the classical three-point representation and each gene codifies the lateral displacement of each point with respect to the initial value. Appropriate constraints on the variation intervals of the genes ensure the partition integrity in [15] [16], while in [17] a new semantic interpretability index, which aggregates three different metrics aimed at preserving as much as possible the original meaning of the membership functions, is used as objective during the evolutionary process.

In [18], the initial KB is generated in two steps: first, the heuristic proposed in [13] is used to generate a set of candidate rules with multiple granularities. Then, a single granularity is specified for each input variable according to the frequency of employed partitions and the importance of the multiple granularity-based extracted rules. The MF parameters tuning is performed by using the 2-tuple representation [19], while the number of rules, the total number of conditions in each rule and the percentage of corrected classified patterns are concurrently optimized by using NSGA-II.

3.4 MOE RB Learning

Unlike MOE RB selection, which starts from an initial RB, MOE RB learning generates the RB from scratch. In [20], each gene in the chromosome identifies the integer index $j_{m,f}$ of the linguistic value which has been selected for variable X_f in rule R_m. Accuracy and interpretability are computed in terms of root mean square error and total number of conditions in the rules. In [21], the RB learning approach has been extended for generating fuzzy rule based classifiers, in the context of imbalanced and cost sensitive datasets. In order to deal with such

datasets, the accuracy has been calculated in terms of sensitivity and specificity associated with the classifiers, while the interpretability has been assessed in terms of RB complexity. The three-objective evolutionary optimization has been carried out by using NSGA-II. Finally, in order to select the best classifiers, the ROC convex hull technique has been exploited.

In [22], since conditions can be formed by disjunctions of linguistic values, each rule is encoded by a binary string for the antecedent and an integer for the consequent. For each linguistic variable, the binary string indicates which linguistic value is used in the antecedent.

3.5 MOE KB Learning

The objective of MOE KB learning is to learn both the RB and the DB during the evolutionary process. The chromosome is typically formed by different parts which codify the RB and some parameters of the DB.

The technique proposed in [23] represents an approximation of the ideal approach to learn simultaneously the RB and the DB of a set of fuzzy rule-based classifiers. Indeed, a hybrid version between Michigan and Pittsburgh learning methods has been employed to generate sets of rules with multiple granularities as in [13]. The number of rules, the total number of conditions in each rule and the percentage of corrected classified patterns are concurrently optimized by using NSGA-II.

In [24], an integer part codifies the RB as described in [20] and a real part determines the position of the cores of triangular membership functions obtained by uniformly partitioning the universes of each linguistic variable with a pre-fixed number of fuzzy sets. Accuracy and interpretability are computed in terms of mean square error and total number of conditions in the rules. Further, DB integrity is preserved by allowing cores to be positioned in specific intervals.

In [25] [26], an integer part is added to the chromosome to determine the granularity of the partitions. Further, in [26] the MOE DB learning is performed by applying piecewise linear transformations to the uniformly partitioned linguistic variables. To preserve the shapes of the membership functions, only the extremes of the supports and the cores of the fuzzy sets are transformed. Further, the piecewise linear transformation is defined for each linguistic variable by using a number of changes of slopes equal to the maximum possible granularity minus 1. Thus, the length of the chromosome is fixed. The rule learning is performed by using the approach proposed in [20] to rules defined on virtual partitions, that is, partitions generated with the maximum granularity. The actual granularity determined by the evolutionary process is used only in the computation of the fitness: an appropriate mapping strategy maps the virtual partitions into concrete partitions and updates accordingly the rules.

In [27], as regards RB, integer and real genes are used to codify the antecedent conditions and the consequent singleton fuzzy sets, respectively. As regards DB, both the parameters of the membership functions and the positions of the singletons are codified by real genes. Semantic partition interpretability is ensured by dynamic constraints.

4 Current Trends and Future Directions

One of the most critical aspects in applying MOEFSs is the computation of the fitness, especially when the dataset is large. To reduce the computational cost, some solutions have been investigated such as employing parallel evolutionary algorithms, exploiting fitness approximation and adopting instance selection techniques [28]. Another critical aspect is to manage high-dimensional datasets since the search space increases with the number of features. To limit this drawback, feature selection during the evolutionary process and ad-hoc modified multi-objective evolutionary algorithms have being explored [11]. Finally, interpretability is now evaluated by using a number of indexes. Classical multi-objective evolutionary algorithms, which have been used so far in MOEFSs, do not result to be too suitable for handling more than 3-4 objectives. Thus, appropriate algorithms recently proposed in the multi-objective evolutionary framework will have to be experimented.

References

1. Guillaume, S.: Designing fuzzy inference systems from data: An interpretability-oriented review. IEEE Trans. Fuzzy Syst. 9(3), 426–443 (2001)
2. Herrera, F.: Genetic fuzzy systems: taxonomy, current research trends and prospects. Evolutionary Intelligence 1, 27–46 (2008)
3. Mencar, C., Fanelli, A.M.: Interpretability constraints for fuzzy information granulation. Information Sciences 178, 4585–4618 (2008)
4. de Oliveira, J.V.: Semantic constraints for membership function optimization. IEEE Trans. Syst. Man. Cybern. Part A 29(1), 128–138 (1999)
5. Zhou, S.M., Gan, J.Q.: Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modeling. Fuzzy Sets and Systems 159, 3091–3131 (2008)
6. Alonso, J.M., Magdalena, L., Gonzalez-Rodriguez, G.: Looking for a good fuzzy system interpretability index: An experimental approach. Int. J. Approx. Reason. 51, 115–134 (2009)
7. Gacto, M.J., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences (2011) (in press)
8. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Trans. Syst. Man. Cybern. 22(6), 1414–1427 (1992)
9. Botta, A., Lazzerini, B., Marcelloni, F., Stefanescu, D.: Context adaptation of fuzzy systems through a multiobjective evolutionary approach based on a novel interpretability index. Soft Comput 13(5), 437–449 (2009)
10. Coello, C., Lamont, G., Van Veldhuizen, D.: Evolutionary algorithms for solving multi-objective problems. Springer, Heidelberg (2007)
11. Alcala, R., Gacto, M.J., Herrera, F.: A fast and scalable multi-objective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems. IEEE Trans. Fuzzy Syst. (in press) doi:10.1109/TFUZZ.2011.2131657
12. Ishibuchi, H., Murata, T., Turksen, I.B.: Single-objective and two objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets and Systems 89(2), 135–150 (1997)

13. Ishibuchi, H., Yamamoto, T.: Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets and Systems 141, 59–88 (2004)
14. Antonelli, M., Ducange, P., Lazzerini, B., Marcelloni, F.: Multi-objective evolutionary generation of Mamdani fuzzy rule-based systems based on rule and condition selection. In: Proc. of the 5th IEEE GEFS 2011, Paris (France), April 11 - 15, pp. 47–53 (2011)
15. Alcalá, R., Gacto, M.J., Herrera, F., Alcalá-Fdez, J.: A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. Int. J. Uncertainty, Fuzziness Knowl.-Based Syst. 15(5), 539–557 (2007)
16. Gacto, M.J., Alcalá, R., Herrera, F.: Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. Soft Computing 13(5), 419–436 (2009)
17. Gacto, M.J., Alcalá, R., Herrera, F.: Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems. IEEE Trans. Fuzzy Syst. 18(3), 515–531 (2010)
18. Alcala, R., Nojima, Y., Herrera, F., Ishibuchi, H.: Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions. Soft Computing (in press) doi: 10.1007/s00500-010-0671-2
19. Alcalá, R., Alcalá-Fdez, J., Herrera, F.: A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. IEEE Trans. Fuzzy Syst. 15(4), 615–635 (2007)
20. Cococcioni, M., Ducange, P., Lazzerini, B., Marcelloni, F.: A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems. Soft Computing 11(11), 1013–1031 (2007)
21. Ducange, P., Lazzerini, B., Marcelloni, F.: Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets. Soft Computing 14(10), 713–728 (2010)
22. Casillas, J., Martnez, P., Bentez, A.D.: Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems. Soft Computing 13(5), 451–465 (2009)
23. Ishibuchi, H., Nojima, Y.: Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning, Internat. J. Approx. Reason. 44(1), 4–31 (2007)
24. Alcalá, R., Ducange, P., Herrera, F., Lazzerini, B., Marcelloni, F.: A multi-objective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy rule-based systems. IEEE Trans. Fuzzy Syst. 17(5), 1106–1122 (2009)
25. Antonelli, M., Ducange, P., Lazzerini, B., Marcelloni, F.: Learning concurrently partition granularities and rule bases of Mamdani fuzzy systems in a multi-objective evolutionary framework. Int. J. Approx. Reason. 50(7), 1066–1080 (2009)
26. Antonelli, M., Ducange, P., Lazzerini, B., Marcelloni, F.: Multi-objective evolutionary learning of granularity, membership function parameters and rules of Mamdani fuzzy systems. Evolutionary Intelligence 2(1-2), 21–37 (2009)
27. Pulkkinen, P., Koivisto, H.: A dynamically constrained multiobjective genetic fuzzy system for regression problems. IEEE Trans. Fuzzy. Syst. 18(1), 161–177 (2010)
28. Antonelli, M., Ducange, P., Marcelloni, F.: Exploiting a coevolutionary approach to concurrently select training instances and learn rule bases of Mamdani fuzzy systems. In: Proc. of the IEEE World Congress on Computational Intelligence, Barcelona (Spain), July 18–23, pp. 1366–1372 (2010)

Interpretability, Interpolation and Rule Weights in Linguistic Fuzzy Modeling

Andri Riid[1] and Ennu Rüstern[2]

[1] Laboratory of Proactive Technologies, Tallinn University of Technology, Ehitajate tee 5, Tallinn, 19086, Estonia
andri@dcc.ttu.ee
[2] Department of Computer Control, Tallinn University of Technology, Ehitajate tee 5, Tallinn, 19086, Estonia
ennu.rystern@dcc.ttu.ee

Abstract. Linguistic fuzzy modeling that is usually implemented using Mamdani type of fuzzy systems suffers from the lack of accuracy and high computational costs. The paper shows that product-sum inference is an immediate remedy to both problems and that in this case it is sufficient to consider symmetrical output membership functions. For the identification of the latter, a numerically efficient method is suggested and arising interpretational aspects are discussed. Additionally, it is shown that various rule weighting schemes brought into the game to improve accuracy in linguistic modeling only increase computational overhead and can be reduced to the proposed model configuration with no loss of information.

1 Introduction

It is generally acknowledged that of the two prevailing types of fuzzy systems, Mamdani systems are more interpretable than Takagi-Sugeno (TS) systems. Mamdani systems provide a better (more intuitive) mechanism for integration of expert's knowledge into the system (and vice versa) as fuzzy rules in Mamdani systems closely follow the format of natural languages and deal with fuzzy sets exclusively. On the other hand, it is generally observed that Mamdani systems suffer heavily from the curse of dimensionality because these easily interpretable rules yield less representative power than TS rules and thus more rules are needed to maintain even approximately same level of accuracy (according to many researchers [1,2], complexity is an important aspect of interpretability, therefore it may be stated that interpretability of Mamdani systems is undermined by exponentially increasing complexity requirements).

The representative power of fuzzy rules depends on how many adjustable parameters (of membership functions) are available for the consequent part of a rule. Extended representative power of fuzzy rules also makes the way the inference operators influence rule interpolation more pronounced. For somewhat uncertain reasons there are many inference operators developed for Mamdani systems (a thorough study [3] counts over 40 different operators for fuzzy implication alone) even though in practice only a handful of them (and perhaps rightfully so) have found wider use.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 91–98, 2011.

In this paper we show that application of product implication and sum aggregation for the inference function of Mamdani systems provides an analytical expression for its input/output relationship and is useful for several reasons. First, it decreases the computational overhead in computing the numerical relationship between system variables. Secondly, it permits to apply computationally efficient methods for the identification of the extended consequent parameter set. From interpretational aspect, however, we argue that some parameters in this extended set do not lend themselves to universally unambiguous interpretation. Moreover, it must be noted that in order to fully exploit the increased adaptation potential we need support from the method used for input partition determination. To conclude the paper, the role of rule weights in linguistic fuzzy modeling is discussed.

2 Preliminaries

Generally, fuzzy rules in Mamdani-type fuzzy systems are based on the disjunctive rule format

$$\begin{array}{c}\text{IF } x_1 \text{ is } A_{1r} \text{ AND } x_2 \text{ is } A_{2r} \text{ AND } \ldots \ldots \text{ AND } x_N \text{ is } A_{Nr} \text{ THEN } y \text{ is } B_r \\ \text{OR } \ldots,\end{array} \tag{1}$$

where A_{ir} denote the linguistic labels of the i-th input variable associated with the r-th rule ($i = 1,\ldots,N$), and B_r is the linguistic label of the output variable, associated with the same rule.

Each A_{ir} has its representation in the numerical domain - the membership function μ_{ir} (the same applies to B_r that is represented by γ_r) and in general case the inference function that computes the fuzzy output $F(y)$ of the system (1) has the following form

$$F(y) = \bigcup_{r=1}^{R} \left(\left(\bigcap_{i=1}^{N} \mu_{ir}(x_i) \right) \cap \gamma_r \right), \tag{2}$$

where $\cup_r^R$ denotes the aggregation operator (corresponds to OR in (1)), $\cap$ is the implication operator (THEN) and $\cap_i^N$ is the conjunction operator (AND). In order to obtain crisp output, (2) is generally defuzzified with center-of-gravity method

$$y = Y_{cog}(F(y)) = \frac{\int_Y yF(y)dy}{\int_Y F(y)dy}. \tag{3}$$

In the following, the activation degree of r-th rule - the result of the conjunction operation in (2) - is denoted as

$$\tau_r = \bigcap_{i=1}^{N} \mu_{ir}(x_i). \tag{4}$$

In a normal fuzzy system the number of membership functions (MFs) per i-th variable (S_i) is relatively small - this number is rarely equal to R as the notation style in (1) implies, moreover, it is often desired that all possible unique combinations of input MFs are represented ($R = \prod_{i=1}^{N} S_i$). MFs of the system are thus shared between the rules and a

separate $R \times N$ dimensional matrix that accommodates the identifiers $m_{ri} \in \{1,2,...,S_i\}$ maps the existing input MFs μ_i^s to the rule slots. Unless unique output MFs are assigned to all rules, they also need some external allocation mechanism (below it will be shown that there is an alternative to it).

In linguistic modeling context, Mamdani systems are subject to transparency constraints to ensure valid interpretation of fuzzy rules [4].

The most convenient way to satisfy transparency constraint ($\sum_{s=1}^{S_i} \mu_i^s(x_i) = 1$) for input MFs μ_i^s $(s = 1,...,S_i)$ is to use the following definition:

$$\mu_i^s(x_i) = \begin{cases} \frac{x_i - a_i^{s-1}}{a_i^s - a_i^{s-1}}, & a_i^{s-1} < x_i \le a_i^s \\ \frac{a_i^{s+1} - x_i}{a_i^{s+1} - a_i^s}, & a_i^s < x_i < a_i^{s+1} \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

For output MFs $\gamma_r(y)$ the respective constraint requires that

$$Y_{cog}(\gamma_r(y)) = \frac{\int_{y_{min}}^{y_{max}} y\gamma_r(y)dy}{\int_{y_{min}}^{y_{max}} \gamma_r(y)dy} = \text{core}(\gamma_r(y)), \tag{6}$$

meaning that γ_r must be symmetrical. The latter property is satisfied by default by symmetrical triangular MFs given by

$$\gamma_r(y) = \max(\min(\frac{2y - 2b_r + s_r}{2}, \frac{2b_r + s_r - 2y}{2}), 0), \tag{7}$$

where s_r is the width of γ_r and $b_r = \text{core}(\gamma_r(y))$ is its center.

3 Product-Sum Fuzzy Systems

It is easy to show that if we fix product implication and sum aggregation then (3) will be significantly reduced

$$y = \frac{\int_{y_{\min}}^{y_{\max}} \sum_{r=1}^{R} \tau_r \gamma_r(y) y dy}{\int_{y_{\min}}^{y_{\max}} \sum_{r=1}^{R} \tau_r \gamma_r(y) dy} = \frac{\sum_{r=1}^{R} \tau_r \int_{y_{\min}}^{y_{\max}} \gamma_r(y) y dy}{\sum_{r=1}^{R} \tau_r \int_{y_{\min}}^{y_{\max}} \gamma_r(y) dy} = \frac{\sum_{r=1}^{R} \tau_r C_r S_r}{\sum_{r=1}^{R} \tau_r S_r}, \tag{8}$$

where

$$C_r = \frac{\int_{y_{\min}}^{y_{\max}} \gamma_r(y) y dy}{\int_{y_{\min}}^{y_{\max}} \gamma_r(y) dy}, \tag{9}$$

is the center of gravity of given γ_r and

$$S_r = \int_{y_{\min}}^{y_{\max}} \gamma_r(y) dy, \tag{10}$$

is its area.

(8) implies that with given inference operators it is sufficient consider only the area and center of gravity of γ_r. For example, with (7), $C_r = b_r$ and $S_r = s_r/2$ and (8) can thus be rewritten as

$$y = \frac{\sum_{r=1}^{R} \tau_r b_r s_r}{\sum_{r=1}^{R} \tau_r s_r}. \tag{11}$$

Note that if γ_r are of equal width ($\forall r, s_r = \xi$, where ξ is an arbitrary positive constant) (11) reduces to

$$y = \frac{\sum_{r=1}^{R} \tau_r b_r}{\sum_{r=1}^{R} \tau_r}, \tag{12}$$

which is the inference function for well-known 0-th order Takagi-Sugeno systems, meaning, of course that if the output MFs of the fuzzy system (11) are of equal width, these can be with no loss of generalization reduced to scalars[1].

Furthermore, as $\sum_{s=1}^{S_i} \mu_i^s(x_i) = 1$ (input transparency constraint), (12) would further reduce to

$$y = \sum_{r=1}^{R} \tau_r b_r, \tag{13}$$

which is as basic as it can get.

Not only are (11) and (13) computationally less expensive than (3) they also permit us to use more efficient methods for the identification of s_r and b_r than derivative-free but computationally greedy population based guided search methods such as evolutionary algorithms that are the only viable option for the identification of Mamdani systems in general case [6].

To include the situations where the number of unique output MFs (T) is or is to be expected smaller than R (meaning that they are shared among the rules) let us introduce a $R \times T$ allocation matrix M [7] that maps t-th output MF ($t \in 1, 2, ..., T$) to the r-th rule if the element in r-th row and t-th column of M is equal to one (only one element of this value is permitted per row). If there is no sharing of output MFs ($T = R$), M is the identity matrix and can be neglected.

Using the notations

$$\Gamma = \begin{bmatrix} \tau_1(1) & \tau_2(1) & \ldots & \tau_R(1) \\ \tau_1(2) & \tau_2(2) & \ldots & \tau_R(2) \\ \ldots & \ldots & \ldots & \ldots \\ \tau_1(K) & \tau_2(K) & \ldots & \tau_R(K) \end{bmatrix}, \tag{14}$$

$$\mathbf{s}' = [s_1, s_2, ..., s_T]^T, \mathbf{b}' = [b_1, b_2, ..., b_T]^T, \tag{15}$$

and

$$\mathbf{y} = [y(1), y(2), ..., y(K)]^T, \tag{16}$$

we can show that if M, Γ, $\mathbf{s}'$ and $\mathbf{y}'$ are known, a least squares solution to (11) that lacks an exact solution in terms of $\mathbf{b}'$ is given by.

$$\mathbf{b}' = pinv(\Gamma \cdot M \cdot diag(\mathbf{s}')) \cdot diag(\Gamma \cdot M \cdot \mathbf{s}') \cdot \mathbf{y}, \tag{17}$$

where $diag()$ denotes the operation which transforms a column vector (its argument) into a diagonal matrix and where $pinv()$ is the Moore-Penrose pseudoinverse [8] that is applied for matrix inversion.

[1] An early attempt to improve tractability and computational properties of Mamdani systems [5] reduces (3) to (12) via the application of "weighted average" defuzzification - an approximate substitute of the center-of-gravity method with accounted information loss.

If M, Γ, $\mathbf{b}'$ and $\mathbf{y}$ are known and $\mathbf{s}'$ is unknown, we face the homogeneous matrix equation

$$(diag(\mathbf{y}) \cdot \Gamma \cdot M - \Gamma \cdot M \cdot diag(\mathbf{b}')) \cdot \mathbf{s}' = 0. \tag{18}$$

To find a non-trivial solution to (18), we can apply singular value decomposition [9], i.e. find matrices U, Σ and V so that

$$U \cdot \Sigma \cdot V^T = diag(\mathbf{y}) \cdot \Gamma \cdot M - \Gamma \cdot M \cdot diag(\mathbf{b}'), \tag{19}$$

where Σ is a diagonal matrix and U and V are orthogonal matrices. The least-squares solution is given by the column of V, which corresponds to the smallest diagonal entry of Σ.

For the parameter identification of (11) we can start with (17)[2] and apply (19) and (17) then repeatedly until the solution converges (less than 10 iterations are needed) or until all elements of $\mathbf{s}'$ maintain the same sign (to preserve the physical meaning of s_r).

Example: consider a simple function[3]

$$y = sin(2x - 0.7), \text{ with } x = [0, 1]. \tag{20}$$

For the identification of (20) we assume that a three-rule Mamdani system would be sufficient. We fix two input MFs at the extremes of input domain (to satisfy coverage property) and let a_1^2 take values from [0,1]. For each position the set of output parameters is identifed using (17) and (19). We also identify corresponding 0-th order TS models. The curves of modeling root mean square error (ε) in respect to the position of a_1^2 for both types of models are depicted in Fig. 1, left.

We can see that as expected, modeling error is dependent on how a_1^2 was positioned, however, a Mamdani model appears to be both more accurate and more robust than a 0-th order TS model. It is quite obvious that one additional parameter per rule offers increased adaptation potential. To exploit this potential fully, however, we need a support from the method that is used for the determination of input partition. The otherwise efficient semi-heuristic approach [10] suggests the value of a_1^2 that is denoted by a_1^{2*} in Fig. 1 and the latter seems to be more suitable for 0-th order TS systems.

Interpretation of b_r in a fuzzy system follows naturally from the definition of transparency of Mamdani systems [4] - each b_r is the value of y when $\tau_r = 1$. Interpretation of s_r, however, becomes ambiguous. As can be seen from Fig. 1, right, a larger s_r acts like a magnet that pulls the data samples between a_1^{r-1} and a_1^{r+1} toward b_r - and the value of s_r could be therefore interpreted as the "magnetic" value of a rule.

There are two reasons, though, why this (or any other) interpretation of s_r cannot be exploited very well. First, unlike b_r, the absolute value of a given s_r is meaningless, what counts is its difference from widths of output MFs of its neighboring rules (and do not forget that all s_r can be scaled up and down proportionally without any effect to

[2] For the first iteration, $\mathbf{b}'$ is identified with $\mathbf{s}'$ that is a $T \times 1$ vector of ones. If our goal is to obtain consequent parameters for a 0-th order TS system (12) then one-time application of (17) concludes the identification procedure.

[3] The goal is not to provide exhaustive demonstration of the performance of the algorithm but to illustrate some of its basic characteristics.

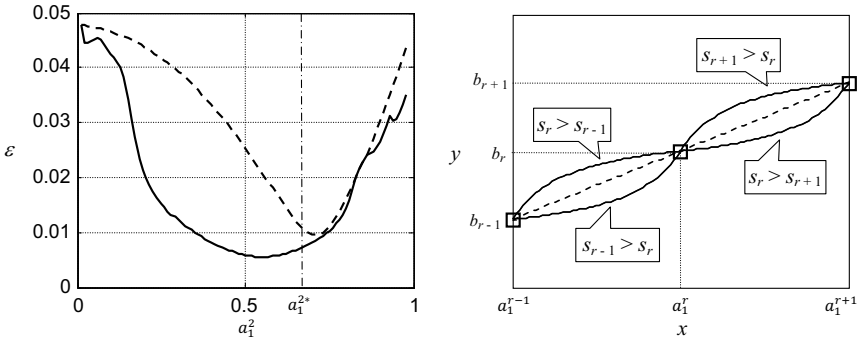


Fig. 1. Left: Dependence between the position of a_1^2 and modeling error for (20). Mamdani models (solid line), 0-th order TS models (dashed line). Right: Interpolation between neighboring rules in Mamdani systems (dashed line depicts the linear interpolation when $s_{r-1} = s_r = s_{r+1}$).

the inferred y). Secondly, each rule interacts with up to $3^N - 1$ neighboring rules (where N is the number of inputs) so we have to deal with an exponentially growing set of parameters in this comparison as the number of inputs increases and the whole effort quickly becomes impractical.

It can be also argued that if M is not the identity matrix, the increased adaptation potential of (11) will be cancelled out by possible contradictory requirements that derive from extensive sharing of consequent parameters (this no doubt influences b_r as well but to a lesser degree).

4 Weighted Fuzzy Systems

Using rule weights in fuzzy systems has been often considered as a big improvement to the way in which the rules interact [11,12,13]. Rule weights can be applied to complete rules or only to the consequent part of the rules [14]. In first case, the corresponding weigth w_r is used to modulate the activation degree of a rule and in second case it is used to modulate the rule conclusion. Consider a weighted fuzzy system ($r = 1, ..., R$)

$$\text{IF } x_1 \text{ is } A_{1r} \text{ AND } ... ... \text{ AND } x_N \text{ is } A_{Nr} \text{ THEN } y \text{ is } P_r \text{ with } w_r, \tag{21}$$

where P_r labels the output fuzzy set associated with r-th rule (having the parameters p_r and z_r) and w_r is the rule weight.

Interpretation of rule weights has long been a controversial issue [14] with many possible interpretations being tossed around (credibility, importance, influence, reliability etc.). With product-sum inference, however, the inference functions corresponding to (21) appear as

$$y = \frac{\sum_{r=1}^{R} \tau_r w_r p_r z_r}{\sum_{r=1}^{R} \tau_r w_r z_r}, \tag{22}$$

(if weights are applied to complete rules) or

$$y = \frac{\sum_{r=1}^{R} \tau_r w_r p_r z_r}{\sum_{r=1}^{R} \tau_r z_r}, \tag{23}$$

(if applied only to rule consequents). It is easy to see that both (22) and (23) are equivalents of (11), as $s_r = w_r z_r, b_r = p_r$ in first and $b_r = w_r p_r, z_r = s_r$ in second case. The only principal difference between (11) and (21) is that even if the original MFs in (21) may have been shared between the rules, the modulated MFs are now unique to each rule (as were the weights w_r, for that matter).

Similar reduction is obtained if the weights are used in multiple-consequent rules [15] in which a rule appears as

$$\begin{gathered}\text{IF } x_1 \text{ is } A_{1r} \text{ AND ... ... AND } x_N \text{ is } A_{Nr} \text{ THEN} \\ y \text{ is } P_1 \text{ with } w_1 r \text{ AND } P_2 \text{ with } w_{2r}, \text{ AND ... AND } P_T \text{ with } w_{Tr}.\end{gathered} \tag{24}$$

The inference functions corresponding to (24) and transformation formulas that reduce (24) to (11) are given in Table 1. There is no clearer way to demonstrate redundancy of rule weights.

Table 1. Transformation of multiple-consequent weighted fuzzy systems into (11)

weights applied to	inference function	b_r	s_r
rule conclusion	$y = \frac{\sum_{r=1}^{R} \tau_r \sum_{j=1}^{T} w_{jr} p_j z_j}{\sum_{r=1}^{R} \tau_r \sum_{j=1}^{T} z_j}$	$\sum_{j=1}^{T} w_{jr} p_j z_j / \sum_{j=1}^{T} z_j$	$\sum_{j=1}^{T} z_j$
complete rules	$y = \frac{\sum_{r=1}^{R} \tau_r \sum_{j=1}^{T} w_{jr} p_j z_j}{\sum_{r=1}^{R} \tau_r \sum_{j=1}^{T} w_{jr} z_j}$	$\sum_{j=1}^{T} w_{jr} p_j z_j / \sum_{j=1}^{T} w_{jr} z_j$	$\sum_{j=1}^{T} w_{jr} z_j$

5 Conclusions

At first glance, expansion of expressive power of fuzzy rules in linguistic modeling seems like a good idea, in particular, when we have an efficient identification mechanism for those extra parameters in (11) readily available. Moreover, various rule weighting schemes can be reduced to the very same configuration.

Closer look, however, reveals problems of inherent nature. For example, interpretation of those extra parameters is not intuitive at all. It is possible to ignore the interpretational aspect and treat the problematic parameters as purely computational parameters, however, in order to fully exploit increased adaptation potential we need to match the output MF identification procedure with an adequate input partition determination method. As has been suggested in the paper, finding a suitable method for that is presently a matter of further research and therefore the fuzzy system configuration (12) (0-th order TS system) still remains a very reasonable choice in linguistic modeling despite its other shortcomings.

Note that these conclusions are valid for applications of linguistic fuzzy systems in identification and control. In classification, the situation is altogether different because the output MFs of a fuzzy classifier are labels and the inference algorithm is generally different from (3). Investigation of similar issues (adaptability, rule weights etc.) in the context of classification is therefore a whole separate line of research.

References

1. Alonso, J.M., Magdalena, L.: A Conceptual Framework for Understanding a Fuzzy System. In: Joint 2009 IFSA World Congress and EUSFLAT Conference, IFSA-EUSFLAT, pp. 119–124 (2009)
2. Ishibuchi, H., Kaisho, Y., Nojima, Y.: Complexity, interpretability and explanation capability of fuzzy rule-based classifiers. In: IEEE Int. Conf. Fuzzy Systems, pp. 1730–1735. IEEE Press, New York (2009)
3. Cordon, O., Herrera, F., Peregrin, A.: Applicability of the fuzzy operators in the design of fuzzy controllers. Fuzzy Sets and Systems 86, 15–41 (1997)
4. Riid, A., Rüstern, E.: Transparent fuzzy systems and modeling with transparency protection. In: IFAC Symp. on Artificial Intelligence in Real Time Control, pp. 229–234. Elsevier Science, New York (2000)
5. Zhang, B.S., Edmunds, J.M.: On Fuzzy Logic Controllers. In: IEEE Int. Conf. Control, pp. 961–965. IEEE Press, New York (1991)
6. Cordon, O., Herrera, F., Hoffmann, F., Magdalena, L. (eds.): Genetic Fuzzy Systems. Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific, Singapore (2001)
7. Riid, A., Rüstern, E.: Interpretability Improvement of Fuzzy Systems: Reducing the Number of Unique Singletons in Zeroth order Takagi-Sugeno Systems. In: IEEE Int. Conf. Fuzzy Systems, pp. 2013–2018. IEEE Press, New York (2010)
8. Penrose, R.: A generalized inverse for matrices. Proc. Cambridge Philosophical Society 51, 406–413 (1955)
9. Golub, G.H., Kahan, W.: Calculating the singular values and pseudo-inverse of a matrix. Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis 2(2), 205–224 (1965)
10. Riid, A., Rüstern, E.: A Method for Heuristic Fuzzy Modeling in Noisy Environment. In: IEEE Int. Conf. Intelligent Systems, pp. 468–473. IEEE Press, New York (2010)
11. Pal, N.R., Pal, K.: Handling of inconsistent rules with and extended model of fuzzy reasoning. J. Intelligent and Fuzzy Syst. 7, 55–73 (1999)
12. Cho, J.S., Park, D.J.: Novel fuzzy logic control based on weighting of partially inconsistent rules using neural network. J. Intelligent and Fuzzy Syst. 8, 99–110 (2000)
13. Yu, W., Bien, Z.: Design of fuzzy logic systems with inconsistent rule base. J. Intelligent and Fuzzy Syst. 2, 147–159 (1994)
14. Nauck, D., Kruse, R.: How the Learning of Rule Weights Affects the Interpretability of Fuzzy Systems. In: IEEE Int. Conf. Fuzzy Systems, pp. 1235–1240. IEEE Press, New York (1998)
15. Ishibuchi, H., Yamamoto, T.: Rule Weight Specification in Fuzzy Rule-Based Classification Systems. IEEE Trans. Fuzzy Systems 13(4), 428–435 (2005)

A Double Axis Classification of Interpretability Measures for Linguistic Fuzzy Rule-Based Systems*

M.J. Gacto[1], R. Alcalá[2], and F. Herrera[2]

[1] Dept. Computer Science, University of Jaén, 23071 Jaén, Spain
mjgacto@ugr.es
[2] Dept. Computer Science and Artificial Intelligence,
University of Granada, 18071 Granada, Spain
{alcala,herrera}@decsai.ugr.es

Abstract. In this paper, we present a simple classification of the papers devoted to interpretability of Linguistic Fuzzy Rule-Based Systems attending to the type of interpretability measures and the part of the system for which they are applied, i.e., a double axis classification. A taxonomy considering this double axis is used to easily categorize the proposals in the existing literature. In this way, this work also represents a simple summary of the current state-of-the-art to assess the interpretability of Linguistic Fuzzy Rule-Based Systems.

1 Introduction

Linguistic fuzzy modelling, developed by linguistic Fuzzy Rule-Based Systems (FRBSs), allows us to deal with the modelling of systems by building a linguistic model which could become interpretable by human beings. Linguistic fuzzy modelling comes with two contradictory requirements [10,11]:

- *Interpretability*: This is the capability to express the behavior of the real system in an understandable way. This is a subjective property depending on a large amount of factors and still represents an open problem.
- *Accuracy*: This is the capability to faithfully represent the real system. This property represents the similarity between the responses of the real system and the fuzzy model.

In recent years the interest of researchers in obtaining more interpretable linguistic fuzzy models has grown [8,35,47], since this is still an open problem. In this paper, we present a categorization of the interpretability of fuzzy systems focused on the framework of linguistic FRBSs attending to the type of interpretability measures and the part of the system for which they are applied, i.e., a double axis classification. To this end, we will show a taxonomy with four quadrant (complexity or semantic interpretability at the level of Rule Base or

* Supported by the Spanish Ministry of Education and Science under grant no. TIN2008-06681-C06-01.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 99–106, 2011.

fuzzy partitions) with the main aim of helping readers have a good global vision of interpretability measures for linguistic FRBSs in the existing literature. This work constitutes a short version of the paper presented in [21], which includes additional complementary references in some quadrants and the detailed descriptions of the measures and approaches proposed or used in all the referred contributions.

This paper is arranged as follows. The next section presents the used taxonomy for categorizing interpretability in the linguistic FRBSs area. Section 3 summarizes the current state-of-the-art to assess the interpretability of linguistic FRBSs highlighting some facts on the different quadrants. Finally, in section 4 we draw some conclusions.

2 A Double Axis Classification of Interpretability Measures for Linguistic FRBSs

In this section, we present a specific classification that can help us better understand how the interpretability aspect has been taken into account in the particular framework of linguistic FRBSs. Different works [8,35,47] have proposed interesting taxonomies as a way to study interpretability aspects within the more general area of fuzzy systems.

In this particular case of linguistic FRBSs, the two main kinds of approaches to take into account the interpretability of linguistic FRBSs are:

1. Complexity-based Interpretability: These approaches are devoted to decreasing the complexity of the obtained model (usually measured as number of rules, variables, labels per rule, etc.).
2. Semantics-based Interpretability: These approaches are devoted to preserving the semantics associated with the Membership Functions (MFs). We can find approaches trying to ensure semantic integrity by imposing constraints on the MFs or approaches considering measures such as distinguishability, coverage, etc.

Since both kinds of measures, complexity-based interpretability and semantic-based interpretability, should be considered in both Knowledge Base (KB) components, linguistic fuzzy partition and Rule Base (RB), we will follow a taxonomy based on a double axis:

- Complexity versus Semantic interpretability.
- Rule Base versus Fuzzy Partition.

In this way, the said taxonomy comes from combining both axes. This leads to the appearance of the following quadrants devoted to analyzing the interpretability of linguistic FRBSs (see Table 1):

- Q_1: Complexity at the RB level.
- Q_2: Complexity at the fuzzy partition level.
- Q_3: Semantics at the RB level.
- Q_4: Semantics at the fuzzy partition level.

Table 1. A taxonomy to analyze the interpretability of linguistic FRBSs

	Rule Base level	Fuzzy Partition level
Complexity-based Interpretability	Q_1 Number of rules Number of conditions	Q_2 Number of Membership Functions Number of Features
Semantic-based Interpretability	Q_3 Consistency of rules Rules fired at the same time Transparency of rule structure Cointension	Q_4 Completeness or Coverage Normalization Distinguishability Complementarity Relative measures

Most of the works in the literature consider measures included inside quadrants Q_1 and Q_2 which are considered as the classic interpretability measures. These measures are widely well-known as the number of rules, number of conditions, number of features and number of MFs. This is not the case of semantic interpretability measures: there are still no widely accepted measures, because many of them appear as new measures in the literature. In the following, we present a short description of these less known measures included in the quadrants Q_3 and Q_4. Some of the more commonly used measures included in the quadrant Q_3 are the following:

- *Consistency of the RB*, is the absence of contradictory rules in RB, i.e., rules with similar premise parts should have similar consequent parts.
- *Number of rules fired at the same time*, which consists of minimizing the number of rules firing that are activated for a given input.
- *Cointension*, is a proximity of the input/output relations of the object of modeling and the model. A model is cointensive if its proximity is high.

The most common constraints or absolute measures in the quadrant Q_4 are:

1. *Completeness or Coverage*: The universe of discourse of a variable should be covered by the MFs, and every data point should belong to at least one of the fuzzy sets and have a linguistic representation, i.e., it is required that membership values should not be zero for all the linguistic variable domains.
2. *Normalization*: MFs are normal if there is at least one data point in the universe of discourse with a membership value equal to one, in respect to the maximum membership degree.
3. *Distinguishability*: An MF should represent a linguistic term with a clear semantic meaning and should be easily distinguishable from the remaining MFs of the corresponding variable.
4. *Complementarity*: For each element of the universe of discourse, the sum of all its membership values should be near to one. This guarantees a uniform distribution of the meanings among the elements.

There is an additional possibility for Q_4 which is considering relative measures, i.e., using a measure that takes into account the user MF definitions if they are available.

3 Categorizing the Current State-of-the-Art to Assess the Interpretability of Linguistic FRBSs

In order to show the existent works chronologically, Table 2 shows a summary of the works that consider the interpretability for Linguistic FRBSs grouped by quadrants and by publication date: by years and within each year by alphabetical order. This table represents a categorization of the current state-of-the art to assess the interpretability of linguistic FRBSs.

Taking into account the situation depicted in Table 2, most of the works consider measures included inside the quadrants Q_1 and Q_2 which are considered

Table 2. Summary of the current state-of-the-Art to assess the interpretability of linguistic FRBSs by years and by quadrants

				Using	Q_1		Q_2		Q_3			Q_4	
Authors	Ref.	Year	Type	MOEAs	NR	NC	Nmf	Nfeat.	Cons.	Rfired.	Coin.	Ctr.	Meas.
Ishibuchi et al.	[24,27]	1997, 1995	Clas.	√, X	√								
Pedrycz et al.	[41]	1996	Reg.										Am
Oliveira et al.	[39,40]	1999	Ctl.				√						Am
Jin et al.	[30]	1999	Ctl.						√			√	
Cheong et al.	[13]	2000	Ctl.							√		√	
Espinosa et al.	[17]	2000	Reg.				√					√	
Jin	[31]	2000	Ctl.		√		√		√			√	
Cordón et al.	[15,16]	2001	Reg.		√		√						
Ishibuchi et al.	[25]	2001	Clas.	√	√	√							
Suzuki et al.	[46]	2001	Reg.	√									Am
Guillaume et al.	[22]	2003	Clas.			√	√						
Ishibuchi et al.	[28]	2003	Reg.		√	√							
Nauck et al.	[38]	2003	Clas.			√	√						Am
Pedrycz et al.	[42]	2003	Reg.						√				
Guillaume et al.	[23]	2004	Clas.		√	√	√						Am
Ishibuchi et al.	[29]	2004	Clas.	√	√	√							
Casillas et al.	[12]	2005	Clas.		√								
Mikut et al.	[37]	2005	Clas.		√			√					
Alcalá et al.	[2]	2007	Reg.		√		√						
Alcalá et al.	[1]	2007	Reg. & Ctl.		√								
Alcalá et al.	[4]	2007	Reg.	√	√								
Cococcioni et al.	[14]	2007	Reg.	√		√							
Fazendeiro et al.	[18]	2007	Ctl.	√									Am
Ishibuchi et al.	[26]	2007	Clas.	√	√	√							
Liu et al.	[32]	2007	Clas.		√				√				
Mencar et al.	[36]	2007	Clas.									√	
Alonso et al.	[7]	2008	Clas.		√	√			√				
Pulkkinen et al.	[43]	2008	Clas.	√	√	√	√		√				Am
Pulkkinen et al.	[44]	2008	Clas.	√	√	√	√						
Alcalá et al.	[3]	2009	Reg.	√	√								
Alonso et al.	[8]	2009	Clas.		√	√			√				
Botta et al.	[9]	2009	Reg.	√									Am
Gacto et al.	[19]	2009	Reg.	√	√								
Alonso et al.	[5]	2010	Clas.	√		√				√			
Alonso et al.	[6]	in press	Clas.		√	√		√	√				
Gacto et al.	[20]	2010	Reg.	√	√								Rm
Márquez et al.	[33]	2010	Reg.	√	√					√			
Mencar et al.	[34]	in press	Clas.								√	√	
Pulkkinen et al.	[45]	2010	Reg.	√		√						√	

NR = Number of Rules, NC = Number of Conditions, Nmf = Number of membership functions, Nfeat. = Number of Features, Cons. = Consistency, Rfired. = Number of Rules fired at the same time, Coin. = Cointension, Ctr. = Constraints, Meas. = Measures; Clas. = Classification, Reg. = Regression, Ctl. = Control, Am = Absolute Measures, Rm = Relative Measures.

as the classic interpretability measures. The number of rules in the quadrant Q_1 is one of the more used measures in the literature, for which reason it is possible to consider it as a good measure of complexity at the rule base level. However the total number of conditions seems to be a more complete way since it can consider both the length of the rules and the number of rules, in an simple measure. In the quadrant Q_2 most of the works simply impose restrictions on the maximum number of MFs allowed, even though, depending on the problem (particularly in high dimensional problems) decreasing the number of features should be preferred. The large quantity of works published in 2003 was motivated by the following two books [10,11] on the interpretability-accuracy trade-off in the field of FRBSs. However, apart from this, the interest of researchers has increased particularly from 2007, giving rise to the appearance of many works from this year to the present.

Inside quadrant Q_3 there are a few measures but some works propose promising measures such as the consistency of the rules and the more recent number of rules fired at the same time and cointension. In quadrant Q_4 there are a lot of works imposing constraints. However, recently new absolute or relative semantic interpretability measures have arisen, which are more suitable to be taken into account for optimization processes. For Q_3 and Q_4 quadrants, there are still no widely accepted measures, which will arise with their use as happened with the number of rules for Q_1. Additionally, the use of Multi-Objective Evolutionary Algorithms (MOEAs) has emerged as a good way to handle interpretability since they allow both complexity and semantic interpretability measures to be optimized together by also taking into account the accuracy of the model.

4 Conclusions

In this work, we have presented a categorization of the interpretability of fuzzy systems focused on the framework of linguistic FRBSs. To this end, we have shown a taxonomy with four quadrants (complexity or semantic interpretability at the level of RB or fuzzy partitions) as a way of organizing the different measures or constraints that we can find in the literature to control interpretability of linguistic FRBSs.

The interpretability of linguistic FRBSs is still an open problem and our aim by presenting this categorization of the works in the existing literature is to help researchers in this field to propose the most appropriate measure depending on the part of the KB in which they want to maintain/improve interpretability. Additionally, this work represents exhaustive list of papers devoted to assess the interpretability of linguistic FRBSs.

References

1. Alcalá, R., Alcalá-Fdez, J., Herrera, F.: A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. IEEE Trans. Fuzzy Syst. 15(4), 616–635 (2007)

2. Alcalá, R., Alcalá-Fdez, J., Herrera, F., Otero, J.: Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation. International Journal of Approximate Reasoning 44(1), 45–64 (2007)
3. Alcalá, R., Ducange, P., Herrera, F., Lazzerini, B., Marcelloni, F.: A multi-objective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy rule-based systems. IEEE Transactions on Fuzzy Systems 17(5), 1106–1122 (2009)
4. Alcalá, R., Gacto, M.J., Herrera, F., Alcalá-Fdez, J.: A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15(5), 539–557 (2007)
5. Alonso, J., Magdalena, L., Cordón, O.: Embedding HILK in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers. In: 4th International Workshop on Genetic and Evolving Fuzzy Systems (GEFS 2010), pp. 15–20. IEEE, Mieres (2010)
6. Alonso, J.M., Magdalena, L.: Hilk++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. Soft Computing (2011) (in press) doi:10.1007/s00500-010-0628-5
7. 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(7), 761–794 (2008)
8. Alonso, J.M., Magdalena, L., Rodríguez, G.G.: Looking for a good fuzzy system interpretability index: An experimental approach. International Journal of Approximate Reasoning 51, 115–134 (2009)
9. Botta, A., Lazzerini, B., Marcelloni, F., Stefanescu, D.C.: Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. Soft Computing 13(5), 437–449 (2009)
10. Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.): Accuracy improvements in linguistic fuzzy modeling. Studies in Fuzziness and Soft Computing, vol. 129. Springer, Heidelberg (2003)
11. Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.): Interpretability issues in fuzzy modeling. Studies in Fuzziness and Soft Computing, vol. 128. Springer, Heidelberg (2003)
12. Casillas, J., Cordón, O., del Jesus, M.J., Herrera, F.: Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. IEEE Trans. Fuzzy Syst. 13(1), 13–29 (2005)
13. Cheong, F., Lai, R.: Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm. IEEE Trans. Syst., Man, Cybern. - Part B: Cybernetics 30(1), 31–46 (2000)
14. Cococcioni, M., Ducange, P., Lazzerini, B., Marcelloni, F.: A pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems. Soft Computing 11, 1013–1031 (2007)
15. Cordón, O., Herrera, F., Magdalena, L., Villar, P.: A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. Information Science 136, 85–107 (2001)
16. Cordón, O., Herrera, F., Villar, P.: Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base. IEEE Transactions on Fuzzy Systems 9(4), 667–674 (2001)
17. Espinosa, J., Vandewalle, J.: Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. IEEE Trans. Fuzzy Syst. 8(5), 591–600 (2000)

18. Fazendeiro, P., de Oliveira, J.V., Pedrycz, W.: A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller. IEEE Trans. on Biomedical Engineering 54(9), 1667–1678 (2007)
19. Gacto, M.J., Alcalá, R., Herrera, F.: Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. Soft Computing 13(5), 419–436 (2009)
20. Gacto, M.J., Alcalá, R., Herrera, F.: Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems. IEEE Transactions on Fuzzy Systems 18(3), 515–531 (2010)
21. Gacto, M., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences (2011) (in press) doi: 10.1016/j.ins.2011.02.021
22. Guillaume, S., Charnomordic, B.: A new method for inducing a set of interpretable fuzzy partitions and fuzzy inference systems from data. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.) Interpretability issues in fuzzy modeling, pp. 148–175. Springer, Heidelberg (2003)
23. Guillaume, S., Charnomordic, B.: Generating an interpretable family of fuzzy partitions from data. IEEE Transactions on Fuzzy Systems 12(3), 324–335 (2004)
24. Ishibuchi, H., Murata, T., Türksen, I.B.: Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets and Systems 89(2), 135–150 (1997)
25. Ishibuchi, H., Nakashima, T., Murata, T.: Three-objective genetics-based machine learning for linguistic rule extraction. Information Sciences 136, 109–133 (2001)
26. Ishibuchi, H., Nojima, Y.: Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. International Journal of Approximate Reasoning 44(1), 4–31 (2007)
27. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans. Fuzzy Syst. 3(3), 260–270 (1995)
28. Ishibuchi, H., Yamamoto, T.: Interpretability issues in fuzzy genetics-based machine learning for linguistic modelling. In: Lawry, J., Shanahan, J.G., Ralescu, A.L. (eds.) Modelling with Words. LNCS, vol. 2873, pp. 209–228. Springer, Heidelberg (2003)
29. Ishibuchi, H., Yamamoto, T.: Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets and Systems 141(1), 59–88 (2004)
30. Jin, Y., von Seelen, W., Sendhoff, B.: On generating FC^3 fuzzy rule systems from data using evolution strategies. IEEE Transactions on Systems, Man, and Cybernetics 29(6), 829–845 (1999)
31. Jin, Y.: Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement. IEEE Transactions on Fuzzy Systems 8(2), 212–221 (2000)
32. Liu, F., Quek, C., Ng, G.S.: A novel generic hebbian ordering-based fuzzy rule base reduction approach to Mamdani neuro-fuzzy system. Neural Computation 19, 1656–1680 (2007)
33. Marquez, A., Márquez, F., Peregrin, A.: A multi-objective evolutionary algorithm with an interpretability improvement mechanism for linguistic fuzzy systems with adaptive defuzzification. In: IEEE World Congress on Computational Intelligence, pp. 277–283 (2010)

34. Mencar, C., Castiello, C., Cannone, R., Fanelli, A.M.: Interpretability assessment of fuzzy knowledge bases: a cointension based approach. International Journal of Approximate Reasoning (2011) (in press) doi: 10.1016/j.ijar.2010.11.007
35. Mencar, C., Fanelli, A.: Interpretability constraints for fuzzy information granulation. Information Sciences 178(24), 4585–4618 (2008)
36. Mencar, C., Castellano, G., Fanelli, A.M.: Distinguishability quantification of fuzzy sets. Information Sciences 177, 130–149 (2007)
37. Mikut, R., Jakel, J., Grall, L.: Interpretability issues in data-based learning of fuzzy systems. Fuzzy Sets and Systems 150, 179–197 (2005)
38. Nauck, D.: Measuring interpretability in rule-based classification systems. In: Proc. of the 12th IEEE International Conference on Fuzzy Systems, vol. 1, pp. 196–201 (2003)
39. de Oliveira, J.V.: Semantic constraints for membership function optimization. IEEE Trans. Syst., Man, Cybern. - Part A: Systems and Humans 29(1), 128–138 (1999)
40. de Oliveira, J.V.: Towards neuro-linguistic modeling: constraints for optimization of membership functions. Fuzzy Sets and Systems 106(3), 357–380 (1999)
41. Pedrycz, W., de Oliveira, J.V.: Optimization of fuzzy models. IEEE Trans. Systems Man Cybernet. Part B 26(4), 627–636 (1996)
42. Pedrycz, W.: Expressing relevance interpretability and accuracy of rule-based systems. In: Casillas, J., Cordón, O., Herrera, F., Magdalena, L. (eds.) Interpretability issues in fuzzy modeling, pp. 546–567. Springer, Heidelberg (2003)
43. Pulkkinen, P., Hytönen, J., Koivisto, H.: Developing a bioaerosol detector using hybrid genetic fuzzy systems. Engineering Applications of Artificial Intelligence 21(8), 1330–1346 (2008)
44. Pulkkinen, P., Koivisto, H.: Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. International Journal of Approximate Reasoning 48(2), 526–543 (2008)
45. Pulkkinen, P., Koivisto, H.: A dynamically constrained multiobjective genetic fuzzy system for regression problems. IEEE Transactions on Fuzzy Systems 18(1), 161–177 (2010)
46. Suzuki, T., Kodama, T., Furuhashi, T., Tsutsui, H.: Fuzzy modeling using genetic algorithms with fuzzy entropy as conciseness measure. Information Sciences 136(1-4), 53–67 (2001)
47. Zhou, S.M., Gan, J.Q.: Low-level interpretability and high-level interpretability: A unified view of data-driven interpretable fuzzy system modelling. Fuzzy Sets and Systems 159(23), 3091–3131 (2008)

Tagging Ontologies with Fuzzy WordNet Domains

Angela Locoro

Computer Science Department, University of Genova, Via Dodecaneso 35, 16146 Genova, Italy
Angela.Locoro@unige.it

Abstract. The use of WordNet Domains is confined in the present days to Text Mining field. Moreover, the tagging of WordNet synsets with WordNet Domain labels is a crisp one. This paper introduces an approach for automatically tagging both ontologies and their concepts with WordNet domains in a fuzzy fashion, for topic classification purposes. Our fuzzy WordNet Domains model is presented as well as our domain disambiguation procedure. Experiments show promising results and are introduced in this paper as well as a final discussion on envisioned scenarios for our approach.

1 Introduction

When dealing with short texts ontologies are the only viable solution to extract machine readable semantics, whereas statistical and other structural approaches are unfeasible if a massive amount of information is not available.

The number of ontologies available online is facing a swift growth thanks to the advent of the Semantic Web, Semantic Web Services and Multi Agent Systems era. Universities, domain experts communities, research groups all endeavour to make their conceptualisation efforts available online. Ontology Repositories and libraries, general purpose semantic search engines as well as specific-purpose networks of ontologies aim at fostering the reuse of widely shared, well formalised and maintained knowledge representation standards and resources. It is well known that ontologies may formalise the world knowledge from different granularities and perspectives. Conceptualisations spread across upper-domains layers, specific-domains descriptions as well as parts of them, with the purpose of keeping things as simple and well focused as possible.

Projects for providing ontologies with metadata for easy description, maintenance, storage and retrieval [4] are based on human experts supervision[1]. Some efforts have already been made in the direction of annotating ontological content with lexical knowledge (see for example the mapping provided between the SUMO ontology and WordNet synsets [7]) and of investigating the interplay between lexical semantics and formal conceptualisations as a means to harness linguistic models and powerful reasoning services coming from both computational linguistics and ontological theories [8].

In this paper we report a preliminary work on automatically tagging ontologies with WordNet Domains in a fuzzy fashion, in order to determine the linguistic domains of one ontology and of its concepts. Stemming from an assignment of domain labels to each concept of an ontology and applying our fuzzy computational model, we were able to assign membership values to domain labels and propagate it to the whole ontology.

[1] http://www.semanticdesktop.org/ontologies/nao/

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 107–114, 2011.

The paper is organised as follows: Section 2 introduces WordNet Domains and its use in the Computational Linguistic field. Section 3 depicts our approach in details while Section 4 reports the experiments and results of our procedure. A final discussion on application scenarios in reported in Section 5.

2 Background

WordNet Domains[2] is a project developed by the Fondazione Bruno Kessler (FBK), Trento [5,2]. The project has the goal of better characterising a word meaning within its use in the language and in texts, hence reducing ambiguity.

The operation of tagging WordNet synsets with domain labels has been conducted by the FBK team partially by hand (by annotating a subset of synsets) and by automatically extending such labels to related synsets through the WordNet hierarchy, fixing the automatic procedure with corrections where necessary. The domain labels assigned to WordNet synsets are those of the Dewey Decimal Classification system[3], a standard largely adopted by library systems.

Gliozzo and Strapparava [1,3] propose to use a Domain Model based on WordNet Domains and exploit such model for a Word Sense Disambiguation methodology named "Domain Driven Disambiguation (DDD)" [6].

2.1 Create the WordNet Domains Ontology

As a first step of our approach we take the WordNet Domains taxonomy[4] and encode it in OWL using the Protégé tool. The *WNDO* ontology result in 160 domain labels concepts divided as follows:

- 11 top level concepts which represent the upper layer of domains classification. They are: *applied_science*, *doctrines*, *factotum*, *free_time*, *metrology*, *number*, *person*, *pure _science*, *quality*, *social_science*, *time_period*;
- 42 mid level concepts that are used to tag synsets representing concepts used at the more general level of a super domain (e.g. *medicine*, *economy*, *sport*, and so on);
- 107 low level concepts that are subclasses of one of the 42 mid level concepts or belong to a further level of specialisation and are also used to tag synsets, which are relative to concepts used in more specialised domains (e.g. *psychiatry*, *banking*, *athletics* and so on);

The *factotum* label applies for all those WordNet synsets whose usage in the language spread over different linguistic domains and hence we exclude it from our domain disambiguation results (see Sect. 3.3 for details).

2.2 Amend Discrepancies

While comparing results from the computation of all WordNet 3.0 synsets domain labels with the official WordNet taxonomy we have noticed that there are differences

[2] http://wndomains.fbk.eu/index.html

[3] http://www.oclc.org/dewey/versions/default.htm

[4] Available at http://wndomains.fbk.eu/hierarchy.html, Last accessed on 3rd January 2010.

between them. To avoid inconsistencies we have translated, where possible, those domain labels not available in *WNDO* into the labels available in it (for example the label *psychological_features* could be easily reconducted to *psychology* and *graphics_arts* to the super domain of *arts*) and where not possible we simply deleted the label itself (for example, we are not aware of the exact position in the hierarchy of the new label *animal_husbandry* - is it just under *animal* label? Or is it at the same level?).

3 Tagging Ontologies with Fuzzy WordNet Domains Labels

3.1 Motivations

WordNet Domains labels are assigned to a WordNet synset according to a crisp approach. This prevents from capturing the underlying degree of membership of a word to a domain of discourse, especially in case a word is tagged with more than one domain label. WordNet Domains are structured in a taxonomy and membership values of a word should also reflect its relations to a grouped set of domains under a more general one. Moreover, a classification model should express the fact that a membership value can dinamically change as the contexts (e.g. ontologies) in which words appear change. A fuzzy domain model like the one we propose tries to explicit all of these aspects of instrinsic vagueness that are characterising the problem at hand.

3.2 A Fuzzy WordNet Domains Model for Ontology Tagging and Classification

A tagging activity consists in associating WordNet domains labels with words in a document. This process results in the creation of a list of domains labels for each document of the form:

$$L_r = [d_1, d_2, d_2, ..., d_n] \ . \tag{1}$$

The list may contain duplicate entries, as more words in the document may be tagged with the same domain label, as exemplified by the repeated d_2 entry in L_r.

According to Zadeh's theory on fuzzy sets [9], in our model we create the fuzzy set D_r from the list L_r, by avoiding duplicate entries in the list, defined as

$$D_r = \{d_1, d_2, ..., d_n\} \tag{2}$$

with membership function

$$\mu_{\mathcal{D}_\mathcal{R}} : D_r \mapsto [0, 1], \forall d_i \in D, i = 1, ..., n \tag{3}$$

defined as follows

$$\mu_{\mathcal{D}_\mathcal{R}}(d_i) = \frac{count_{L_r}(d_i)}{\mid L_r \mid} \ . \tag{4}$$

The membership value for each d_i is computed as the ratio between the number of occurrences of label d_i in the list L_r and the total number of labels in the whole list L_r.

As a specialisation of the procedure above we define the activity of tagging concepts from an ontology O with the concepts of the ontology *WNDO* (that represents the WordNet domain labels in a hierarchical fashion, as depicted in Sect. 2.1) as the

creation of a list L_O of domain labels and of the fuzzy set D_O obtained as in (2), with membership function:

$$\mu_{\mathcal{D}_\mathcal{O}} : D_O \mapsto [0,1], \forall do_x \in L_O, x = 1, ..., m \tag{5}$$

calculated as in (4) over the list L_O of domains labels associated with concepts in ontology O.
Our tagging procedure includes an inference-assignment step from domain label $do_x \in$ *WNDO* assigned to a concept in ontology O to its direct super-class domain label $do_{x_{super}} \in$ *WNDO*, which is also assigned to the concept (see next Section for details). The list L_O will hence contain not only the list of domain labels analogous to the one resulting from (1), but also as many occurrences of $do_{x_{super}}$ as do_x ones.

The membership value of $do_{x_{super}}$ is then constrained by the following condition that holds for the membership function of (5):

$$\forall do_x \sqsubseteq do_{x_{super}} : \sum \mu_{\mathcal{D}_\mathcal{O}}(do_x) \leq \mu_{\mathcal{D}_\mathcal{O}}(do_{x_{super}}) \ . \tag{6}$$

The results of our tagging procedure allow the fuzzy classification of ontology O at concept level as well as at ontology level. Moreover, the hierarchical model of domain labels in the *WNDO* ontology and the property stated in (6) enables to extract the most relevant domain according to its level in the hierarchy.

Let $conc(\textit{WNDO})$ be the set of concepts representing the domain labels in *WNDO*, and let $topLevC(\textit{WNDO})$, $midLevC(\textit{WNDO})$ and $lowLevC(\textit{WNDO})$ be the set of top level, mid level and bottom level concepts respectively, representing a partition of $conc(\textit{WNDO})$, such that

$$topLevC(\textit{WNDO}) \cap midLevC(\textit{WNDO}) \cap lowLevC(\textit{WNDO}) = \varnothing \tag{7}$$

$$conc(\textit{WNDO}) = topLevC(\textit{WNDO}) \cup midLevC(\textit{WNDO}) \cup lowLevC(\textit{WNDO}) \ . \tag{8}$$

In order to extract the most relevant top level domain label d for any ontology O we use the formula:

$$winnerTopLev_O = max_{d \in topLevC(\textit{WNDO})}(\mu_{\mathcal{D}_\mathcal{O}}(d)) \tag{9}$$

whereas to extract the most relevant mid level domain label d for any ontology O we use the formula:

$$winnerMidLev_O = max_{d \in midLevC(\textit{WNDO})}(\mu_{\mathcal{D}_\mathcal{O}}(d)) \ . \tag{10}$$

To extract the most relevant top and mid level domain labels d_{c_z} for any ontology concept $c_z \in O$ we may use the fuzzy subset of D_O that contains all the domain labels associated with c_z, and the following formulas:

$$winnerTopLev_{O_{c_z}} = max_{d_{c_z} \in topLevC(\textit{WNDO})}(\mu_{\mathcal{D}_\mathcal{O}}(d_{c_z})) \tag{11}$$

and

$$winnerMidLev_{O_{c_z}} = max_{d_{c_z} \in midLevC(\textit{WNDO})}(\mu_{\mathcal{D}_\mathcal{O}}(d_{c_z})) \ . \tag{12}$$

3.3 Automatic Domain Disambiguation of Ontologies

For each concept from one ontology we perform the following pre-processing operation:

- tokenise, eliminate stop words and lemmatise the concepts in order to obtain canonical word forms either from composite concepts or derivationally formed ones. To obtain the lemma (or the set of lemmas if the concept is composite) from one concept we use WordNet 3.0;
- look into "WordNet Domains - WordNet 3.0 synsets" mappings files (WnToWnD)[5] and assign to each ontology concept lemma (or the resulting sets of lemmas if the concept is composite) its domain labels as they result from the mappings file;
- if one lemma does not have any domain labels, we tag it with the domain labels of its super-concepts, if they exist. Figure 1 shows the inference procedure starting from the ontology hierarchy and producing a domain label for the sub-class of a tagged super-class;

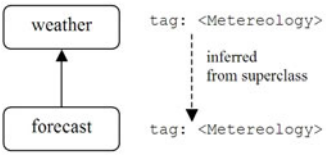


Fig. 1. Inferring a domain tag for class *Forecast* from the super-class *Weather*

- we do the necessary domain labels translations (see Sect. 2.2);
- we add the super-domain labels of domain labels assigned, looking at the WordNet Domains ontology hierarchy (through an inference procedure on the whole domain hierarchy space). Figure 2 exemplifies the procedure for the two concepts *Weather* and *Forecast*;

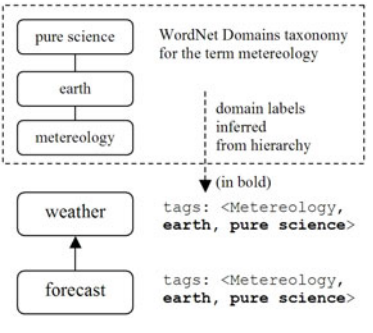


Fig. 2. Inferring super domain tags for class *Forecast* from the super-class *Weather*

[5] We recall that the WordNet Domain version we used is 3.1. For conversion from WordNet 2.0 to WordNet 3.0 synsets we use mappings files available at http://www.lsi.upc.es/~nlp/tools/mapping.html

- we instantiate our computational fuzzy model on the whole set of labels assigned as above and obtain the fuzzy domain tagging for each concept and for each ontology.
- we extract classification results. If the maximum membership value is assigned to the *factotum* domain, we pick up the second higher membership value. If we want to visit the sub-domains and take the best of them with highest membership value we can again combine our fuzzy WordNet Domains model with *WNDO* hierarchy and obtain the most specific winner domains for each c_i (and for the whole ontology respectively).

4 Experiments and Results

We conducted our tests by running the pre-processing procedure for 17 ontologies codified in OWL and randomly chosen from the Swoogle portal[6]. It is worth noting that each class of each ontology turned out to have a label or a super-class with a label correctly assigned. The total number of classes that missed a domain tag is 71 over the 17 ontologies, namely the 3,7% of the total number of concepts. They were all successfully tagged through their super-classes. The number of concepts for each ontologies goes from a minimum of 19 concepts (the Vertebrate ontology) to a maximum of 349 (MPEG7 ontology).

Table 1 shows, for each ontology, the results of the automatic domain disambiguation procedure. In particular we report the winner top level domain and its direct subdomains (mid level), according to the *WNDO* hierarchy and to our fuzzy computational model applied for each ontology. If the highest membership value for one ontology results to be assigned to the *factotum* label we have considered the second higher membership value.

A general consideration that holds for the results is that the top level domain labels correctly reflect the topic description[7] of the most part of the ontologies. With very few exceptions, mid level domain labels were able to better reflect the ontology description by identifying its more specialistic domains. From our experiments we notice that low level domains are less meaningful for our classification purpose. Most results show that mid level domain labels with highest membership value are related to their top level labels (except for two cases, MPEG7 and Space ontologies, whose mid level labels are indicated in italic in the table, that are characterised by concepts belonging to disparate domains and or whose concepts are associated with the factotum domain, respectively). As a result of our experiments we may state that mid level domain labels of *WNDO* are the best candidate to properly classify ontologies in a domain disambiguation task such as the one we presented.

[6] Experiments are conducted on a notebook Toshiba, with Windows 7 64-bit OS, Processor Intel Core i5 2.27 GHz, 4 GB RAM. The time complexity of the approach has been estimated for the ontology with the fewest number of concepts (15) as well as for the ontology with the greatest number of concepts (349) and is of 3 seconds and 7 seconds respectively.

[7] Ontologies and their manual topic descriptions are available at `http://www.disi.unige.it/people/LocoroA/download/wilfontologies/`

Table 1. Domain labels automatically assigned for each ontology with highest membership values

Ontology	**Automatic Domain Classification**	
	top level domain labels	mid level domain labels
Agent	social_science	pedagogy
Bibtex	social_science	publishing
Biosphere	pure_science	biology
Ecology	pure_science	biology
Food	applied_science	alimentation
Geofile	social_science	politics
HL7_RBAC	social_science	commerce
Ka	social_science	pedagogy
MPEG7	social_science	*sport*
Restaurant	applied_science	alimentation
Resume	social_science	economy
Space	social_science	*earth*
Subject	doctrines	literature
Top-bio	pure_science	chemistry
Travel	social_science	transport
Vacation	social_science	economy
Vertebrate	pure_science	biology

5 Discussion on Scenarios

In this paper we have provided a procedure for the automatic domain disambiguation of one ontology and its concepts by means of Fuzzy WordNet Domains tagging. Promising results enable us to draw some application scenarios of our approach in the next paragraphs.

Ontology Classification and Retrieval. The Semantic Web relies on ontologies that describe contents and resources from different perspectives and with different granularities. Knowing the domain of ontologies in advance may provide agents and systems, in charge of storing and retrieving such knowledge, with search and classification hints for identifying domain-related ontologies hence giving an idea of what exist for a particular topic of discourse and at which levels of detail it may be provided.

Ontology Matching. During the phase of meaning negotiation, organising and managing semantic interoperability through schema translation, matching or query expansion may rely on domain knowledge. For example if two ontologies have the same domain labels a system may decide that they can be matched or otherwise it may decide that matching them would result in a useless time consuming effort. The domains assigned at a concept level may provide an information to be used for disambiguating concepts in combination with other semantic features in order to refine matching algorithms.

Ontology Analysis for Engineering and Evolution. In some real scenarios ontologies are used to properly classify users' contents. In such realms knowledge evolves and

domain ontologies have to be maintained and constantly fed with unstructured knowledge. A linguistic analysis of the entities chosen to conceptualise a domain may help in verifying if a linguistic domain balance has been reached by terms used for defining concepts and hence it may empower the ontology expressivity, from the point of view of the language use, as well as reducing the ambiguities towards correct and complete domains description.

Acknowledgments. I would like to thank my colleagues Viviana Mascardi and Stefano Rovetta for their precious advice on the modelling phase of WordNet domains and for their careful reading of my work. This work was partly supported by the Italian research project Iniziativa Software CINI-FinMeccanica.

References

1. Agirre, E., Edmonds, P.: Word Sense Disambiguation - Algorithms and Applications. Springer, Heidelberg (2007)
2. Bentivogli, L., Forner, P., Magnini, B., Pianta, E.: Revising WordNet domains hierarchy: Semantics, coverage, and balancing. In: Proc. of the 21st International Conference on Computational Linguistics (COLING 2004), pp. 101–108 (2004)
3. Gliozzo, A., Strapparava, C.: Semantic Domains in Computational Linguistics. Springer, Heidelberg (2009)
4. Hartmann, J., Palma, R., Sure, Y.: Omv - ontology metadata vocabulary. In: ISWC 2005 Workshop on Ontology Patterns for the Semantic Web (2005)
5. Magnini, B., Cavagliá, G.: Integrating Subject Field Codes into WordNet. In: Gavrilidou, M., Crayannis, G., Markantonatu, S., Piperidis, S., Stainhaouer, G. (eds.) Proc. of the 2nd International Conference on Language Resources and Evaluation (LREC-2000), pp. 1413–1418 (2000)
6. Magnini, B., Strapparava, C., Pezzulo, G., Gliozzo, A.: The role of domain information in word sense disambiguation. Natural Language Engineering 8(4), 359–373 (2002)
7. Niles, I., Pease, A.: Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology. In: Arabnia, H.R. (ed.) IKE, pp. 412–416. CSREA Press (2003)
8. Oltramari, A., Vetere, G., Lenzerini, M., Gangemi, A., Guarino, N.: Senso comune. In: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (eds.) LREC, European Language Resources Association (2010)
9. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)

On the Notions of Residuated-Based Coherence and Bilattice-Based Consistence

Carlos V. Damásio[1], Nicolás Madrid[2], and M. Ojeda-Aciego[2]

[1] CENTRIA. Universidade Nova de Lisboa. Portugal
cd@di.fct.unl.pt
[2] Dept. Matemática Aplicada. Universidad de Málaga. Spain*
{nmadrid,aciego}@ctima.uma.es

Abstract. Different notions of coherence and consistence have been proposed in the literature on fuzzy systems. In this work we focus on the relationship between some of the approaches developed, on the one hand, based on residuated lattices and, on the other hand, based on the theory of bilattices.

1 Introduction

Although inconsistency is essentially considered as an undesirable feature, it arises naturally when considering databases and, in many cases, seems to be unavoidable. As a result, some efforts have been made in order to develop some mechanism to tolerate inconsistent information.

Paraconsistent logics were introduced several decades ago as an *inconsistency-tolerant approach* which allows for efficiently handling inconsistent information. Among the different approaches in the literature, we emphasize the approaches related to *consistence restoring* [2,3], *fix-point semantics* [9,16] and *inconsistent information measuring* [11,15].

It is noticeable that there is not a consensus on the notion of inconsistency in the fuzzy logic framework: one approach, given in [7], considers that a knowledge base is potentially inconsistent, or incoherent, if there exists a piece of input data that respects integrity constraints and leads to logical inconsistency when added to the knowledge base; in [16] the authors consider the problem of revising extended programs, and base their approach on the coherence theory initially advocated by Gardenfors for belief revision.

Our contribution in this work is based on two additional approaches, previously developed separately by the authors, the notion of coherence [14] in residuated logic programming [5] and the notions of consistence on a paraconsistent extension of logic programming [1].

2 Preliminary Definitions

In order to make this paper self-contained, the notions of residuated-based coherence and of bilattice-based consistence are recalled here.

* Partially supported by Spanish Ministry of Science project TIN09-14562-C05-01 and Junta de Andalucía project P09-FQM-5233.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 115–122, 2011.

2.1 *L*-Interpretations and Coherence

Definition 1. *A* residuated lattice with negation *is a tuple* $\mathcal{L} = (L, \leq, *, \leftarrow, n)$ *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* $y * z \leq x \quad \forall x, y, z \in L$.
4. n *is an antitonic mapping* $n : L \to L$ *satisfying* $n(0) = 1$ *and* $n(1) = 0$.

The operator n in a residuated lattice with negation $\mathcal{L} = (L, \leq, *, \leftarrow, n)$, is called the *negation operator* of $\mathcal{L}$. Let us define now the syntax of our logic. Let Π be a set of propositional symbols, then the set of *well-formed formulas* is defined inductively as follows:

- every propositional symbol is a well-formed formula (wff).
- if p is a propositional symbol, then $\sim p$ is a wff.
- if ϕ and ψ are wffs then $\neg\phi, \phi * \psi$ and $\phi \leftarrow \psi$ are wffs.

Note that we use four propositional connectives; $*$ to represent the conjunction, $\leftarrow$ to represent the implication, $\sim$ to represent the strong negation and $\neg$ to represent the default negation. The use of these two different negations is highly motivated by the fuzzy answer set semantics [13].

Definition 2. *A literal* ℓ *is either a propositional symbol* p *or a propositional symbol negated by the strong negation* $\sim p$. *The set of literals is denoted by Lit.*

Let us describe the semantics for the syntax introduced previously.

Definition 3. *Let* $\mathcal{L} = (L, \leq, *, \leftarrow, n)$ *be a residuated lattice with negation, an L-interpretation is a mapping* I: $Lit \to L$. *The domain of each L-interpretation can be inductively extended to non-literal wffs as follows:*

- *if* ϕ *and* ψ *are wff then*
  - $I(\neg\phi) = n(\phi)$
  - $I(\phi * \psi) = I(\phi) * I(\psi)$
  - $I(\phi \leftarrow \psi) = I(\phi) \leftarrow I(\psi)$

It is important to point out the semantical difference between strong and default negation in this logic framework. The semantics is compositional (i.e the truth value of $\neg p$ depends univocally of the truth value of p) with respect to default negation but not necessarily with respect to the strong negation (i.e the truth values of p and $\sim p$ are, a priori, independent). As a result, it might happen that the truth-value of two opposite literals, which are assigned directly by one L-interpretation, represent contradictory information and consider the possibility of rejecting those cases; for instance, in classical logic programming inconsistent interpretations are rejected. In [13,14], we introduced the notion of *coherence* as a suitable generalization of consistence in the residuated framework.

Definition 4. *Let* $\mathcal{L} = (L, \leq, *, \leftarrow, n)$ *be a residuated lattice with negation, an L-interpretation is* coherent *if* $I(\sim p) \leq n(I(p))$ *holds for every propositional symbol* $p \in \Pi$.

It is remarkable that the formula $\neg p \leftarrow \sim p$ has the value 1 as truth-value with respect to every coherent L-interpretation; since $I(p \leftarrow q) = 1$ if and only if $I(p) \geq I(q)$ holds in every residuated lattice [12]. The formula above states a relationship between the two types of negation, specifically, it states that *strong negation implies default negation.*

2.2 Bilattices and Consistence

Other approaches to deal with default negation and consistency are based on the notion of bilattice, instead of on a residuated lattice with negation as in the previous section. This is due to the fact that bilattices provide a natural framework in which one can define the notions of *consistence* and *default negation.* For instance, [1] proposed a framework which extends a previous approach to generalized logic programming to an arbitrary complete bilattice of truth-values, where belief and doubt are explicitly represented, as well as a precise definition of important operators found in logic programming. Furthermore, bilattices have been widely used as useful tools to deal with incomplete and/or inconsistent information [4,6].

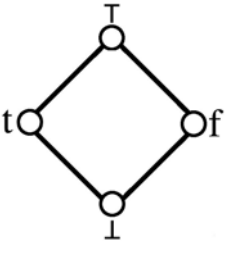
Definition 5. *A* bilattice *is a tuple* $\mathcal{B} = (B, \leq_t, \leq_k)$ *where* B *is a nonempty set, and* $(B, \leq_t)$ *and* $(B, \leq_k)$ *are both bounded lattices.*

Note that the subscript in the ordering relations occurring in the definition stands for *truth* and for *knowledge* as this will be their underlying meaning.

Example 1. A typical example of bilattice is $\mathcal{FOUR}$, which is commonly used in many paraconsistent logics. This bilattice is defined over the set $\{\bot, \mathbf{t}, \mathbf{f}, \top\}$ by considering the ordering $\leq_k$ and $\leq_t$ defined below. These four values are interpreted frequently as follows: $\bot$ by *unknown*; $\mathbf{f}$ by *false*; $\mathbf{t}$ by *true*; and $\top$ by *inconsistent.* $\mathcal{FOUR}$ is usually represented by the following diagram:

$$x \leq_k y \iff \begin{cases} x = \bot \\ y = \top \\ x = y \end{cases}$$

$$x \leq_t y \iff \begin{cases} x = \mathbf{f} \\ y = \mathbf{t} \\ x = y \end{cases}$$



Given a bounded lattice $(L, \leq)$, two standard orderings can be defined in order to provide a bilattice structure on $L \times L$:

$$\langle a, b\rangle \leq_1 \langle c, d\rangle \quad \text{if and only if} \quad a \leq c \ \text{ and } \ d \leq b \tag{1}$$

$$\langle a, b\rangle \leq_2 \langle c, d\rangle \quad \text{if and only if} \quad a \leq c \ \text{ and } \ b \leq d \tag{2}$$

With the orderings above, two different bilattices can be considered, depending on the choice of truth and knowledge, which have been used in the literature:

the bilattice $\mathcal{G}(L) = (L \times L, \leq_1, \leq_2)$ was used by Ginsberg in [10], who proved that it was able to represent both the standard notion of inference and that of assumption-based truth maintenance systems; the underlying idea in this bilattice consists in constructing pairs of the form $\langle a, b\rangle$ where a is interpreted as the degree of truth and b as the degree of falsity. The second bilattice $\mathcal{F}(L) = (L \times L, \leq_2, \leq_1)$ can be seen in [10] again, as well as in [8] and the idea here is to represent intervals.

In order to introduce the notions of consistence and default negation in a bilattice, the operators of *negation* and *conflation* will be required.

Definition 6. *Let $\mathcal{B}$ be a bilattice $(B, \leq_t, \leq_k)$*

1. *A* negation *operator over $\mathcal{B}$ is a mapping $n\colon B \to B$ such that:*
   (a) $a \leq_k b$ implies $n(a) \leq_k n(b)$;
   (b) $a \leq_t b$ implies $n(b) \leq_t n(a)$;
   (c) $n(n(a)) = a$
2. *A* conflation *operator over $\mathcal{B}$ is a mapping $-\colon B \to B$ such that:*
   (a) $a \leq_k b$ implies $-b \leq_k -a$;
   (b) $a \leq_t b$ implies $-a \leq_t -b$;
   (c) $--a = a$
   If $-$ satisfies just items (2a)-(2b) above, it is called a weak-conflation.

Notice that a negation (resp. conflation) operator reverses the truth (resp. knowledge) ordering, but preserves the knowledge (resp. truth) ordering. Once the notion of bilattice, conflation and negation operator have been introduced, we can provide the bilattice-based semantics for strong and default negated propositional symbols.

Definition 7. *Let $\mathcal{B} = (B, \leq_t, \leq_k)$ be a bilattice, a $\mathcal{B}$-interpretation is a mapping $I\colon \Pi \to B$.*

Note that each $\mathcal{B}$-interpretation is defined on the set of propositional symbols Π while the domain of L-interpretations is the set of literals Lit.

Now, we can give the definition of consistence and default negation in the framework of bilattices.

Definition 8. *Let $\mathcal{B} = (B, \leq_t, \leq_k)$ be a bilattice*

1. *The* default negation *operator is defined as $not(x) = -(n(x))$, where n and $-$ are a negation and a weak-conflation operator defined on $\mathcal{B}$.*
2. *Given a conflation operator $-$, a $\mathcal{B}$-interpretation I is said to be* consistent *if and only if $I(p) \leq_k -I(p)$ for every propositional symbol p.*

Given a $\mathcal{B}$-interpretation I, the truth-values assigned to propositional symbols negated by the strong and default negation are defined as follows:

1. $I(\sim p) = n(I(p))$
2. $I(\neg p) = not(I(p)) = -n(I(p))$

Note that under this semantics the truth-value assigned to both strong and default negation, is given compositionally in contrast to the residuated semantics.

3 Default Negation, Consistence and Coherence

We start this section by showing how we can link the residuated-based semantics (given in Section 2.1) to the bilattice-based semantics (given in Section 2.2).

Let $\mathcal{L} = (L, \leq, *, \leftarrow, n)$ be a residuated lattice with negation, every L-interpretation can be considered as a $\mathcal{G}(L)$-interpretation, and vice versa, via the following (reversible) transformation:

$$\begin{array}{c} \Omega\colon L\text{-interpretations} \to \mathcal{G}(L)\text{-interpretations} \\ \left.\begin{array}{c} p \mapsto I(p) \\ \sim p \mapsto I(\sim p) \end{array}\right\} \Rightarrow p \mapsto (I(p), I(\sim p)) \end{array} \tag{3}$$

Once the relationship between L-interpretations and $\mathcal{G}(L)$-interpretations has been fixed, the underlying mathematical structures in both frameworks can be easily related by the following result.

Proposition 1. *Let $(L, \leq, *, \leftarrow, n)$ be a residuated lattice, then $(L \times L, \leq_1, *, \leftarrow, n)$ and $(L \times L, \leq_2, *, \leftarrow, n)$ are residuated lattices, where the operators $*, \leftarrow$ and n of $\mathcal{L}$ are extended componentwise to $L \times L$ and the orderings $\leq_1$ and $\leq_2$ are those defined in (1) and (2).*

As a result of the proposition above, we have a residuated structure on $(L \times L, \leq_i, *, \leftarrow, n)$ for $i = 1, 2$, as well as the bilattice one presented above. The natural step now would be to compare both semantics for strong and default negated propositional symbols; however, in principle, the bilattice structure does not have suitable operators for negation and conflation.

Note that n does not define a negation operator on $\mathcal{G}(L)$, since it is antitonic with respect to both orderings; this is not really a problem, for we can always define in $\mathcal{G}(L)$ a "natural" negation operator by $\overline{n}(\langle a, b\rangle) = \langle b, a\rangle$. On the other hand, given the negation operator n from $(L, \leq, *, \leftarrow, n)$, we can define the following weak-conflation in $\mathcal{G}(L)$:

$$-\langle a, b\rangle = \langle n(b), n(a)\rangle$$

With these definitions we obtain the following interesting result.

Proposition 2. *The default negation in $\mathcal{G}(L)$, defined as not$(x) = -(\overline{n}(x))$, coincides with the componentwise extension of the negation operator from L.*

In order to establish consistence in $\mathcal{G}(L)$, a non-weak conflation is needed. Thus, to ensure that the operator $-$ defined above is actually a conflation, we have to assume that the negation operator n defined on L is involutive. In that case the notions of coherence and consistence are equivalent, that is:

Proposition 3. *Let $(L, \leq, *, \leftarrow, n)$ be a residuated lattice where n is an involutive operator. Then I a coherent interpretation in L if and only if $\Omega(I)$, as defined by (3), is a consistent interpretation in $\mathcal{G}(L)$ with respect to the conflation operator $-$.*

Note that, at first sight, the definition of consistence in a $\mathcal{G}(L)$-interpretation $\Omega(I)$ implies two different inequalities in I, namely:

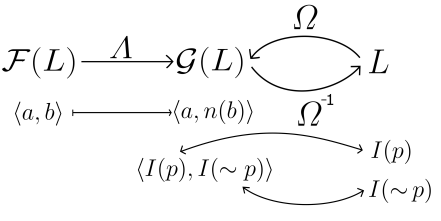
$$I(\sim p) \leq n(I(p)) \quad \text{(coherence)} \tag{4}$$
$$I(p) \leq n(I(\sim p)) \quad \text{(dual-coherence)} \tag{5}$$

but under the hypothesis of Proposition 3, that is when n is involutive, coherence implies dual-coherence. So, the second inequality imposed by the definition of consistence in $\mathcal{G}(L)$ is unnecessary in that case. The following question concerning the previous proposition arises now: what is the relationship between coherence and consistence when the negation operator in $\mathcal{L}$ is not involutive?

The answer is not straightforward, as there is not a natural conflation in $\mathcal{G}(L)$, in the sense of the negation $\overline{n}$, which is independent of the negation in L. Obviously, defining a conflation on $\mathcal{G}(L)$ by using a negation operator different from n seems inadequate. We have chosen to use a bilattice structure which admits a natural conflation, namely, $\mathcal{F}(L)$. The problem here is that we cannot identify L-interpretations one-to-one with $\mathcal{F}(L)$-interpretations, as in the case of $\mathcal{G}(L)$. But, by using the negation operator defined on $\mathcal{L}$, we can identify every element in $\mathcal{F}(L)$ with another in $\mathcal{G}(L)$ by preserving both orderings. In other words, we can define the following operator:

$$\begin{aligned} \Lambda\colon \mathcal{F}(L) &\to \mathcal{G}(L) \\ \langle a, b\rangle &\mapsto \langle a, n(b)\rangle \end{aligned}$$

Note that Λ is not necessarily a one-to-one mapping, since n need not be injective. Note as well that, by using the mapping Ω, we can assign to each $\mathcal{F}(L)$-interpretation an L-interpretation. Pictorially, we have the following mappings among $\mathcal{F}(L)$-interpretations, $\mathcal{G}(L)$-interpretations and L-interpretations:



The advantage of using $\mathcal{F}(L)$ is that we can define a natural conflation operator without using the operator n:

$$-(\langle a, b\rangle) = \langle b, a\rangle$$

As a result, the definition of consistence in this structure is the following: a pair $\langle a, b\rangle$ is consistent in the bilattice of intervals if and only if $a \leq b$ (that is, if and only if $\langle a, b\rangle$ actually defines a interval).

The following proposition shows the relationship between coherence in L and consistence in $\mathcal{F}(L)$:

Proposition 4.

1. *If J is a consistent $\mathcal{F}(L)$-interpretation, then there exists a coherent L-interpretation I such that $\Omega(I) = \Lambda(J)$.*
2. *If I is a coherent L-interpretation, then there exists a consistent $\mathcal{F}(L)$-interpretation J such that $\Omega(I) \leq_k \Lambda(J)$.*

Proof. (Sketch)

1. Assuming that $J(p) = \langle a, b\rangle$, we define $I(p) = a$ and $I(\sim p) = n(b)$.
2. Given I, the $\mathcal{F}(L)$-interpretation J is defined as $J(p) = \langle I(p), I(p)\rangle$. □

It is important to recall that inconsistency is linked to the knowledge ordering in the following sense: let I_1 and I_2 be two $\mathcal{B}$-interpretations such that $I_1 \leq_k I_2$ and I_2 is consistent, then I_1 is consistent as well. Thus, we have the following:

Corollary 1. *Assume that we have a conflation in $\mathcal{G}(L)$ such that Λ assigns consistent $\mathcal{F}(L)$-interpretations to consistent $\mathcal{G}(L)$-interpretations, then I is a coherent L-interpretation if and only if $\Omega(I)$ is a consistent $\mathcal{G}(L)$-interpretation.*

This corollary allows to further justify why the definition of coherence given in [13,14] was given in terms of inequality (4) above, and not using inequality (5) as an alternative. Even when the negation in the residuated lattice is not involutive, the following example shows that none of the implications of Proposition 4 hold under the sole hypothesis of inequality of dual-coherence (5).

Example 2. Consider the negation operator $n(x) = 1-\sqrt{x}$ on $[0, 1]$. The $\mathcal{F}([0, 1])$-interpretation defined by $J(p) = \langle 0.64, 0.64\rangle$ is consistent but does not satisfy in Proposition 4(1): the only $[0, 1]$-interpretation I satisfying $\Omega(I) = \Lambda(J)$ is the $[0, 1]$-interpretation given by $I(p) = 0.64$ and $I(\sim p) = 0.2$, but the dual-coherence inequality does not hold, since $I(p) = 0.64 \not\leq 1 - \sqrt{I(\sim p)} \approx 0.55$.

For the other item, consider the negation operator $n(x) = 1 - x^2$, again on $[0, 1]$, and the L-interpretation given by $I(p) = 0.75$ and $I(\sim p) = 0.46$. In order to check that Proposition 4(2) does not hold, it suffices to exhibit an inconsistent $\mathcal{F}(L)$-interpretation satisfying $\Omega(I) \geq_k \Lambda(J)$, since all the upper bounds of $\Lambda(J)$ would be inconsistent as well. Simply define $J(p) = \langle 0.75, 0.74\rangle$, it is straightforward to check that J is inconsistent and, furthermore, the inequality $\Omega(I) \geq_k \Lambda(J)$ holds.

4 Conclusions

The relationship between some of the approaches developed, on the one hand, based of residuated lattices and, on the other hand, based on the theory of bilattices has been investigated. Specifically, we have established links between the notions of coherence in the residuated-based approach and consistence in the bilattice-based approaches (those based on Ginsberg's $\mathcal{G}(L)$ and on Fitting's $\mathcal{F}(L)$), and proved the equivalence of both approaches under certain reasonable conditions.

References

1. Alcântara, J., Damásio, C.V., Pereira, L.M.: An encompassing framework for paraconsistent logic programs. J. Applied Logic 3(1), 67–95 (2005)
2. Balduccini, M., Gelfond, M.: Logic programs with consistency-restoring rules. In: International Symposium on Logical Formalization of Commonsense Reasoning. AAAI 2003 Spring Symposium Series, pp. 9–18 (2003)
3. Barceló, P., Bertossi, L.: Logic programs for querying inconsistent databases. In: Dahl, V. (ed.) PADL 2003. LNCS, vol. 2562, pp. 208–222. Springer, Heidelberg (2002)
4. Cornelis, C., Arieli, O., Deschrijver, G., Kerre, E.E.: Uncertainty modeling by bilattice-based squares and triangles. IEEE T. Fuzzy Systems 15(2), 161–175 (2007)
5. Damásio, C.V., Pereira, L.M.: Monotonic and residuated logic programs. In: Benferhat, S., Besnard, P. (eds.) ECSQARU 2001. LNCS (LNAI), vol. 2143, pp. 748–759. Springer, Heidelberg (2001)
6. Deschrijver, G., Arieli, O., Cornelis, C., Kerre, E.E.: A bilattice-based framework for handling graded truth and imprecision. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15(1), 13–41 (2007)
7. Dubois, D., Prade, H., Ughetto, L.: Checking the coherence and redundancy of fuzzy knowledge bases. IEEE T. Fuzzy Systems 5(3), 398–417 (2002)
8. Fitting, M.: Bilattices and the semantics of logic programming. Journal of Logic Programming 11, 91–116 (1991)
9. Fitting, M.: Fixpoint semantics for logic programming - a survey. Theoretical Computer Science 278, 25–51 (1999)
10. Ginsberg, M.: Multivalued logics: A uniform approach to inference in artificial intelligence. Computational Intelligence 4, 265–316 (1988)
11. Grant, J., Hunter, A.: Measuring inconsistency in knowledgebases. Journal of Intelligent Information Systems 27(2), 159–184 (2006)
12. Hájek, P.: Metamathematics of Fuzzy Logic. Trends in Logic. Kluwer Academic, Dordrecht (1998)
13. Madrid, N., Ojeda-Aciego, M.: Towards a fuzzy answer set semantics for residuated logic programs. In: Proc. of WI-IAT 2008 Workshop on Fuzzy Logic in the Web, pp. 260–264 (2008)
14. Madrid, N., Ojeda-Aciego, M.: On coherence and consistence in fuzzy answer set semantics for residuated logic programs. In: Di Gesù, V., Pal, S.K., Petrosino, A. (eds.) WILF 2009. LNCS, vol. 5571, pp. 60–67. Springer, Heidelberg (2009)
15. Madrid, N., Ojeda-Aciego, M.: Measuring inconsistency in fuzzy answer set semantics. IEEE Transactions on Fuzzy Systems (2011) (accepted)
16. Yuan, L., You, J.: Coherence approach to logic program revision. IEEE Tr. Knowledge and Data Engineering 10(1), 108–119 (2002)

Investigation of Evolving Fuzzy Systems Methods FLEXFIS and eTS on Predicting Residential Prices

Bogdan Trawiński[1], Krzysztof Trawiński[2], Edwin Lughofer[3], and Tadeusz Lasota[4]

[1] Wrocław University of Technology, Institute of Informatics, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
bogdan.trawinski@pwr.wroc.pl

[2] European Centre for Soft Computing, Edificio Científico-Tecnológico, planta 3, C. Gonzalo Gutiérrez Quirós, s/n, 33600 Mieres, Spain
krzysztof.trawinski@softcomputing.es

[3] Johannes Kepler University Linz, Department of Knowledge-based Mathematical Systems, Altenbergerstrasse 69, A-4040 Linz, Austria
edwin.lughofer@jku.at

[4] Wrocław University of Environmental and Life Sciences, Dept. of Spatial Management Ul. Norwida 25/27, 50-375 Wrocław, Poland
tadeusz.lasota@up.wroc.pl

Abstract. In this paper, we investigate on-line fuzzy modeling for predicting the prices of residential premises using the concept of evolving fuzzy models. These combine the aspects of incrementally updating the parameters and expanding the inner structure on demand with the concepts of uncertainty modeling in a possibilistic and linguistic manner (via fuzzy sets and fuzzy rule bases). The FLEXFIS and eTS approaches are evolving fuzzy models used to compare with an expert-based property valuating method as well as with a classic genetic fuzzy system. We use a real-world dataset taken from a cadastral system for that comparison.

1 Introduction

Nowadays, the professional real estate appraisers are more and more supported by computer systems called Automated Valuation Models (AVM) and Computer Assisted Mass Appraisal (CAMA), which incorporate statistical multiple regression models [20], neural networks [21], and the combination of soft computing and spatial analysis methods using data provided by geographic information systems (GIS) [8]. Other intelligent methods have been also developed to assist professional valuers including: case-based reasoning [4], rough set theory [6], and hybrid approaches [13]. However, the need to provide meaningful and interpretable models resulted in the proposals of fuzzy and neuro-fuzzy systems [10,11] as alternative solutions. The application of evolving fuzzy models could be especially useful because each day one cadastral information centre in a big city register dozens of real estate sales transactions which are the base of sales

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 123–130, 2011.

comparison valuation methods, the most commonly used by the appraisers. If all the data would be ordered by the transaction date, they would constitute some form of a data stream which in turn could reflect the changes of real estate market in the course of time. This motivated us also to use evolving models, which are able to process streams of data and to learn, update, expand their memory on-line on demand. So far we have studied genetic and evolving fuzzy real estate appraisal models [15,16]. In this paper we make one step forward, we compare two evolving fuzzy techniques FLEXFIS and eTS with one of the most common state-of-the-art valuation approaches of professional appraisers based on sliding windows, nearest neighbors elicitation and averaging of past prices to estimate a new price. As the second reference method we employed also a classic genetic fuzzy rule-based system trained in off-line mode.

We conduct our experiments by using data-driven fuzzy rule-based systems (FRBS) as a specific model architecture to build reliable models in the field of residential premises. Data-driven fuzzy systems have are characterized by three important features. 1.) They are able of approximating any real continuous function on a compact set with an arbitrary accuracy [5,14]. 2.) FRBSs have the capability of knowledge extraction and representation when modeling complex systems in a way that they could be understood by a human being called Interpretability [1]. 3.) Data-driven fuzzy systems can be permanently updated on demand based on new incoming samples as is the case for on-line measurements or data streams. Technologies for providing such updates with high performance (both in computational times and predictive accuracy) are called evolving fuzzy systems approaches or simply evolving fuzzy systems [19].

In this work, we exploit two evolving fuzzy systems approaches dealing with incremental on-line modeling demands. The first approach, the so-called FLEXFIS method [20] incrementally evolves clusters (associated with rules) and performs a recursive adaptation of consequent parameters by using local learning approach. The second approach, the so-called eTS approach [2] is also an incremental evolving approach based on recursive potential estimation in order to extract the most dense regions in the feature space as cluster centers (rule representatives). In [3], both are compared in terms of methodological and empirical aspects.

2 Fuzzy Modeling in On-line Mode with Evolving Fuzzy Systems

In the following, we present FLEXFIS and eTS approaches as the two most widely used EFS approaches, which are based on the so-called Takagi-Sugeno (TS) fuzzy systems [22] (in a specific form called fuzzy basis function networks [23]), in order to cope with the dynamics of (on-line) streaming data over time and applied to the prediction of residential premises in the evaluation section.

2.1 The FLEXFIS Approach

The FLEXFIS approach, short for FLEXible Fuzzy Inference Systems was first introduced in [19] and significantly extended version in [17], and designed for

the purpose of incremental learning of TS fuzzy systems from data streams in a sample-wise single-pass manner. This means that always one sample can be loaded, sent into the FLEXFIS learning engine where the model is updated and immediately discarded, afterwards. In this sense, the method needs low resources 1.) with respect to computational complexity and 2.) with respect to virtual memory and hence is feasible for on-line modeling applications, where models should be kept-up-to-date as early as possible in order to account for new operating conditions, systems states etc. and to prevent extrapolation situations when performing predictions on new samples. The basic steps in FLEXFIS approach can be summarized as follows:

1. Rule evolution and updating of antecedent parameters in the cluster space with the help of an incremental evolving clustering variant.
2. Recursive adaptation of consequent parameters exploiting the local learning approach (parameters are updated for each rule separately).
3. Balancing out a non-optimal situation by adding a correction vector to the vector of consequent parameters.
4. In the extended version: Detecting redundant rules with the help of specific overlap measures (two variants: one-dimensional intersection points of fuzzy sets and inclusion metric) and performing on-line rule merging/pruning after each incremental update cycle.

To get a deep insight of FLEXFIS we recommend the reader to look into [17] and for pruning rules on demand into [18].

2.2 The eTS Approach

The basic difference to the FLEXFIS approach lies in the way in which the antecedent parameters in the fuzzy sets are modified respectively new rules are added and evolved during the on-line mode based on new incoming samples. In doing so, a product-space clustering approach is also exploited, using a recursive calculation of potentials of data samples respectively cluster prototypes. Subtractive clustering is extended to the incremental on-line case (called *eClustering*) by associating the first data sample with the center of the first cluster and calculating the potential P of new incoming data samples recursively. Those data samples having higher potentials than all already existing clusters are candidates for new cluster centers. Two different situations can be distinguished: 1.) the new sample having highest potential is close to an existing cluster, then the existing cluster center is replaced to the new data sample; 2.) the new sample having highest potential is far away from all existing clusters, then a new rule is born with center coordinates equal to the coordinates of the new sample (rule evolution step).

The consequent parameters in TS fuzzy models are either estimated by recursive fuzzily weighted least squares (local learning) or recursive least squares with integration of dynamic merging of the inverse Hessian matrix (global learning). To get a deep insight of eTS we recommend the reader to look into [2].

3 Experimental Setup

The main goal of experiment was to compare evolving data driven fuzzy modeling techniques, i.e. conventional FLEXFIS(c), FLEXFIS(p) with pruning, and eTS with genetic fuzzy systems. The techniques were applied to real-world regression problem of predicting the prices of residential premises based on historical data of sales transactions obtained from a cadastral system and registry of real estate transactions. Moreover, we aimed to compare the soft computing algorithms with a property valuating method employed by professional appraisers in reality.

Real-world dataset used in experiments was drawn from the registry of real estate transactions accomplished in one Polish big city within 11 years from 1998 to 2008. After cleansing, the final dataset selected by the experts counted 5213 records. Five following features were pointed out as the main drivers of premises prices: usable area of premises, age of a building, number of rooms in a flat, number of storeys in a building, and distance from the city centre. As target values total prices of premises were used.

Table 1. Training/test data portions used for validation of evolved fuzzy models

Dataset	Training portions	# train.	Test	# test
01	199802-200002	1008	200101	228
02	199901-200101	1034	200102	235
03	199902-200102	1056	200201	267
04	200001-200201	1059	200202	263
05	200002-200202	1160	200301	267
06	200101-200301	1260	200302	386
07	200102-200302	1418	200401	278
08	200201-200401	1461	200402	268
09	200202-200402	1462	200501	244
10	200301-200501	1443	200502	336
11	200302-200502	1512	200601	300
12	200401-200601	1426	200602	377
13	200402-200602	1525	200701	289
14	200501-200701	1546	200702	286
15	200502-200702	1588	200801	181

The comparative experiments were based on a streaming data context which was simulated by the data of consecutive half-years joined together as they were recorded. This means that in the incremental learning context the same order of the data samples is supported as the temporal order in which the samples were stored onto hard-disc, yielding a one-to-one simulation of the real on-line case. In order to achieve a data stream with a significant length, we joined five consecutive half-years to form one training data set (e.g. half-years 199802 to 200002 to form a data set of 1008 samples) and used the subsequent half-year (e.g. half-year 200101 with 228 samples) as separate test data set in order to verify the predictive power of the models. Training and test data portions used for on-line testing of evolved fuzzy models are shown in Table 1. The normalization of data was accomplished using the min-max approach, as performance functions the root mean square error (RMSE) was applied.

To each actual transactional price, we had the corresponding value of premises estimated by the experts at our disposal. In this sense, we are able to conduct a performance comparison among three evolving fuzzy modeling and the expert-based estimations, which can be seen as the main focus in this paper.

In order to compare evolutionary machine learning algorithms with techniques applied to property valuation we asked experts to evaluate premises using their pairwise comparison method to historical data of sales/purchase transactions recorded in a cadastral system. The expert's method is based on sliding windows, nearest neighbors elicitation and averaging of past prices to estimate a new price. The experts worked out a computer program which simulated their routine work and was able to estimate the experts' prices of a great number of premises automatically.

Moreover, we wanted to obtain a clue how good evolving fuzzy modeling approaches perform compared with alternative data-driven modeling techniques with similar interpretation capabilities, but operating in the off-line mode. Therefore, we also applied genetic fuzzy systems (GFS) [12] in two experimental procedures. The first provided the apparent error that is GFS was learned and tested employing the whole half-year dataset, it was denoted as GFS(a). The second, in turn, used the random splits of each half-year dataset into training and test sets in the proportion 70% to 30%. In order to avoid the effect of lucky or unfortunate division, all learning/testing cycles were repeated 50 times on different holdout splits and the final outcome was averaged. This procedure was named GFS(h). Both GFSs were Mamdani type with the same inputs and outputs as evolving techniques, 3 triangular membership functions in each input and 5 in the output. The other main parameters were as follows: number of rules, population size, and number of generations were equal to 15, 50, and 100, respectively.

For verifying statistical significance of differences among all modeling methods and the expert-based technique, we conducted the pairwise non-parametric Wilcoxon test. We applied also advanced non-parametric approaches which control the error propagation of making multiple comparisons. They included the rank-based non-parametric Friedman test and its Iman-Davenport correction followed by Nemenyi's, Holm's, Shaffer's, and Bergmann-Hommel's post-hoc procedures which are recommended for $n \times n$ comparisons [7,9].

4 Results of Experiment

The performance of compared techniques is shown in Table 2. Comparing FLEXFIS(c) against eTS the former outperforms the latter in 5 out of 15 cases (+ 5 ties), whereas comparing FLEXFIS(p) against eTS the latter outperforms the former in 7 out of 15 cases (+5 ties) in terms of accuracy. Considering a number of rules, both FLEXFIS variants obtained lower value than eTS. Conventional FLEXFIS did so in 9 out of 15 times (+1 tie) and FLEXFIS with pruning did so in 9 out of 15 times (+2 ties). The Expert and GFS methods provided apparently greater predictive errors. Execution times of individual algorithms are not presented because they were incomparable: evolving fuzzy completed their

Table 2. Performance of tested techniques by predicting the premises prices (#r stands for the number of rules)

Dataset	Flexfis(c)		Flexfis(p)		eTS		Expert	GFS(a)	GFS(h)
	RMSE	#r	RMSE	#r	RMSE	#r	RMSE	RMSE	RMSE
01	0.0332	12	0.0332	12	0.0300	6	0.1485	0.0725	0.0966
02	0.0224	2	0.0224	2	0.0283	8	0.1010	0.0948	0.1136
03	0.0265	1	0.0265	1	0.0265	3	0.1019	0.0969	0.1148
04	0.0265	1	0.0265	1	0.0265	6	0.1504	0.0681	0.0871
05	0.0245	3	0.0245	2	0.0245	4	0.0936	0.0948	0.1118
06	0.0300	6	0.0300	4	0.0300	8	0.0933	0.0936	0.0937
07	0.0436	4	0.0500	4	0.0447	8	0.1118	0.1003	0.1210
08	0.0374	3	0.0374	2	0.0374	4	0.1128	0.1018	0.1166
09	0.0490	3	0.0490	3	0.0374	4	0.1270	0.0951	0.1091
10	0.0400	2	0.0400	2	0.0387	5	0.1053	0.0966	0.1093
11	0.0794	9	0.0800	7	0.0819	4	0.1012	0.1083	0.1222
12	0.1030	8	0.1010	5	0.1049	5	0.1043	0.1017	0.1114
13	0.1300	4	0.1425	4	0.1327	4	0.0858	0.0674	0.0853
14	0.1170	8	0.1292	6	0.1114	5	0.0968	0.0822	0.1008
15	0.1000	16	0.0911	7	0.0693	6	0.1458	0.0824	0.1072

Table 3. Adjusted p-values for n×n comparisons of fuzzy models

Alg vs Alg	pWilcox	pNeme	pHolm	pShaf	pBerg
eTS vs GFS(h)	*0.0026*	*0.0004*	*0.0004*	*0.0004*	*0.0004*
FLEXFIS(c) vs GFS(h)	*0.0031*	*0.0006*	*0.0006*	*0.0004*	*0.0004*
FLEXFIS(p) vs GFS(h)	*0.0031*	*0.0021*	*0.0018*	*0.0014*	*0.0010*
eTS vs Expert	*0.0026*	*0.0136*	*0.0109*	*0.0091*	*0.0091*
FLEXFIS(c) vs Expert	*0.0022*	*0.0192*	*0.0141*	*0.0128*	*0.0091*
FLEXFIS(p) vs Expert	*0.0031*	0.0512	*0.0341*	*0.0341*	*0.0137*
GFS(a) vs GFS(h)	*0.0007*	0.1675	0.1005	0.0782	0.0670
Expert vs GFS(a)	*0.0038*	1.0000	0.7769	0.6798	0.3884
eTS vs GFS(a)	*0.0146*	1.0000	0.7769	0.6798	0.6798
FLEXFIS(c) vs GFS(a)	*0.0146*	1.0000	0.7769	0.7105	0.6798
FLEXFIS(p) vs GFS(a)	*0.0409*	1.0000	1.0000	0.8182	0.6798
Expert vs GFS(h)	0.7333	1.0000	1.0000	1.0000	1.0000
FLEXFIS(p) vs eTS	0.1141	1.0000	1.0000	1.0000	1.0000
FLEXFIS(c) vs FLEXFIS(p)	0.3454	1.0000	1.0000	1.0000	1.0000
FLEXFIS(c) vs eTS	0.5076	1.0000	1.0000	1.0000	1.0000

calculations within several seconds whereas the genetic fuzzy systems needed a few hours to produce results. For Friedman and Iman-Davenport tests, the calculated values of χ^2 and F statistics were 31.95 and 10.39, respectively, whereas the critical values at $\alpha = 0.05$ are $\chi^2(5) = 12.83$ and $F(5, 140) = 2.28$, so that the null-hypotheses were rejected showing that there were statistically significant differences among modeling methods in respect of their predictive accuracy. Average ranks of individual methods placed algorithms in the following order: eTS (2.40), FLEXFIS(c) (2.47), FLEXFIS(p) (2.67), GFS(a) (3.53), Expert (4.67), GFS(h) (5.27), where the lower rank value the better model. It could be seen there are minor differences among all evolving fuzzy algorithms and these take clearly lower positions than genetic fuzzy systems and the Expert's method. In Table 3 p-values for Wilcoxon test and adjusted p-values for Nemenyi, Holm, Shaffer, and Bergmann-Hommel tests are placed for $n \times n$ comparisons for all possible 15 pairs of modeling methods. The p-values below 0.05 indicate that

respective algorithms differ significantly in prediction errors (italic font). According to the Wilcoxon tests all three evolving fuzzy techniques revealed significantly better performance than others. In any case statistically significant differences among eTS, FLEXFIS(c), and FLEXFIS(p) methods could not be observed. However, the post-hoc procedures showed that evolving fuzzy modeling techniques provided significantly better price estimations than GFS(h) and Expert's methods.

5 Conclusions and Future Work

In this paper, we investigated the usage of incremental data-driven fuzzy modelling techniques (FLEXFIS and eTS) for the purpose to estimate the future prices of residential premises based on the past recordings. A comprehensive data set including over 5200 samples and recorded during the time span 1998 to 2008 served as basis for benchmarking FLEXFIS and eTS. Our investigation showed that evolving fuzzy methods produce models with prediction accuracy greater than the experts' valuation method routinely employed in reality and a classic genetic fuzzy rule-based system. Moreover, they allow for linguistic interpretation of the modeled dependencies as could be verified when using FLEXFIS in connection with rule merging/pruning techniques. Thus, we can conclude that evolving fuzzy modeling methods are a reliable and powerful alternative for valuation of residential premises.

Acknowledgements. This work was funded partially by the Austrian fund for promoting scientific research (FWF, contract number I328-N23, acronym IREFS) as well as by the Polish National Science Centre under grant no. N N516 483840.

References

1. Alonso, J.M., Magdalena, L., González-Rodríguez, G.: Looking for a good fuzzy system interpretability index: An experimental approach. Int. J. of Approximate Reasoning 51, 115–134 (2009)
2. Angelov, P., Filev, D.: An approach to online identification of Takagi-Sugeno fuzzy models. IEEE Trans. Syst., Man, Cybern. B 34(1), 484–498 (2004)
3. Angelov, P., Lughofer, E.: Data-driven evolving fuzzy systems using eTS and FLEXFIS: Comparative analysis. Int. J. of General Systems 37(1), 45–67 (2008)
4. Bonissone, P., Cheetham, W.: Financial applications of fuzzy case-based reasoning to residential property valuation. In: Proceedings of the Sixth IEEE Int. Conference on Fuzzy Systems 1997, vol. 1, pp. 37–44 (July 1997)
5. Castro, J., Delgado, M.: Fuzzy systems with defuzzification are universal approximators. IEEE Trans. Syst., Man, Cybern. B 26, 149–152 (1996)
6. D'Amato, M.: Comparing rough set theory with multiple regression analysis as automated valuation methodologies. Int. Real Estate Review 10(2), 42–65 (2007)
7. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. J. of Machine Learning Research 7, 1–30 (2006)

8. García, N., Gámez, M., Alfaro, E.: ANN+GIS: An automated system for property valuation. Neurocomputing 71(4–6), 733–742 (2008)
9. García, S., Herrera, F.: An extension on statistical comparisons of classifiers over multiple data sets. J. of Machine Learning Research 9, 2677–2694 (2008)
10. González, M.A.S., Formoso, C.: Mass appraisal with genetic fuzzy rule-based systems. Property Management 24(1), 20–30 (2006)
11. Guan, J., Zurada, J., Levitan, A.S.: An adaptive neuro-fuzzy inference system based approach to real estate property assessment. J. of Real Estate Research 30(4), 395–422 (2008)
12. Kempa, O., Lasota, T., Telec, Z., Trawiński, B.: Investigation of bagging ensembles of genetic neural networks and fuzzy systems for real estate appraisal. In: Nguyen, N.T., Kim, C.-G., Janiak, A. (eds.) ACIIDS 2011, Part II. LNCS (LNAI), vol. 6592, pp. 323–332. Springer, Heidelberg (2011)
13. Kontrimas, V., Verikas, A.: The mass appraisal of the real estate by computational intelligence. Applied Soft Computing 11(1), 443–448 (2011)
14. Kosko, B.: Fuzzy systems as universal approximators. IEEE Trans. on Computers 43(11), 1329–1333 (1994)
15. Król, D., Lasota, T., Trawiński, B., Trawiński, K.: Investigation of evolutionary optimization methods of tsk fuzzy model for real estate appraisal. Int. J. Hybrid Intell. Syst. 5(3), 111–128 (2008)
16. Lasota, T., Telec, Z., Trawiński, B., Trawiński, K.: Investigation of the ets evolving fuzzy systems applied to real estate appraisal. J. of Multiple-Valued Logic and Soft Computing 17(2-3), 229–253 (2011)
17. Lughofer, E.: FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. IEEE Trans. on Fuzzy Systems 16(6), 1393–1410 (2008)
18. Lughofer, E., Bouchot, J.L.: On-line elimination of local redundancies in evolving fuzzy systems, evolving systems. Evolving Systems (2011) (in revision)
19. Lughofer, E., Klement, E.: FLEXFIS: A variant for incremental learning of Takagi-Sugeno fuzzy systems. In: Proceedings of FUZZ-IEEE 2005, Reno, Nevada, U.S.A, pp. 915–920 (2005)
20. Nguyen, N., Cripps, A.: Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. J. of Real Estate Research 22(3), 313–336 (2001)
21. Selim, H.: Determinants of house prices in turkey: Hedonic regression versus artificial neural network. Expert Systems with Applications 36, 2843–2852 (2009)
22. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst., Man, Cybern. 15(1), 116–132 (1985)
23. Wang, L., Mendel, J.: Fuzzy basis functions, universal approximation and orthogonal least-squares learning. IEEE Trans. Neural Networks 3(5), 807–814 (1992)

An Empirical Study on Interpretability Indexes through Multi-objective Evolutionary Algorithms

R. Cannone[1], J. M. Alonso[2], and L. Magdalena[2]

[1] Department of Informatics, University of Bari, Bari 70125, Italy
raffaelecannone@di.uniba.it
[2] European Centre for Soft Computing, Mieres 33600, Spain
{jose.alonso,luis.magdalena}@softcomputing.es

Abstract. In the realm of fuzzy systems, interpretability is really appreciated in most applications, but it becomes essential in those cases in which an intensive human-machine interaction is necessary. Accuracy and interpretability are often conflicting goals, thus we used multi-objective fuzzy modeling strategies to look for a good trade-off between them. For assessing interpretability, two different interpretability indexes have been taken into account: Average Fired Rules (AFR), which estimates how simple the comprehension of a specific rule base is, and Logical View Index (LVI), which estimates how much a rule base satisfies logical properties. With the aim of finding possible relationships between AFR and LVI, they have been used in two independent experimental sessions against the classification error. Experimental results have shown that the AFR minimization implies the LVI minimization, while the opposite is not verified.

Keywords: interpretability, fuzzy modeling, multi-objective evolutionary algorithm, interpretability indexes.

1 Introduction

In the realm of fuzzy systems, interpretability is well-appreciated in most applications, but it becomes essential in those cases in which an intensive human-machine interaction is necessary. For example, decision support systems in medicine must be interpretable, indeed each suggestion provided by the system must be comprehensible for both physicians and patients with the aim of being accepted.

Recently, a complete taxonomy on the existing interpretability measures has been proposed in [5]. In a nutshell, authors identify four groups of indexes by combining two different criteria, i.e. the nature of the interpretability index (complexity vs. semantic) and the elements of the fuzzy knowledge base taken into account (partitions vs. rule base).

The four groups are: (1) Complexity at rule base level; (2) complexity at partition level; (3) semantic interpretability at rule base level; and (4) semantic interpretability at partition level. Most well-known existing interpretability

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 131–138, 2011.

indexes correspond to groups (1) and (2) as they only focus on complexity, e.g. counting the number of membership functions, rules, etc. In group (4), indexes usually measure the degree of fulfillment of semantic constraints. Few indexes proposed in literature belong to group (3) and are mainly devoted to evaluating the rule base consistency. Moreover, the effectiveness of such indexes is still not proved in experimental trials. As far as we know, this is the first empirical study aimed to evaluate the goodness of existent indexes for assessing semantic interpretability at rule base level, following a comparative approach.

Accuracy and interpretability are often conflicting goals, thus looking for a good trade-off between them is one of the most difficult and challenging tasks in fuzzy system modeling. The use of multi-objective fuzzy modeling strategies perfectly fits with this problem and it has become a very fruitful research field [7].

This paper is focused on analyzing two different interpretability indexes belonging to group (3): Average Fired Rules (AFR), which estimates how simple (or difficult) the comprehension of a specific rule base is by counting the mean number of rules fired at the same time [2] and Logical View Index (LVI), which estimates how much a rule base satisfies logical properties [8]. The objective of this analysis is to evaluate possible relationships between the two indexes. If the two indexes are unrelated then they are both required for assessing the interpretability of a fuzzy system. The underlying idea is to exploit a Multi-Objective Evolutionary Algorithm (MOEA) for producing several rule bases with different accuracy-interpretability trade-offs. Integration of Multi-Objective Evolutionary Algorithms and Fuzzy Systems is a recent and growing research area [1,7]. We restrict our study on fuzzy rule-based classifiers (FRBCs) since current implementation of LVI only can be applied to these models.

Our contribution is organized as follows. Section 2 presents both AFR and LVI. Section 3 introduces the proposed evolutionary framework. Section 4 summarizes the experiments carried out along with the achieved results. Finally, some conclusions and future works are sketched in Section 5.

2 Indexes for Assessing Semantic Interpretability

Semantic interpretability assessment estimates how much the user is able to understand the knowledge embedded into a system. Considering fuzzy systems, it is known that the most effective way to express and communicate knowledge is by means of composition of linguistic terms, i.e. fuzzy rules. Hence, the rule base can be considered as the main communication interface to users. In this section we consider two indexes for evaluating semantic interpretability at rule base level.

2.1 Average Fired Rules

Average Fired Rules (AFR) is an index for estimating how difficult (or easy) the comprehension of a rule base could be by counting the mean value of rules fired at the same time [2].

The evaluation is made as follows: for each sample in the dataset, the inference system is used for carrying out the strength of each rule. Rules with strength greater than a threshold "delta" (0.1) are considered as "fired". The mean number of "fired" rules is finally computed.

When AFR is small on the average the number of fired rules is small. As a consequence, the rule base provides for specific knowledge on each sample, hence the user understanding of the inferred result is facilitated. On the other hand, a high AFR indicates that a high number of rules is required for inferring the output for a given input. This makes difficult for the user to understand the system behavior from a local point of view.

2.2 Logical View Index

A new technique for semantic interpretability evaluation, called *Logical View* (LV), has been recently introduced in [8]. LV identifies two different semantics associated to a FRBC: explicit semantics, which is defined by fuzzy sets and operators used in the fuzzy system, and implicit semantics, which is the knowledge gathered by the user in his/her mind while reading and interpreting the rule base.

Interpretability evaluation is made by comparing explicit and implicit semantics through the cointesion concept. In few words, both implicit and explicit semantics are cointensive if they almost refer to the same entities. Since implicit semantics is encoded into the user's mind, LV is based on logical properties, which are assumed to hold for user's knowledge. The general approach for LV evaluation is to apply truth-preserving operators, and to assess the eventual change in classification ability.

To sum up, the LV approach for interpretability assessment is composed of the following steps: the rule base is transformed into several truth tables without any change of semantics, then the truth tables are minimized by applying truth-preserving operators. The reduced truth tables are then transformed into propositions constituting a new rule base, which is compared with the original one on the basis of their classification performance. If they do not differ too much, we recognize that the logical view of the original FRBC is in agreement with the explicit semantics exhibited by the fuzzy rules, so the cointension with the user's knowledge is verified and the FRBC can be deemed as interpretable. On the other hand, if the two rule bases are characterized by notably different classification behavior, then the logical view of the FRBC is not compatible with the explicit semantics of fuzzy rules, therefore the knowledge base is not cointensive with user's knowledge and it can be deemed as not interpretable.

3 Evolutionary Framework

This section describes thoroughly all the components of the proposed evolutionary framework. The starting point is the HILK (Highly Interpretable Linguistic Knowledge) fuzzy modeling methodology [3] which focuses on making easier the

design process of interpretable FRBCs. As described in [2], embedding HILK into a multi-objective evolutionary algorithm yields a set of non-dominate FRBCs which are exhibiting a wide range of interpretability-accuracy trade-offs.

Each chromosome consists of a fixed-length string with a number of genes equal to the number of input variables. Each gene represents the number of linguistic terms defined for the related input. It can take integer values between one and nine, where one means *Don't Care*, i.e., the input is filtered and not taken into account during the rule base generation.

Given a chromosome, HILK is able to generate a whole FRBC in three main steps: (1) partition design, (2) rule base learning, and (3) knowledge base improvement. Uniformly distributed strong fuzzy partitions (SFPs) with fuzzy sets of triangular shape are generated considering the number of terms codified by each gene. Such partitions, by definition, satisfy most properties demanded to be interpretable (distinguishability, coverage, etc).

Then, the Fast Prototyping Algorithm (FPA) [6] is run to generate a set of linguistic rules, taking as inputs both the previously defined fuzzy partitions and the training data set. Firstly, a grid with all possible combinations of input labels is generated, then, outputs are defined removing redundancies and inconsistencies in an iterative process. The last step consists of reducing the set of rules by means of the logical minimization process carried out as part of the semantic cointension analysis.

At the end, the generated FRBCs are made up of linguistic variables and linguistic rules with the usual fuzzy classification rule structure. For crisp classification, the class attached to the rule with highest activation strength is considered. In this work we consider Mamdani FRBCs with "min" function for t-norm, "max" for s-norm and "maximum criteria" for defuzzification.

3.1 Multi-objective Optimization

With the aim of yielding a high quality Pareto set of solutions characterized by good interpretability-accuracy trade-offs, we have chosen the well-known NSGA-II algorithm [4], which represents a generational approach with a multi-objective elitist replacement strategy. Among the possible options, we have chosen a two-point crossover and Thrift mutation.

For both crossover and mutation operations we check that resultant chromosomes have always at least one gene with value greater than one, because at least one input partition must be considered for generating rules.

Our goal is minimizing the following three indexes:

1. *Error*. We minimize the rate of classification error (ERR) which is computed as the percentage of examples that are wrongly classified:

$$ERR = \frac{\sum_{i=1}^{\#DS} (1 - \chi(INF_{RB}(DS_i), C_i))}{\#DS} \tag{1}$$

where $\#DS$ is the cardinality of the dataset, DS_i is the i-th sample, C_i is the observed output value for DS_i, $\chi(a, b)$ returns one if $a = b$, zero otherwise.

$INF_{RB}(DS_i)$ represents the inferred class for the sample DS_i and the rule base RB.

2. *Average Fired Rules.* We minimize AFR which is measured as follows:

$$AFR = \frac{\sum_{i=1}^{\#DS} FR_{RB}(DS_i)}{\#DS - \#NOFR} \tag{2}$$

where $FR_{RB}(DS_i)$ is the number of fired rules of the rule base RB for the sample DS_i, and $\#NOFR$ is the number of samples DS_i such that $FR_{RB}(DS_i) = 0$.

3. *Logical View Index.* We minimize LVI which is computed as follows:

$$LVI = 1 - \frac{\sum_{i=1}^{\#DS} \chi(INF_{BM}(DS_i), INF_{AM}(DS_i))}{\#DS} \tag{3}$$

where BM and AM represents the rule base before and after logical minimization. Hence, we are counting the percentage of cases in which both rule bases are exhibiting a different classification behavior, i.e. they are inferring different output classes (no matter if they are inferring the right class or not).

4 Experimental Analysis

The proposed evolutionary framework has been used for setting up two independent experimental sessions with two objectives: ERR vs. AFR, ERR vs. LVI. For each experiment, we have run NSGA-II [1] with a population of 50 individuals and a stopping condition of 17,000 evaluations. Crossover probability was set to 0.6 while mutation probability was set to 0.1. Notice that, the latter has been set to 0.01 when a gene takes the value one because we want to maximize the number of *Don't Care* conditions with the aim of reducing the number of inputs, and as a side effect, the size of the generated rule bases. Notice that, maximizing DC is likely to increase AFR.

Our experiments deal with the well-known PIMA benchmark classification problem whose data set is freely available at the KEEL[2] machine-learning repository. It contains 768 instances determining if a subject shows signs of diabetes according to the criteria of World Health Organization. Eight numerical input variables have been considered. The output class determines if the diabetes test is negative or positive.

One evaluation means that a FRBC is built for a chromosome and then evaluated according to all defined objectives. FRBC generation is made as described in the previous section using the free software tool GUAJE[3]. First, SFPs are defined for all inputs. Then, rule generation is made using FPA. Afterwards, logical

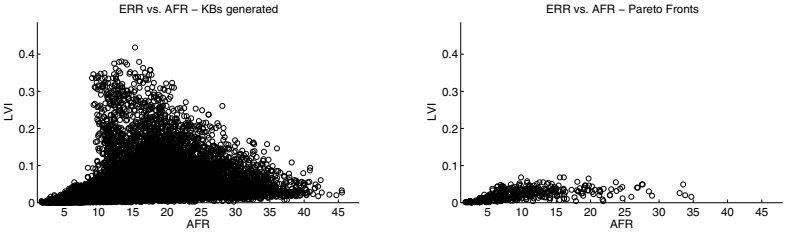
[1] We used the free software jMetal (Metaheuristic Algorithms in Java) [http://jmetal.sourceforge.net/].

[2] A free software tool available online at [http://sci2s.ugr.es/keel/].

[3] A free software tool available at [http://www.softcomputing.es/guaje].

Table 1. Experimental Results

		ERR vs. AFR		ERR vs. LVI	
		Mean	SD	Mean	SD
train	ERR (min range)	0.191	0.006	0.195	0.006
	ERR (max range)	0.261	0.043	0.237	0.025
	AFR (min range)	7.8	3.8	28.6	14.8
	AFR (max range)	44.9	15.4	94.9	21.7
	LVI (min range)	0.011	0.010	0.001	0.001
	LVI (max range)	0.040	0.012	0.021	0.012
test	ERR (min range)	0.233	0.033	0.240	0.043
	ERR (max range)	0.348	0.046	0.333	0.040
	AFR (min range)	5.1	2.5	28.3	15.2
	AFR (max range)	21.3	6.2	94.8	24.3
	LVI (min range)	0.013	0.017	0.001	0.002
	LVI (max range)	0.072	0.027	0.038	0.032
Number of solutions		12.7	4.4	7.0	2.0
Execution time (min.)		64.3	13.5	71.9	15.7



(a) All the generated KBs (b) KBs in the Pareto fronts

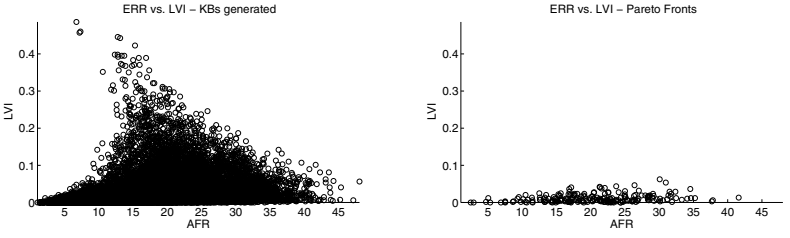
Fig. 1. Relation between AFR and LVI in ERR vs. AFR

minimization is carried out with ESPRESSO[4]. As evaluation methodology we have selected 10-fold cross-validation, executed three times (3x10CV).

Experimental results are shown in table 1. They were carried out by an Intel Xeon E5420 (2.5GHz, four cores) with 8GB of RAM and two hard disks in RAID 1. For each objective the minimum (min range) and the maximum (max range) values have been considered for estimating the objective range. As previously mentioned, we performed a 3x10CV producing 30 Pareto fronts, so mean and standard deviation (SD) values have been provided. For each session, the results for all the three indexes are reported. If an index is not included in a session (i.e. it is not minimized), its values have been computed using the solutions in the Pareto fronts.

In comparing the two experimental sessions, we found quite similar error (ERR) values, both on train and test set. It should be noticed that in ERR vs. AFR a larger number of solutions is found in a shorter time.

[4] Available at [http://www.cs.columbia.edu/~cs4861/s07-sis/]



(a) All the generated KBs (b) KBs in the Pareto fronts

Fig. 2. Relation between AFR and LVI in ERR vs. LVI

Concerning the comparison between AFR and LVI, which is the goal of this work, we found that for all the solutions in the Pareto fronts of the session ERR vs. AFR, LVI values appear to be similar to LVI values in ERR vs. LVI. Hence, it appears that the AFR minimization implies somehow the LVI minimization. On the other hand, for all the solutions in the Pareto fronts of the session ERR vs. LVI, AFR values appear to be large compared to the AFR values in ERR vs. AFR. In this case, minimizing LVI does not imply AFR minimization.

Since this relationship has been observed only on few KBs (solutions in the Pareto fronts), and with the aim of providing more evidence, we computed the three indexes on all the KBs produced through MOEA. In this way, it is possible to compare all the generated solutions with those that are not dominated.

In fig. 1 (a) (ERR vs. AFR) it is shown that all the KBs have wide ranged LVI values, while in fig. 1 (b) the Pareto fronts solutions have remarkable smaller LVI values. As a consequence, the relation holds, i.e. AFR minimization implies LVI minimization. Similarly, figure 2 (ERR vs. LVI) displays that the generated KBs and those in the Pareto fronts share very similar AFR ranges. It is immediate to notice that LVI minimization does not imply AFR minimization.

In our opinion, the empirical relationship we found between AFR and LVI can be partially explained by observing that KBs with small AFR are characterized by quite specialized rules, which are therefore activated only in very limited conditions. This makes logical minimization very limited, thus reducing the LVI value. On the other hand, the minimization of LVI does not depend on the degree of specificity of rules in a KB, hence AFR values are not minimized consequently.

From the point of view of designing highly interpretable rule bases, AFR objective should be used because results have shown a consequent LVI reduction surrounded by shorter execution time, and a larger set of non-dominated solutions.

5 Conclusions

In this work, an empirical study on interpretability indexes through multi-objective evolutionary algorithms has been carried out. A relation between two interpretability indexes, namely AFR and LVI, has been observed. We found

that the AFR minimization implies the LVI minimization, while the opposite is not verified. In our opinion, it could depend on the intrinsic characteristics of the rule bases which have small values of AFR.

This relation should be verified by increasing the number of datasets used in the experimental sessions. As a future work, other interpretability indexes could be taken into account for finding other possible hidden relations.

From the designing perspective, we suggest to adopt the AFR objective in a in a two-objective evolutionary algorithm (ERR vs. AFR) because experimental results have shown that it implies the LVI reduction. Moreover, AFR appears to be less expensive than LVI in terms of computational time, providing at the same time a larger set of non-dominated solutions.

Acknowledgment. This work has been partially funded by the European Centre for Soft Computing (ECSC) located at Mieres (Spain) and the Spanish government (CICYT) under project TIN2008-06890-C02-01.

References

1. Alcalá, R., Ducange, P., Herrera, F., Lazzerini, B., Marcelloni, F.: A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems. Trans. Fuz. Sys. 17, 1106–1122 (2009), http://dx.doi.org/10.1109/TFUZZ.2009.2023113
2. Alonso, J.M., Magdalena, L., Cordón, O.: Embedding hilk in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers. In: Fourth Int. Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS), pp. 15–20 (2010)
3. Alonso, J.M., Magdalena, L., Guillaume, S.: HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. Int. Journal of Intelligent Systems 23(7), 761–794 (2008)
4. Deb, K., Pratap, A., Agarwal, S., Meyarevian, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. on Evolutionary Computation 6(2), 182–197 (2002)
5. Gacto, M., Alcalá, R., Herrera, F.: A double axis classification of interpretability measures for linguistic fuzzy rule-based systems. In: Ninth International Workshop on Fuzzy Logic (accepted) (2011)
6. Glorennec, P.Y.: Algorithmes d'apprentissage pour systèmes d'inférence floue. Editions Hermès, Paris (1999)
7. Herrera, F.: Genetic fuzzy systems: taxonomy, current research trends and prospects. Evolutionary Intelligence 1, 27–46 (2008), http://dx.doi.org/10.1007/s12065-007-0001-5, 10.1007/s12065-007-0001-5
8. Mencar, C., Castiello, C., Cannone, R., Fanelli, A.: Interpretability assessment of fuzzy knowledge bases: A cointension based approach. International Journal of Approximate Reasoning 52(4), 501–518 (2010)

Team Performance Evaluation Using Fuzzy Logic

Mauro Nunes and Henrique O'Neill

Instituto Universitário de Lisboa (ISCTE-IUL), ADETTI-IUL,
Av[a] das Forças Armadas, 1649-026 Lisboa, Portugal
{mauro.nunes,henrique.oneill}@iscte.pt

Abstract. In this paper we describe an experiment where team performance is evaluated by intelligent agents with fuzzy logic reasoning. Although not paramount to the study, which seeks to formally define where and how can intelligent agents help assessing team performance, fuzzy logic was implemented using a set of performance evaluation rules. Results show that the intelligent agents are able to perceive and critically evaluate a team's performance.

Keywords: virtual teams, team performance, intelligent agents, fuzzy logic.

1 Introduction

Researchers have applied agent technology to different areas of teamwork, such as teamwork simulation, team communication and team selection/formation.

Based on a comprehensive review of research about human performance in team settings, Paris et al. [1] maintains that there is still a need to "develop more dynamic measurement systems that allow for on-line assessment of teamwork," particularly because the long-established observational methods (i.e., behavioural checklists or videotaped observation, etc.) have been insufficient to measure teamwork processes [2]. These methods are labour intensive and time consuming, which decreases the speed of analysis and reporting of team performance, becoming unsuitable for large-scale team settings and to fully capture the dynamic nature of teamwork. We believe that agents may offer a value-added contribution to such a measurement system and even extend it, with proactive performance assessment, intelligent analysis and feedback.

It should be noted that our approach is focused on the synergistic relationship between intelligent agents and teams composed by humans. This relationship is established with the purpose of assessing the team's performance. Agents will primarily support teamwork, rather than aiding on individual tasks, following findings from Lenox et al. [3], which suggest that the direct support of teamwork is more beneficial than aiding individual tasks.

To critically assess a team's performance, agents require a set of team performance measures. However, these measures must be adequate and, preferably, linked with a team performance model. Building on research that examined dynamic modelling of team functioning, Marks et al. [4] proposed a time-based conceptual framework of team processes, which can help measure team performance. Agents should use performance measures based on process and outcome

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 139–146, 2011.

variables, which have to be introduced into the reasoning mechanism. This reasoning must take into account not only previously established standard measure values, but also values from existing performance data (history), which will act as a dynamic reference value.

Our study seeks to establish where and how can intelligent agents help assessing a virtual team's performance. We have formalized a possible answer to this question by purposing an agent mediated team performance model – aTeam [5]. We have also implemented the Promus agent system [6], a test bed for evaluating the aTeam model. This paper describes the results obtained when fuzzy logic was used as the reasoning mechanism for evaluating team performance. It should be considered as a by-product of a broader study on team performance and it is not intended as a comparison study between reasoning systems.

1.1 Related Work

From the literature, it is possible to identify work related to this research. The work from Lenox et al. [3] has some similarities to the proposed, notably, the use of agents to support team performance. However, in their study, no geographically dispersed teams are considered, which decreases the communication and coordination effort. The agents also work as specialized facilitators rather than representing each member working as a team, circumscribing the work of the intelligent agents to support and promote teamwork dimensions (Team situation awareness, Support behaviour, Communication and Team initiative/leadership) in a target identification task.

The approach from Chung et al.[2] towards the development of a real-time measurement system for teamwork skills is also related to our approach. They focus their work on assessing team processes used by a group of individuals responsible for jointly constructing a knowledge map. Team processes were measured according to a taxonomy developed from previous work [see 7]; likewise, we also measure team processes according to a similar taxonomy [see 4]. However, their system was not proactive towards team performance, rather just an automatic measurement system. Although a team member identity was anonymous, they were all collocated at the same time, whereas in our study we are imposing team member time dispersion, as it is a key characteristic in a Virtual Team.

The work from Miller and Volz [8] also relates to our study, in the sense that they use a model of teamwork to identify weaknesses in team interaction skills in a human trainee within a team of agents. Although their study's subject is focused on training mixed teams of agents and humans, they state that it could also be used as a tracking or monitoring tool for human teams.

2 Promus Agent System, Beer Game and Fuzzy Logic Application

The agent platform was built based on the generic requirements gathered from past experience and the specific requirements for a team assessment tool from

Paris et al. [1]. It presents a unique set of team-oriented features (based on the aTeam model) like capturing team performance metrics and proactive analysis of team performance. It measures both processes and outcomes in order to ensure consistently effective performance [9].

The Promus agent system is organized in two major components, the Promus Server and Promus Client. The Promus Server design follows an object-oriented approach based on software design patterns.

The Beer Game was selected as a test case, inspired by the approach from Rafaeli and Ravid [10]. The game simulates a distribution system where the team's objective is maximizing net profit by managing inventories appropriately in the face of uncertain demand. Players are linked in a distribution chain, depending on each other for the right flow of orders and goods. Each team member takes a position (Retailer, Wholesaler, Distributor or Factory) in the supply chain. This role-playing simulation of an industrial production and distribution system was originally created as a board game and used to introduce students of management to the concepts of economic dynamics and computer simulation.

Promus agents evaluate performance using relative concepts between the standard high, average and low concepts. Specifically, the agent reasoning is based on fuzzy logic[1] with linguistic variables to evaluate the performance measures. Zadeh [11, 12] was the author of the first publication on fuzzy logic. Since that time, fuzzy logic has been used extensively on a variety of applications. These include, for instance, supply chain management or credit evaluation. In this study, it enabled the agents to show some intelligent behaviour by critically assessing performance within a broad spectrum of concepts to which they were not programmed.

2.1 Team Performance Analysis with Fuzzy Logic

The individual level analysis is conducted through the weighed sum of the performance measure values. This is to differentiate the influence of each performance measure. Formally it can be defined as:

A team (tpx) in a collaborative environment ≡ a set of persons that are working together on a task.

Individual Performance Value (IPVt) ≡ the weighed (cf) sum of the performance values of a team member (tm), at a particular moment t.

Individual Performance Measure Value (IPMVt) ≡ a value for a particular Performance Measure (PM) from a team member, at a particular moment t.

The following equation (1) shows how IPVt was calculated. Please consider #PM as the total number of performance measures.

$$IPVt(tm) = \sum_{i=1}^{\#PM} cf * IPMVt(i). \quad (1)$$

[1] Fuzzy logic was implemented with an Open Source Fuzzy Engine available at: http://people.clarkson.edu/ esazonov/FuzzyEngine.htm

A team member is underperforming within his team (locally) when his IPV is considered lower in comparison with the remaining team members, at a particular moment in time. Because the concept of "how low" is hard to quantify it is specified as a fuzzy function.

Local Underperforming Team Members (LUPTMt) $\equiv \forall tm \in tpx$: is low(IPVt(tm)).

Likewise, global underperforming team members are identified when the respective individual performance values (IPVs) are compared with a reference global performance value (GIPV) from all remaining teams belonging to the team group (TG) being analyzed. For the purpose of this experiment, the GIPV is the average value of all the IPVt (2). Please consider #tm as the total number of team members.

$$GIPVt(tmx) = \frac{\sum_{i=1}^{\#tm} IPMVt(tm)}{\#tm}, \forall tpx \in TG, \forall tm \in tpx \backslash tmx. \quad (2)$$

Instead of trying to combine all the relevant factors into one large fuzzy rule base, the evaluation is decomposed into smaller segments. The decomposition makes the system development manageable and allows evaluation of smaller subcomponents. The following sub-set of fuzzy rules was used for the evaluation of the individual performance data:

```
if performance_value is low then warning is on
if performance_value is avg then warning is off
if performance_value is high then warning is off
```

These rules evaluate the proportional difference between the performance value and a reference performance value. The fuzzy functions low(), avg() and high() are shown in the next graph (Fig. 1).

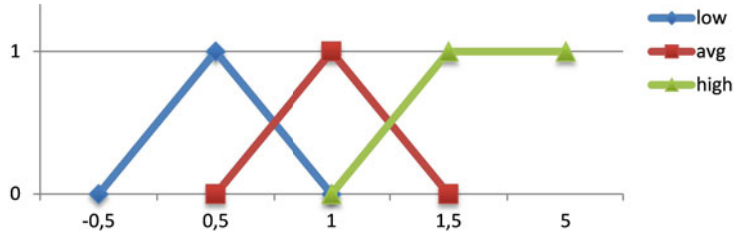


Fig. 1. Performance value fuzzy evaluation functions

The fuzzy functions represented above attempt to provide a reasonable response on the evaluation of the performance measure values, in particular, when these values are at the borderline of each function. For instance, when the performance value is between 0,5 and 1 it will be considered a low and average value. Defuzzification follows the centroid method.

3 Experiment Results

The Promus Beer Game implementation was used in four trials. The first three trials consisted of a challenge for undergraduate (Management) students, where in each trial, 20 participants in five teams played the game at the same time (60 participants in total). These trials were run on the same network of the Promus Server. The last trial consisted of a challenge for postgraduate (Logistics) students and had 12 participants in three teams. This trial was run on a remote location outside the Promus Server network.

Each trial was expected to take on average 1h10m (36 game days), and although the players were in the same room they were not allowed to engage in oral communication. Teams were randomly assembled for the duration of the trial. It was not possible to control if team members already knew each other or had previously worked together as a team. It was the first time that participants played the Beer Game.

3.1 Analysis

Promus Server successfully supported the trials, processing more than 22,500 system messages, generating 12,825 performance data records and 2,192 game plays. Table 1 sums up the performance evaluation of each game for the different trials. Trial t1 was used for platform training and fine-tuning. The number of performance warnings is shown at day 18 (halfway through the game) and day 36 in order to perceive performance evolution.

For trial t2 and t4 the fuzzy evaluation was able to reflect the team's performance by having a higher number of warnings on the games that had the highest total cost.

However, for trial t3 game 80 gets the highest number of warnings instead of game 84. If the outliers (games 80 and 84) were not considered then the performance evaluation would be correct. Looking closer at the performance measures from both games, this may be an example where flawed processes occasionally result in successful outcomes. Meaning that despite having better team processes in game 84, the followed game strategy did not achieve the best results. On the other hand, we also believe that the reasoning and the metrics capturing process can be improved.

3.2 Evaluation

The Promus agents are able to evaluate a team's performance based on the captured performance data. The performance evaluation process, using fuzzy logic, was able to reflect some of the team's performance traces, particularly in the form of performance warnings. While this was acceptable on a simple task like the Beer Game, it is not necessarily the behaviour in a real life situation. Even so, there is some evidence to show that agents are able to evaluate team performance as formalized in the aTeam model.

The Promus agent's performance feedback, based on the aTeam model using a drill-down approach, successfully helped break down a team's performance into

Table 1. Performance warnings per game

		#Warnings		Total Cost		
Trial	Game	Day 18	Day 36	Day 18	Day 36	Evaluation
	20	285		6.320,0 €		Training
	53	294		3.487,5 €		Training
t1	54	287		4.908,0 €		Training
	55	309		3.707,0 €		Training
	56	265		3.450,0 €		Training
	65	215		6.514,5 €	41.659,50 €	
t2	66	299	599	7.572,0 €	87.686,50 €	OK
	67	263	508	6.116,0 €	21.664,50 €	
	80	321	672	3.642,0 €	11.786,00 €	NOK
	81	271	611	3.366,5 €	24.171,50 €	
t3	82	261	561	5.585,0 €	18.079,50 €	
	83	250	582	3.986,0 €	17.789,50 €	
	84	266	501	7.608,0 €	41.965,50 €	
	85	321	686	5.154,5 €	24.217,00 €	
	86	320	675	5.601,0 €	21.815,00 €	
t4	87	341	721	5.385,0 €	47.257,50 €	OK
	88	280	611	3.785,5 €	18.527,50 €	
	89	322	641	8.992,0 €	18.636,00 €	

a set of individual performances and specific performance measures. This has helped not only to identify potentially underperforming teams but also to trace back to the performance of team members, giving some evidence that the agent's performance feedback can help in determining potential underperforming teams. It also reveals that the quality of the performance feedback is very important, i.e., although a consolidated view of the results is necessary, it must also be possible to disaggregate the results into the respective contributing parts.

Limitations. This experiment has some limitations that must be considered before generalising the results. First, the experiment domain and test case do not take into account all the variables and constraints that are present in real life virtual team settings. This simplification however, enables the analysis to focus on the key aspects required by this study.

Having only students as participants can also be considered a limitation, in the sense that they do not fully represent the target population. However, for the purpose of this study, this sample provided enough data for the analysis of team performance within the experiment settings and scope.

4 Conclusions

Even though technology has been used extensively within virtual team settings, it still remains passive towards team performance. Hence, there is a need to

adapt automation to team settings, not only automating the synthesis and interpretation of performance data, but also better measurement tools that can cope with high levels of complexity. This is where the application of the emerging agent technology may help.

Intelligent agent technology has been very helpful in different fields. However, it was unclear where and how this technology could be applied to the problem of managing team performance. This was the motivation for this study.

Unlike normal performance evaluation tools, the Promus agent platform conducts performance evaluation, using fuzzy logic, in real time, providing performance warnings while the task is being carried out. This potentially provides a higher degree of control to a team manager (a target client for this type of application), who can perceive how a team is currently performing and likely to perform in the future, hence acting almost as an early warning system.

Practical applications of this research can be used, for instance, in the evaluation of the performance of operators in a Call Center, where agents monitor the performance of each team of operators and provide real-time feedback to team managers, not only with potentially underperforming operators, but also with the work patterns of those that are better performing.

Acknowledgments. This research work has been partially supported by Associação para o Desenvolvimento das Telecomunicações e Técnicas da Informática (ADETTI) and Lisbon University Institute (ISCTE-IUL).

References

1. Paris, C.R., Salas, E., Cannon-Bowers, J.A.: Teamwork in multi-person systems: a review and analysis. Ergonomics 43, 1052–1075 (2000)
2. Chung, G.K.W.K., O'Neil, H.F., Herl, H.E.: The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. Computers in Human Behavior 15, 463–493 (1999)
3. Lenox, T., Hahn, S., Lewis, M., Roth, E.: Improving performance: Should we support individuals or teams? In: Human Factors and Ergonomics Society 43rd Annual Meeting (HFES), Santa Monica, CA, pp. 223–227 (1999)
4. Marks, M.A., Mathieu, J.E., Zaccaro, S.J.: A temporally based framework and taxonomy of team processes. The Academy of Management Review 26, 356–376 (2001)
5. Nunes, M., O'Neill, H.: aTeam- Agent Mediated Team Performance Model. In: IADIS International Conference WWW/Internet 2004, ICWI 2004, Madrid, Spain (2004)
6. Nunes, M., O'Neill, H.: The Promus Agent System, a tool to assess a virtual team's performance. In: International Conference Applied Computing, pp. 417–424. IADIS, Algarve (2005)
7. O'Neil, H.F.J., Chung, G., Brown, R.: Use of networked simulations as a context to measure team competencies. In: O'Neil, H.F.J. (ed.) Workforce readiness: competencies and assessment, pp. 411–452. Erlbaum, Mahwah (1997)
8. Miller, M.S., Volz, R.A.: Training for Teamwork. In: 5th World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, Florida, pp. 52–57 (2001)

9. Meyers, C.: How the right team measures help teams excel. Harvard Business Review, 95–103 (1994)
10. Rafaeli, S., Ravid, G.: Information sharing as enabler for the virtual team: an experimental approach to assessing the role of electronic mail in disintermediation. Information Systems Journal 13, 191–206 (2003)
11. Zadeh, L.A.: Fuzzy sets. Information and Control, vol. 8, pp. 338–353. Academic Press, New York (1965)
12. Zadeh, L.A.: Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transactions on Systems, Man, and Cybernetics, 28–44 (1973)

Interpretable Fuzzy Modeling for Decision Support in IgA Nephropathy

Marco Lucarelli and Ciro Castiello

Department of Informatics, University of Bari, Italy
marco.lucarelli@uniba.it, castiello@di.uniba.it

Abstract. The aim of the work is to show the potential usefulness of interpretable fuzzy modeling for decision support in medical applications. For this pursuit, we present an approach for designing interpretable fuzzy systems concerning the prognosis prediction in Immunoglobulin A Nephropathy (IgAN). To deal with such a hard problem, prognosis has been granulated into three classes; then, a number of fuzzy rule based classifiers have been designed so that several interpretability constraints are satisfied. The resulting classifiers have been evaluated in terms of classification accuracy (also compared with a standard neural network), some of interpretability indexes, and in terms of unclassified samples. Experimental results show that such models are capable to provide both a first estimation of prognosis and a readable knowledge base that can be inspected by physicians for further analyses.

1 Introduction

Immunoglobulin A Nephropathy (IgAN) is a renal disease: it represents the most frequent primitive glomerulonephritis in the world and one of the main causes of the terminal chronic renal failure[5]. Within about 20 years, 25-30% of subjects with IgAN diagnosis evolve in an End-Stage Renal Disease (ESRD), thus requiring renal transplantation or dialysis replacement. On the one hand, such circumstances imply a number of issues negatively interfering with the patient's living standards; on the other hand, the ESRD treatments represent a considerable cost for National Health Systems in the world. Therefore, prognosis of IgAN is crucial and, even when it cannot be precisely asserted (which is the most frequent case), its prediction is of greatest importance. As a consequence, several research efforts reported in literature have been addressed to the identification of the features useful for predicting the decline of the renal function, thus providing information about the subsequent ESRD (see e.g. [6]).

The need for a predictive model, which is capable to predict the expected time to reach the ESRD, is straightforward. Such a model should rely both on clinical data and other specific measurements for evaluating the disease risk of a subject, so as to make a diagnosis and, more importantly, to propose a prognosis. Commonly, it is not easy to manage this kind of data which is characterized by non-linear distributions and models very complex interactions. Furthermore, the nature of data itself is not precise, so that predictive models, as long as they are

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 147–154, 2011.

based on either classical statistical methods or classical data mining approaches, are likely to fail. Therefore, a somewhat dedicated intelligent analysis is needed.

The complexity of the prediction problem requires the acquisition of knowledge from available data, possibly with a refinement intervention by domain experts. Also, the acquired knowledge must be *interpretable*, so as to be accessible by physicians for further considerations. Natural language description of acquired knowledge represents for sure a valid mean for interpretability and therefore it should be pursued. Such a requirement becomes more stringent when complexity of the knowledge to be communicated is increasing.

The semantics underlying natural language terms embodies the property of fuzziness, hence it can be modeled through fuzzy information granules. In order to process this type of information, the fuzzy set theory provides the necessary machinery to realize information processing. Indeed, fuzzy information granules are represented as fuzzy relations and fuzzy set operators are used to elaborate information.

Interpretability is not a trivial requirement for a model and a number of issues must be addressed to achieve it. Most of the issues are context dependent, therefore it is quite impractical to define once and for all a single approach to obtain an interpretable model. The adoption of a methodology which is bound to a number of interpretability constraints can facilitate the design of an interpretable model. In this paper we propose a new approach aimed at predicting the evolution time of the ESRD in subjects affected by IgAN. The intent is to set up a process of knowledge discovery from data in order to produce an interpretable linguistic knowledge base providing ESRD-related information. To this aim, we employed the GUAJE tool, a free open-source software which is able to ensure a number of interpretability requirements while producing fuzzy models [1]. As a result, we derive fuzzy rule-based classifiers predicting the ESRD evolution time for each patient in an *interpretable* way. In perspective, this may represent a relevant achievement since the obtained information could enable the discovery of new patterns underlying data, useful for the physicians to propose targeted therapies.

The rest of the paper is structured as follows. Section 2 describes the proposed methodology descending into the details related to the dataset settings, the strategies and the design choices. Section 3 reports the performed experimental session and the obtained results. Finally, in section 4 some conclusive remarks and future proposals are discussed.

2 Methodology

The problem we are dealing with is the prediction of the time elapsing between the renal biopsy and the occurrence of the ESRD for a given subject. We move from the assumption that the disease actually manifested itself (the subject is affected by IgAN). In tackling the predictive problem we put as a central tenet the interpretability requirement, so that the derived fuzzy classifiers can be easily understood and adopted by nephrologists. Hence, the models must respect some

kind of interpretability constraints and must be designed by taking in account the knowledge held by the renal specialists.

In the following sections we are going to detail the successive steps we followed in our research activity: first the pre-processing of the available data, then the data partition phase which is preparatory for the next rule-base learning step, finally the improvement of the derived knowledge base.

2.1 Data Pre-processing

The dataset we analyzed has been provided by the Department of Nephrology of the University of Bari and it is composed of 660 entries related to patients affected by IgAN. For each patient, the reported data refer to clinical information collected at different moments: partly during the Renal Biopsy (RB), that is the moment when the pathology is verified, and others during one or more follow-ups (FU), corresponding to subsequent check-up visits. The medical dataset has been subjected to a pre-processing phase in order to arrange the experimentation.

Firstly, it should be recalled that the predictive models are supposed to provide for the evolution time of the ESRD. Therefore, only a subset of data has been involved in our analysis: 98 out of 660 instances of the dataset have been considered, corresponding to the subjects who actually reached the ESRD. Such a drastic reduction was inevitable, since for all the other entries in the dataset no relevant information about ESRD was available.

Pre-processing of data affected also the number of involved features. Studies on the progression of IgAN provide some predictive factors which can be considered as a significant feature subset [4,9]. Accordingly, the initial total number of 43 features has been reduced. Furthermore, special attention has been paid on the RB/FU distinction of clinical data. Data deriving from FU observations are surely relevant for refining ESRD prediction; however, at least in a preliminary stage, we are much more interested in deriving a predictive model which could prove to be useful during the first contacts with the patients, i.e. when only RB information has been collected. This may assist physicians in formulating a very first diagnosis (which could be later adjusted by monitoring the patients during the check-up visits). Therefore, we resolved to exclude FU features from our analysis.

Some other interventions of ours concerned a better specification of the information reported in the dataset. That is the case of a couple of features, namely RB DATE and ESRD DATE, that can be more conveniently integrated into a new feature representing the numbers of years between the RB and the ESRD. Such a piece of information is very significant since, apart from releasing the dataset from unnecessary timing constraints, it embodies the actual object of the prediction to be performed. Therefore, we decided to combine RB DATE and ESRD DATE into the PROGNOSIS feature, representing the expected output of the predictive model. The set of features retained for experimentation is composed by nine elements. Among them, eight are intended as input features for the predictive problem to be tackled (GENDER, AGE AT RB, AGE AT ONSET, ONSET TYPE, GRADE, RB SCR, RB UPR, HYP) and one is intended as the

Table 1. Description of the involved features

Name	Type	Description
GENDER	Categorical	Patient gender (male/female)
AGE AT RB	Numerical	Patient age at the RB time
AGE AT ONSET	Numerical	Patient age when the disease manifested
ONSET TYPE	Categorical	Type of the disease onset (micro/macro hematuria)
GRADE	Categorical	Histological grade of the RB (G1/G2/G3)
RB SCR	Numerical	Creatinina measure at the RB time
RB UPR	Numerical	Proteinuria measure at the RB time
HYP	Categorical	Patient affected by hypertension (yes/no)
PROGNOSIS	Categorical	Evolution time of the ESRD (short/medium/long)

single output feature to be predicted (PROGNOSIS). All the features are briefly described in Table 1.

Since prediction problems can be translated into classification tasks (by partitioning the features that must be predicted), the feature PROGNOSIS has been processed in order to obtain temporal intervals (information granules). Both the number of intervals and their amplitude have been preliminarily discussed with the experts to maximize the semantic underlying each of them. Once verified that no medical purposeful subdivision of the PROGNOSIS feature is traceable, the number of intervals and their amplitude have been automatically evaluated by a clustering process. It has been used the K-Means clustering algorithm to identify three temporal clusters. For each of them, a linguistic label has been proposed which represents the information underlying the granule, and whose semantic is strictly co-intensive with the physicians' knowledge. The resulting linguistic granules defined for the PROGNOSIS feature are: SHORT, identifying subjects whose renal survive is ranging between zero and (about) 5 years; MEDIUM, ranging between (about) 5 and 13 years; LONG, ranging between (about) 13 and 25 years.

In the following sections we are going to describe the successive steps we followed to design the interpretable predictive model, in according with the HILK++ approach proposed in [2]. In each of them, both the interpretability constraints and the experts' knowledge have been considered, so that a highly interpretable model can be obtained.

2.2 Data Partition and Rule-Base Learning

To carry on the experimentation, we employed a tool named GUAJE (Generating Understandable and Accurate fuzzy models in a Java Environment). It is a free open-source software[1] implementing the HILK++ methodology, which is able to ensure a number of interpretability requirements while producing fuzzy models [3]. GUAJE enables the user to follow a step-by-step procedure starting from the design of fuzzy partitions, going through the rule-based learning and ending up with knowledge base improvement. For each step, GUAJE offers a number

[1] GUAJE is avaible at http://www.softcomputing.es/guaje

of alternative methods so that the user can make his best choice depending on the problem and the data at hand.

As concerning our session of experiments, we resolved to generate partitions of the pre-processed data by means of centroids deriving from the well-known K-Means algorithm. As reported in Table 1, the input features involved in the classification task are both numerical and categorical. The latter ones have been modeled by defining a fuzzy singleton for each category. The numerical features underwent the partitioning process: each numerical feature has been partitioned by five fuzzy sets obtained with the K-Means algorithm. Both the fuzzy singletons and the fuzzy partitions have been labeled by linguistic terms whose semantic is strictly co-intensive with the experts knowledge, in order to maximize the interpretability of the resulting model.

Once obtained the fuzzy partitions, a fuzzy rule base must be automatically derived from data. Among the different methods for rule-based learning available in GUAJE, we adopted the Fuzzy Decision Tree (FDT) algorithm, which proved to produce best results both in terms of interpretability and accuracy of the derived models. FDT generates a neuro-fuzzy decision tree from data, firstly obtaining the fuzzy rules by the fuzzy ID3 algorithm (an information-based method) and then applying an algebraic learning method (analogous to artificial neural learning) to facilitate the tuning of fuzzy rules [8].

2.3 Knowledge Base Improvement

During the phase of knowledge base improvement, an iterative refinement process is applied both on partitions and rules. Such an activity can be performed by following a twofold strategy. On the one hand, this is the time for the domain expert to review the knowledge base, in order to apply real world experience to confirm or modify the discovered data relationships. On the other hand, some kind of automatic procedures can be applied for the sake of model refinement. Of course, in this phase accuracy and interpretability of the classifier must be steadily inspected so as to implement the best trade-off between them. The automatic procedures available in GUAJE distinguish between two kinds of improvement: linguistic simplification and partition optimization. In the first case the aim is to produce a global rule base simplification by reducing the number of rules and linguistic variables. An iterative process looks for redundant elements (labels, inputs, rules, etc.) that can be removed without altering the system accuracy (whose worsening should not exceed a predefined threshold). At each iteration the process operates first on rules, then on partitions and stops when no more interpretability improvement is feasible without penalizing the accuracy of the model. Partition optimization is devoted to increase the system accuracy while preserving the previously achieved interpretability. The process realizes a membership function tuning, constrained to keep the strong fuzzy partition property. The strategy for partition optimization is the Genetic Tuning algorithm, which is an all-in-one optimization procedure based on a global search strategy which draws inspiration from the evolutionary processes that take place in nature [7].

Table 2. Confusion Matrix obtained by *leave-one-out cross-validation*

	Inferred				
Observed	short	medium	long	Unclassified cases	Total cases
short	36.8%	7.1%	1%	4.1%	49%
medium	13.3%	13.3%	7.1%	4.1%	37.8%
long	2%	7.1%	4.1%	0%	13.3%
	52%	27.6%	12.2%	8.2%	100%

Both the linguistic simplification and the partition optimization have been applied to derive the predictive models in our experimentation.

3 Experimental Evaluation

The methodology illustrated in the previous section has been applied to derive interpretable fuzzy classifiers starting from the analysis of data. Due to the limited amount of instances inside the reduced dataset, the models have been tested through a leave-one-out cross-validation. The results of the cross-validation are reported in Table 2, under the guise of a confusion matrix.

It should be observed that the obtained models may not be able to provide a class for each input instance. This happens whenever the classifier response lies on the borderline of different classes (i.e. the fuzzy inference produces two outcomes whose numerical difference is below a certain threshold). Such a circumstance is reported in Table 3 ("Unclassified cases"), amounting to 8.2% of cases. The overall value of classification accuracy is 54.1% and the values in the table show how the most critical situations are well managed by the classifiers. In fact, some particular cases, such as a kind of "false positive" (when a short PROGNOSIS value is predicted in front of a long observed value), and — most importantly — a kind of "false negative" (when a long PROGNOSIS value is predicted in front of a short observed value) are reported with the reduced percentage values of 2% and 1%, respectively. As concerning the interpretability of the obtained fuzzy models, Table 3 depicts the representation of a sample rulebase, corresponding to one of those obtained during the cross-validation session. The reported fuzzy rules, expressed by means of natural language terms, lend themselves to be read by physicians in order to assist their prognosis formulation. Additionally, in Table 4 some indexes are shown to provide information about the interpretability of the models. In particular, the table reports the values (averaged over all the models obtained during the cross-validation session) concerning: the number of rules composing the knowledge bases, the total rule length (i.e. the total number of atoms composing all the rules inside a knowledge base), the averaged number of rules firing at each inference, the Logical View index. The latter represents an original non-structural parameter adopted to evaluate interpretability of a knowledge base by considering cointension with the semantics embedded into the fuzzy rules [10].

Table 3. A sample base of rules

Gender	Age at RB	Age at Onset	Onset Type	Grade	RB sCr	RB uPr	Hyp Prognosis
				G1	very low		long
	very low			G2	very low		long
	average		mH	G2	very low		long
	low			G2	very low		medium
	high			G2	very low		medium
				G3	very low	very low OR low	medium
	low OR average				low	very low	medium
	low OR average	low			low	low	medium
		very high			low	low	medium
	average				low	average	medium
					low	very high	medium
		average	MH		average		medium
	high	average	mH		average		medium
				G3	very low	average OR high	short
	high OR very high				low	very low	short
	very low	low OR average			low	low	short
male	low	average			low	low	short
	average OR high	average			low	low	short
		high			low	low	short
	very low OR low OR average	average	mH		average		short
		high OR very high			average		short
					high OR very high		short

Table 4. Interpretability indexes

Rules	Total rule length	Avg. fired rules	Logical View Index
28.4	93.9	3.6	0.9

The implementation of an Artificial Neural Network (ANN) has been adopted as a tool of comparison: we aimed at comparing the obtained interpretable fuzzy results with those deriving from a black-box model supplied with the same input instances. In particular, the ANN is organized in two feed-forward connection layers, with eight input nodes (corresponding to the input features), one output node (yielding a crisp number of years concerning the evolution time of the ESRD) and 24 hidden nodes. In order to set up a fair comparison, a granulation process has been applied over the output values of the network (so that outcomes are comparable with those provided by the fuzzy inferences). Moreover, the fuzzy inference models have been forced to provide an output for the unclassified cases too (thus yielding an accuracy value equal to 56.12%). In this way, all the data are considered for evaluation, similarly to the evaluation of the network performed on the basis of a 10-fold cross-validation. The ANN accuracy value is 53%: it is slightly lower than the value produced by the fuzzy classifying models which provide physicians with the interpretability added value.

4 Conclusions

The application of a peculiar methodology for deriving interpretable predictive models to a medical prognostic problem has been described in this paper. The problem concerns the prediction of the evolution time of the ESRD for IgAN,

i.e. the progress of a renal disease which leads the patient to undergo renal transplantation or dialysis replacement. The task is not easy and a number of factors are involved in the prognosis, but the availability of real data (supplied by a medical institution) let us set up an experimentation. The data were subjected to a thorough analysis and a pre-processing phase was performed to construct a significant dataset. In order to build an interpretable fuzzy model from it, we followed the steps of the HILK++ methodology which allows to derive fuzzy partitions from data and to learn a rule-based knowledge which can be finally refined to improve readability. In practice, we employed GUAJE, a tool which implements the HILK methodology and offers the possibility to choose among a number of methods to derive the predictive models by following the previously mentioned steps.

The fuzzy classifiers obtained at the end of the experimental session proved to be quite interpretable and their performance is comparable with the accuracy value exhibited by a well established classifier model — an artificial neural network acting as a black box — constructed and evaluated on the basis of the same dataset.

References

1. Alonso, J.M., Magdalena, L.: GUAJE - a java environment for generating understandable and accurate models. In: Peregrin, A. (ed.) XV Spanish Conference for Fuzzy Logic and Technology (ESTYLF), vol. 1, pp. 399–404. University of Huelva Press (2010)
2. Alonso, J.M., Magdalena, L.: HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. Soft Computing, 1–22 (2010)
3. Alonso, J.M., Magdalena, L., Guillaume, S.: KBCT: a knowledge extraction and representation tool for fuzzy logic based systems. In: Proceedings of the 2004 IEEE International Conference on Fuzzy Systems, vol. 2, pp. 989–994 (July 2004)
4. Bartosik, L.P., Lajoie, G., Sugar, L., Cattran, D.C.: Predicting progression in IgA nephropathy. American Journal of Kidney Diseases 38(4), 728–735 (2001)
5. Beerman, I., Novak, J., Wyatt, R.J., Julian, B.A., Gharavi, A.G.: The genetics of IgA Nephropathy. Nature Clinical Practice Nephrology 3, 325–338 (2007)
6. Coppo, R., D'Amico, G.: Factors predicting progression of IgA Nephropathies. Journal of Nephrology 18(5), 503–512 (2005)
7. Cordón, O., Herrera, F.: A three-stage evolutionary process for learning descriptive and approximate fuzzy-logic-controller knowledge bases from examples. International Journal of Approximate Reasoning 17(4), 369–407 (1997)
8. Ichihashi, H., Shirai, T., Nagasaka, K., Miyoshi, T.: Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. Fuzzy Sets and Systems 81, 157–167 (1996)
9. Manno, C., Strippoli, G.F.M., D'Altri, C., Torres, D., Rossini, M., Schena, F.P.: A novel simpler histological classification for renal survival in IgA Nephropathy: A retrospective study. American Journal of Kidney Diseases 49(6), 763–775 (2007)
10. Mencar, C., Castiello, C., Cannone, R., Fanelli, A.M.: Interpretability assessment of fuzzy knowledge bases: A cointension based approach. International Journal of Approximate Reasoning (2010) (in Press), http://dx.doi.org/10.1016/j.ijar,11.007

Experimental Comparative Study of Compilation-Based Inference in Bayesian and Possibilitic Networks

Raouia Ayachi[1,2], Nahla Ben Amor[1], and Salem Benferhat[2]

[1] LARODEC Laboratory, ISG, University of Tunis, Tunisia, 2000
[2] CRIL-CNRS, University of Artois, France, 62307

Abstract. Graphical models are important tools for representing and analyzing uncertain information. Diverse inference methods were developed for efficient computations in these models. In particular, compilation-based inference has recently triggered much research, especially in the probabilistic and the possibilistic frameworks. Even though the inference process follows the same principle in the two frameworks, it depends strongly on the specificity of each of them, namely in the interpretation of handled values (probability\possibility) and appropriate operators (*\min and +\max). This paper emphasizes on common points and unveils differences between the compilation-based inference process in the probabilistic and the possibilistic setting from a spatial viewpoint.

Keywords: Bayesian networks, Qualitative graphical models, Possibility theory, Compilation-based inference.

1 Introduction

Graphical models are a powerful family of models for representing and analyzing uncertain information. They are characterized by their explicitness and clarity. *Bayesian networks* are studied under the broader class of probabilistic graphical models. However, the probability theory in such models is only appropriate when all numerical data are available, which is not always possible. Several non-classical theories of uncertainty have been proposed in order to deal with uncertain and imprecise data. We are in particular interested in *possibility theory* [8,9]. The last decade has seen a virtual explosion of applications of propositional logic. One emerging application is *knowledge compilation*. It consists in preprocessing the propositional theory only once in an off-line phase, with the goal of making frequent on-line queries efficient [2]. One of the most prominent successful applications of knowledge compilation is in the context of graphical models. In fact, in [3], authors focused on compiling Bayesian networks using DNNFs. In [1], we studied the possibilistic adaptation of some compilation-based inference methods using Π-DNNFs. The objective behind these methods is to ensure an efficient computation of a-posteriori probability or possibility

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 155–163, 2011.

degrees given some evidence on some variables. We intend that is considerable to accomplish a comparison study in which we unveil the most compact framework in this context. In this paper, we investigate the extent to which possibility theory can be used to reduce sizes of compiled bases. In fact, we will compare the probabilistic and the possibilistic approaches and prove the importance of the possibilistic setting versus the probabilistic setting. The remaining paper is organized as follows: Section 2 presents a brief refresher on possibility theory and compilation. Section 3 describes both the probabilistic and the possibilistic approaches by focusing on similarities and differences between the two settings. Experimental study is presented in Section 4. Section 5 concludes the paper.

2 Basic Concepts

Let $V = \{X_1, X_2, ..., X_N\}$ be a set of variables. We denote by D_{X_i} the domain associated with the variable X_i. By x_i (resp. x_{ij}), we denote any of the instances of X_i (resp. the j^{th} instance of X_i). When there is no confusion we use x_i to mean any instance of X_i. In the n-ary case, $D_{X_i} = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$ where n is the number of instances of X_i. By v we denote instantiations of all variables $X_i \in V$. Ω denotes the universe of discourse, which is the cartesian product of all variable domains in V. Each element $\omega \in \Omega$ is called an *interpretation*, a possible *world* or a *state* of Ω. $\omega[X_i] = x_i$ denotes an instantiation of X_i in ω.

In this paper, we are interested in two uncertainty frameworks, namely the standard one, i.e., the probabilistic setting and the non-standard possibility theory [8,9]. The basic building block in this theory is the concept of *possibility distribution* π, which is a mapping from the universe of discourse Ω to the unit interval $[0, 1]$ such that $\pi(\omega) = 1$ means that the realization of ω is totally possible and $\pi(\omega) = 0$ means that ω is an impossible state. In the extreme case of total ignorance, $\pi(\omega) = 1, \forall \omega \in \Omega$. It is generally assumed that there exists at least a state ω which is totally possible. In this case, π is said to be normalized. From π, we can compute two dual measures $\Pi(\phi) = max_{\omega \in \phi} \pi(\omega)$ and $N(\phi) = 1 - \Pi(\neg\phi)$ evaluating respectively to which extent ϕ is consistent with the knowledge represented by π and to which level ϕ is certainly implied by this knowledge. Contrarily to the probabilistic case where $P(\neg\phi) = 1 - P(\phi)$, possibility and necessity measures are weakly linked. In possibility theory, conditioning is defined by the following counterpart of the Bayesian rule: $\forall \omega,\ \pi(\omega) = \min(\pi(\omega \mid \psi), \Pi(\psi))$. $\pi(\omega \mid \psi)$ and $\Pi(\psi)$ are combined using a min operation, according to the *ordinal* interpretation of the possibilistic scale[1]. In what follows, we will use some generic notations, i.e., the conjunctive operator $\otimes$ corresponding to $\prod$ and min, the disjunctive operator $\oplus$ corresponding to $\sum$ and max, in the probabilistic and the possibilistic case, resp. $\propto$ denotes the probability or the possibility degree depending on the setting. $\oslash = \{B, \Pi\}$, i.e., if we use $\oslash = B$ we mean the probabilistic setting and if use $\oslash = \Pi$ we mean the possibilistic case.

[1] The *numerical* interpretation of possibility theory uses the product instead of the min, but this is out the scope of the present study.

2.1 Bayesian and Possibilistic Networks

Bayesian and min-based graphical models (denoted by G^B and G^{Π}, resp.) share the same graphical component, i.e., a DAG where nodes represent variables in V and edges encode different (in)dependence relationships. The major difference resides in their numerical component since in G^B different links are quantified via probability distributions while in G^{Π}, we use possibility distributions. Formally, each node X_i in a G^B (resp. G^{Π}) is represented by a local normalized probability (resp. possibility) distribution in the context of its parents, denoted by $U_i = \{U_{i1}, U_{i2}, ..., U_{im}\}$ where m is the number of parents of X_i. In what follows, we use x_i, u_i, u_{ij} to denote, resp. possible instances of X_i, U_i and U_{ij}.

The set of a priori and conditional probability (resp. possibility) distributions induces a unique joint distribution via a chain rule based on the product in the probabilistic setting and min in the possibilistic setting [11].

2.2 Compilation

A logical form qualifies as a *target compilation language* if it supports some set of nontrivial *transformations* and *queries* in polynomial time with respect to the size of compiled bases [7]. We will review in this section the target compilation languages relevant to the present paper. The *Decomposable Negation Normal Form (DNNF)* is a universal language presenting a number of properties that makes it tractable and of a great interest. DNNF, which is qualified as succinct, is an *Negation Normal Form (NNF)* language satisfying the *decomposability* property stating that: conjuncts of any conjunction do not share variables [4]. A set of important properties may be imposed to DNNF, for instance, *determinism* and *smoothness* giving rise to the sd-DNNF. The DNNF compilation language (or one of its variants) supports a rich set of polynomial-time logical operations [7]. For queries, within the most common queries, we cite *model counting*. For transformations, we focus on *conditioning*, *forgetting* and *minimization*:

- *Conditioning*: Let α be a propositional formula. Let ρ be a consistent term, then conditioning α on ρ, denoted by $\alpha|\rho$ generates a new formula where each variable P_i of α is replaced by $\top$ if P_i is consistent with ρ, and by $\bot$ otherwise.
- *Forgetting*: Let α be a propositional formula, let P be a finite set of propositional variables P_i, then the forgetting of P from α, denoted by $\exists P.\alpha$ is a formula that does not mention any variable P_i from P.
- *Minimization*: Let α be a propositional formula, then the minimization of α is a formula β such that the cardinality of all β's models is equal to the minimum cardinality models of α [4]. Recall that the cardinality of a model corresponds to the number of variables set to True ($\top$) or False ($\bot$).

Π-DNNF [1] is a possibilistic version of DNNF in which conjunctions and disjunctions are substituted by minimum and maximum operators, respectively. It is considered as a special case of VNNFs [10].

3 DNNF vs Π-DNNF

It was shown recently that compiling Bayesian networks corresponds to factoring multi-linear functions [5]. In the possibilistic framework, we have shown that compiling possibilistic networks corresponds to factoring possibilistic functions [1]. In this section, we will propose a generic approach handling both probabilistic and possibilistic settings to raise awareness about these approaches and reveal the differences between them.

3.1 From Encoding to Inference

The multi-linear (resp. possibilistic) function $f^{\oslash}$ of a network $G^{\oslash}$ contains two types of propositional variables. An *indicator variable* λ_{x_i} is associated for each value x_i of each X_i in $G^{\oslash}$. Furthermore, a *parameter variable* $\theta_{x_i|u_i}$ is associated for each network parameter $\propto (x_i|u_i)$ in $G^{\oslash}$. Equation (1) expresses the generic function $f^{\oslash}$ s.t. $\oplus$ denotes maximum or probabilistic sum and $\otimes$ denotes minimum or product depending on the setting.

$$f^{\oslash} = \bigoplus_{v} \bigotimes_{(x_i,u_i)\sim v} \lambda_{x_i}\theta_{x_i|u_i} \tag{1}$$

To compute the probability or the possibility of an evidence e, $f^{\oslash}$ should be evaluated after setting appropriate values to indicator variables depending on e. In both settings, $f^{\oslash}$ has an exponential size, so it should be interesting to encode such function using a propositional theory to represent it more compactly. The CNF propositional language is chosen since it is often convenient for compactly encoding knowledge bases [7].

Definition 1. *Let $G^{\oslash}$ be a network $(\oslash = \{B, \Pi\})$, $\lambda_{x_{ij}}$, $(i = 1, \ldots, N)$,$(j = \{1, \ldots, n\})$ be the set of evidence indicators and $\theta_{x_i|u_i}$ be the set of parameter variables, then $C^{\oslash}$ should contain the following clauses:*

- $\forall$ $X_i \in V$, $C^{\oslash}$ *contains the following two clauses (named indicator clauses):*

$$\lambda_{x_{i1}} \vee \lambda_{x_{i2}} \vee \cdots \lambda_{x_{in}} \tag{2}$$

$$\neg\lambda_{x_{ij}} \vee \neg\lambda_{x_{ik}}, j \neq k \tag{3}$$

- $\forall$ $\theta_{x_i|u_i}$ *s.t* $u_i = \{u_{i1}, u_{i2}, ..., u_{im}\}$, $C^{\oslash}$ *contains the following clauses:*

$$\lambda_{x_i} \wedge \lambda_{u_{i1}} \wedge \ldots \wedge \lambda_{u_{im}} \rightarrow \theta_{x_i|u_i} \tag{4}$$

$$\theta_{x_i|u_i} \rightarrow \lambda_{x_i} \tag{5}$$

$$\theta_{x_i|u_i} \rightarrow \lambda_{u_{i1}}, \cdots, \theta_{x_i|u_i} \rightarrow \lambda_{u_{im}} \tag{6}$$

Clauses (2) and (3) state that indicator variables are exclusive, while clauses (4)-(6) encode network's structure. Once $f^{\oslash}$ is represented as $C^{\oslash}$, a compilation step is required to prepare for answering efficiently a large number of inference queries. This process depends on the uncertainty framework. Let $C_c^{\oslash}$ be the compilation result of $C^{\oslash}$. Let x be an instantiation of some variables $X \subseteq V$, then computing $\propto (x)$ using $C_c^{\oslash}$ is ensured as follows:

1. Conditioning $C_c^{\oslash}$ on x by setting each λ_{x_i} to $\bot$ if $\exists\ x_j \in x$ s.t. x_j and x_i disagree on values (i.e., $x_i \nsim x$), and to $\top$ if $x_i \sim x$.
2. Decoding $C_c^{\oslash}$ to have a valued expression, denoted by $\oslash$-circuit (Definition 2),
3. Computing $\propto (x)$ using $\oslash$-circuit.

Definition 2. *A $\oslash$-circuit is a DAG with internal nodes are labeled with $\oplus/\otimes$ and leaves are labeled with propositional variables.*

The first step allows us to exclude terms incompatible with x. The second step decodes the compiled base $\oslash$-circuit depending on the framework. The third step ensures an efficient computation of $\propto (x)$ by applying some query or transformation supported by some target compilation language. In the probabilistic case, inference problems have been effectively translated into *model counting* problems. According to the knowledge map of [7], the appropriate language is the sd-DNNF, which is less succinct than DNNF [7]. In the possibilistic case, inference corresponds to *max-variable elimination* (forgetting using the max), hence the language should support both max-variable elimination and conditioning. The honored language is Π-DNNF [1].

3.2 Which is the Most Compact Method?

In the previous subsection, we have only focused on network's structure and variable's domains and we have not explored parameters values, i.e., the so-called *local structure* which refers to a structure that can be inferred from the specific values of network parameters. Encoding local structure into logic has been under investigation in both probabilistic [3] and possibilistic settings [1]. We are in particular interested in parameters 0, 1 and equal parameters within CPTs. Incorporating local structure into $C^{\oslash}$ differs depending on the framework. In what follows, we will study in depth each case for both frameworks and reveal the differences between the two settings.

- *Parameters equal to 0*: Consider the parameter $\theta_{b_1|a_2} = 0$ which generates the three clauses of Definition 1. Given that this parameter is known to be 0, all terms that contain this parameter must vanish using either $\prod$ or min. Therefore, we can drop it from the encoding and replace its clauses by a shorter clause involving only indicator variables as follows: $\neg\lambda_{x_i} \vee \neg\lambda_{u_{i1}} \vee \cdots \vee \neg\lambda_{u_{im}}$.
- *Equal parameters*: Parameter equality is exploited to reduce at least the number of propositional variables. Due to the fact that no two parameters in the same CPT can ever appear in the same $f^{\oslash}$'s term, the same propositional variable can be used to represent multiple parameters within the same CPT. However, such simplification cannot be applied directly since an inconsistent family instantiations will be evoked at the level of clause (5) and clause (6). The idea is to drop these clauses from the encoding, which introduces additional models into the CNF, allowing multiple parameters from

the same CPT in $f^{\oslash}$'s terms. These unintended models, having a cardinality higher than the cardinality of original models (i.e., $2N$: N evidence indicators and N network parameters), can be filtered by applying the *minimization* transformation that can be ensured in polytime by DNNF [4].

– *Parameters equal to 1*: Let $\theta_{x_i|u_i}$ be a parameter equal to 1. In the probabilistic case, we can omit this parameter and its associated clauses from C^B [5], while in the possibilistic case, this is not the case since the parameter 1, used to satisfy the normalization constraint, is not qualified as a particular value. Indeed, it is a fundamental parameter as it appears several times in each CPT. Hence, in the possibilistic case, parameters equal to 1 should be considered as a set of equal parameters within CPTs.

By exploiting local structure, the CNF encoding $C^{\oslash}$ and its corresponding circuit $\oslash$-circuit should be smaller, especially in the possibilistic case. In fact, the normalization constraint relative to the possibilistic network offers the opportunity to incorporate less propositional variables, encode less CNF clauses, and consequently construct more compact Π-circuits w.r.t. B-circuits, which proves the importance of the possibilistic setting versus the probabilistic setting.

Proposition 1. *Let Nb_{poss} and Nb_{proba} be the number of variables/clauses in the possibilistic and probabilistic cases, respectively. Then $Nb_{poss} \prec Nb_{proba}$.*

Example 1. *Let us consider the bayesian and the possibilistic networks of Fig. 1.*

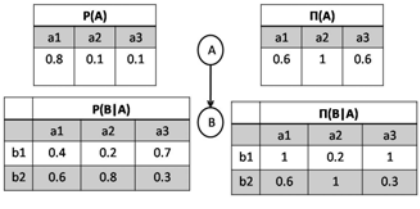


Fig. 1. A bayesian and a possibilistic network

We stress that the number of propositional variables and clauses in C^{Π} is less than than those of C^B. In fact, there are 13 propositional variables and 28 clauses in C^B, while in C^{Π} there are only 11 variables and 22 clauses. The compilation of C^B and C^{Π} give us the following B-circuit and Π-circuit represented by figures 2 and 3, respectively s.t. θ_1 encodes the probability degree 0.1 and θ_1 (resp. θ_2) encode the possibility degree 0.6 (resp. 1). It is prominent that the number of nodes/edges of the Π-circuit is less than those of the B-circuit. Indeed, the number of nodes/edges in the probabilistic (resp. possibilistic) case is equal to 46/66 (resp. 41/60). These numbers include negated propositional variables. After evacuation of such parameters, the resulting compiled bases, which are represented by figures 2 and 3, contain 30/34 (resp. 25/31) nodes/edges in the probabilistic (resp. possibilistic) framework.

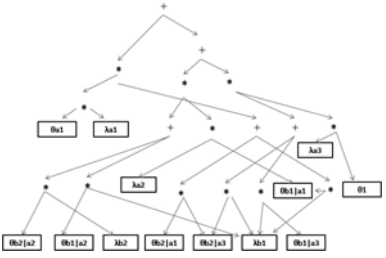


Fig. 2. The B-circuit

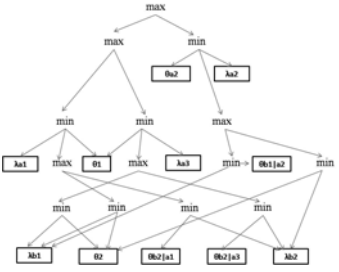


Fig. 3. The Π-circuit

4 Experimental Study

In this section, we will compare the probabilistic DNNF and the possibilistic Π-DNNF approaches. The objective behind this study is to highlight the extent to which possibility theory can reduce sizes of compiled bases. Our experimentation is performed on random bayesian and possibilistic networks generated as follows:

- *Graphical component*: DAGs are generated randomly, by just varying two parameters: the number of nodes and the maximum number of parents per node.
- *Numerical component*: Once the DAG structure is fixed for both bayesian and possibilistic networks, we generate random conditional probability and possibility distributions of each node in the context of its parents, with taking into consideration equal parameters and parameters equal to 0/1. These parameters will be stated as $\%EP$ the percent of equal parameters within the same CPT and $\%ExP$ the percent of extreme parameters within the same CPT. These two values give an idea on the amount of local structure within CPTs.

For each experimentation, we set $\%ExP$ to three values which are 10%, 50% and 90%. For each instantiation of $\%ExP$, we set $\%EP$ to 10%, 30%, 50% and 70%. For each instantiation of these parameters, we generate 100 random bayesian and possibilistic networks having a number of nodes equal to 50 and a maximum number of parents per node equal to 4. For each method, we will compare sizes of CNF encodings in terms of both number of variables and clauses and sizes of compiled bases in terms of both number of nodes and edges. Note that we have used the c2d compiler [6] developed by Darwiche. Each pair of the following figures show the behavior of DNNF and Π-DNNF for each instantiation of $\%EP$ and $\%ExP$. From figures 4, 6 and 8, we can deduce that the number of variables and clauses fall down in both DNNF and Π-DNNF for each instantiation of $\%EP$. Furthermore, it is prominent that Π-DNNF uses less variables and clauses comparing to DNNF. This is due to the normalization constraint offered by the possibilistic setting. Hence, the higher the value of $\%EP$ and $\%ExP$, the better the quality of results (i.e., the lower number of CNF variables and CNF clauses). By just taking a careful look at the scale of figures 4, 6 and 8

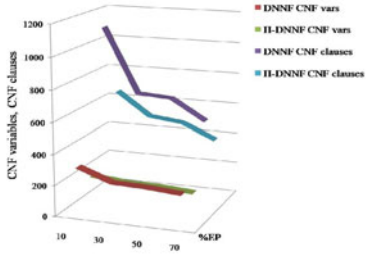


Fig. 4. Variables and clauses for $ExP = 10\%$

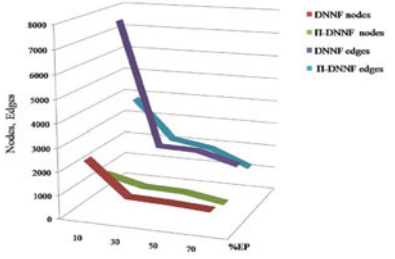


Fig. 5. Nodes and edges for $ExP = 10\%$

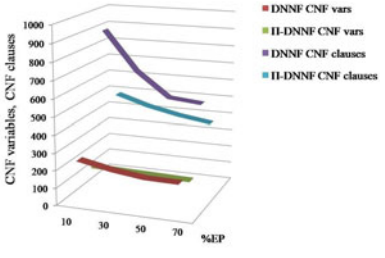


Fig. 6. Variables and clauses for $ExP = 50\%$

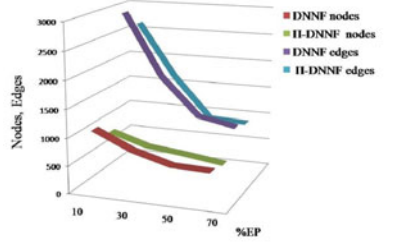


Fig. 7. Nodes and edges for $ExP = 50\%$

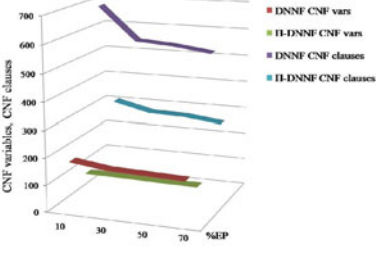


Fig. 8. Variables and clauses for $ExP = 90\%$

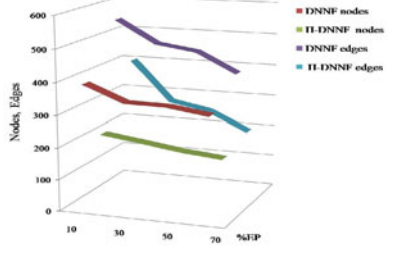


Fig. 9. Nodes and edges for $ExP = 90\%$

which decreases from 1200 to 700 through 1000, we can confirm this key result. Regarding compiled bases parameters, we can notice from figures 5,7 and 9 that Π-DNNF is characterized by a lower number of nodes and edges comparing to those of DNNF. Indeed, the higher the value of $\%EP$, the lower number of nodes and edges for both methods, especially for Π-DNNF. We should also pinpoint that compiled bases parameters and CNF parameters follow the same behavior since by increasing $\%EP$ or $\%ExP$, compiled bases parameters fall down, which is also the case for CNF parameters. Hence, we can conclude that Π-DNNF performs better than DNNF in terms of both CNF parameters and compiled bases parameters, which allows us to make up some extra space.

5 Conclusion

This paper proposed a generic compilation-based inference approach handling both probabilistic and possibilistic settings. We focused on theoretical in common points between the two approaches and unveil the differences between them. Furthermore, we studied the so-called local structure in both approaches. We theoretically show that Π-DNNF is more compact than DNNF, in terms of both number of CNF variables/clauses and nodes/edges of compiled bases, which proves the importance of the possibilistic setting versus the probabilistic setting. These results were confirmed by experimental results.

References

1. Ayachi, R., Ben Amor, N., Benferhat, S., Haenni, R.: Compiling possibilistic networks: Alternative approaches to possibilistic inference. In: Proceedings of 26th Conference on UAI, pp. 40–47. AUAI Press (2010)
2. Cadoli, M., Donini, F.M.: A survey on knowledge compilation. AI Communications–The EJAI (10), 137–150 (1998)
3. Chavira, M., Darwiche, A.: Compiling bayesian networks with local structure. In: Proceedings of the 19th IJCAI, pp. 1306–1312 (2005)
4. Darwiche, A.: Decomposable negation normal form. Journal of the ACM 48(4), 608–647 (2001)
5. Darwiche, A.: A logical approach to factoring belief networks. In: Proceedings of KR, pp. 409–420 (2002)
6. Darwiche, A.: New advances in compiling CNF to decomposable negational normal form. In: Proceedings of ECAI, pp. 328–332 (2004)
7. Darwiche, A., Marquis, P.: A knowledge compilation map. Journal of Artificial Intelligence Research 17, 229–264 (2002)
8. Dubois, D., Prade, H.: Possibility theory:An approach to computerized, Processing of uncertainty. Plenium Press, New York (1988)
9. Dubois, D., Prade, H.: Possibility theory. In: Meyers, R.A. (ed.) Encyclopedia of Complexity and Systems Science, pp. 6927–6939. Springer, Heidelberg (2009)
10. Fargier, H., Marquis, P.: On valued negation normal form formulas. In: Proceedings of the 20th IJCAI, pp. 360–365 (2007)
11. Fonck, P.: Propagating uncertainty in a directed acyclic graph. In: Valverde, L., Bouchon-Meunier, B., Yager, R.R. (eds.) IPMU 1992. LNCS, vol. 682, pp. 17–20. Springer, Heidelberg (1993)

Tuning Graded Possibilistic Clustering by Visual Stability Analysis

Stefano Rovetta[1,⋆], Francesco Masulli[1,2], and Tameem Adel[3]

[1] Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova, Italy
rovetta@disi.unige.it
[2] Center for Biotechnology, Temple University, Philadelphia, USA
[3] Computer and Systems Engineering Dept., University of Alexandria, Alexandria, Egypt

Abstract. When compared to crisp clustering, fuzzy clustering provides more flexible and powerful data representation. However, most fuzzy methods require setting some parameters, as is the case for our Graded Possibilistic c-Means clustering method, which has two parameters in addition to number of centroids. However, for this model selection task there is no well established criterion available. Building on our own previous work on fuzzy clustering similarity indexes, we introduce a technique to evaluate the stability of clusterings by using the fuzzy Jaccard index, and use this procedure to select the most suitable values of parameters. The experiments indicate that the procedure is effective.

Keywords: Possibilistic clustering, Clustering methods, Fuzzy clustering, Fuzzy statistics and data analysis.

1 Introduction

When compared to standard (crisp) clustering, fuzzy clustering provides a more flexible and powerful data representation paradigm. However, most fuzzy clustering methods require setting some parameters, in addition to the number of centroids, which often play the role of degrees of fuzziness. For instance, Bezdek's fuzzy c-means [1] requires to set an exponent m to control fuzziness, and Krishnapuram and Keller's possibilistic c-means [12] needs a set of width parameters β_j, one per cluster. In [14], we have proposed a *graded possibilistic c-means* clustering technique (GPCM) that provides control over the degree of possibility, thus allowing a soft transition between the standard probabilistic and the possibilistic models. This is done through an additional parameter α. No criterion was provided in the original work for setting its value.

In the context of clustering, stability is an important quality criterion, for lack of supervised information allowing objective evaluation. Robust clustering has been the subject of many studies [17]. Intuitively, clusters that are less sensitive to perturbations in the data (e.g., adding/removing data points, outliers, and adding noise to point coordinates) or to variations in the clustering parameters (number of centroids, initialization, other model parameters) are more likely to reflect properties of the data themselves, while clusters subject to greater variability may be due to finite sample effects. The

⋆ Corresponding author.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 164–171, 2011.

purpose of the present work is to introduce a technique to evaluate the stability of clusterings by using the fuzzy Jaccard index, and to use this procedure to select the most suitable values of parameters in a specific clustering method.

To assess and possibly improve the stability of a clustering, several approaches have been proposed. The topic has a long history; the first methods were developed directly in the realm of cluster analysis [18], whereas more recently concepts from robust statistics like breakdown analysis and influence functions [10] have been introduced in specific clustering methods [8,4,7]. However these approaches are specific to a given method, and are not generally applicable.

In contrast, methods exploiting direct comparison between clusterings, e.g., [5], allow an application to more general classes of methods. In [20] we have proposed the generalization of several indexes of similarity between partitions to the fuzzy case. Most fuzzy similarity indexes have the additional property that the level of fuzziness in the partitions being compared is reflected in the maximum attainable value of the index. This is true also for possibilistic clusters, where an added feature is that the "best" clusterings are not only the stablest, but also those with the highest degree of self-similarity. The approach proposed by Hennig in [9] compares crisp clusters using the Jaccard index. This method is related to our proposal, with two important differences: first, we build on previous work [20,3] and use a method for comparing *fuzzy* clusterings, while Hennig's approach is strictly tied to the availability of crisp memberships; second, on the other hand, our method cannot directly compare individual clusters, because these are not clearly defined in the fuzzy case, and compares whole clusterings instead.

In the remainder of this paper, Section 2 introduces the clustering method and Section 3 the clustering similarity measure; then in Section 4 we present the method and in Section 5 the experimental verification.

2 The Graded Possibilistic c-Means Clustering Method

The fuzzy central clustering paradigm is implemented in several, diverse algorithms. In this work we focus on methods based on the following definition of the c cluster centroids Y:

$$\mathbf{y}_j = \frac{\sum_{k=1}^{c} u_{jk}\mathbf{x}_k}{\sum_{k=1}^{n} u_{jk}}, \tag{1}$$

as in the c-Means methods, with u_{jk} membership of data point x_k to the cluster having centroid $\mathbf{y}_j$. Since the memberships will be computed in different ways, the resulting centroids are not the same for all methods. One taxonomy is based on the value of the sum of all membership for any given data point. When this sum is constrained to 1, we are in the standard ("probabilistic") case; when it is essentialy unconstrained, we are in the possibilistic case.

The graded possibilistic membership model presented in [14] takes an in-between approach by constraining the memberships to obey an interval equality, so that when the equality holds the model is effectively possibilistic, whereas, for other combinations of values, a compensation is introduced to constrain the memberships to each other.

We start from a membership function expressed as:

$$u_{jk} = \frac{e^{d_{jk}/\beta}}{Z_k}, \tag{2}$$

where $Z_k = \sum_{l=1}^{c} e^{u_{lk}/\beta}$ is called the partition function.

We define the *free membership* as:

$$v_{jk} = e^{-d_{jk}/\beta_j}, \tag{3}$$

as in the ME and PCM-II algorithms. Then we define the partition function Z_k as:

$$\begin{aligned} Z_k &= \textstyle\sum_{j=1}^{c} v_{jk} && \text{if} \quad \textstyle\sum_{j=1}^{c} v_{jk} > 1 \\ Z_k &= \left(\textstyle\sum_{j=1}^{c} v_{jk}^{\alpha}\right)^{1/\alpha} && \text{if} \quad \textstyle\sum_{j=1}^{c} v_{jk}^{\alpha} < 1 \\ Z_k &= 1 && \text{else.} \end{aligned} \tag{4}$$

For $\alpha = 1$, the representation properties of the method reduce to those of ME, whereas in the limit case for $\alpha = 0$, the representation properties are equivalent to those of PCM-II for low membership values, and to those of ME for higher values. In [14] this particular model has been shown to possess robustness properties, with applicability to stable clustering and outlier analysis. However, no criterion was given to assess the values of parameters α and β.

3 Fuzzy Similarity Indexes

3.1 An Approach to the Comparison of Fuzzy Clusterings

Measuring the agreement between two clusterings amounts to measuring the similarity between two partitions. There are several partition similarities available in the literature. The two main approaches include comparing matching clusters, and comparing co-association information. The first one is not reliable when the partitions are not very similar, and require some criterion for matching subsets. In [20] we focused on the second, defining a set of fuzzy pairwise indexes based on co-association matrices [6].

We are given a data set X and two fuzzy partitions A and B of X. Fuzzy partitions means that $\forall x \in X$ there is a membership $\mu(x, a_i) \in [0, 1]$ for each subset $a_i \in A$ (similarly for B). Normal memberships are obtained from the majority of fuzzy clustering methods, so we make this reasonable assumption.

Each data point is thus represented by a coordinate vector, whose dimension is the number of clusters in the partition, and whose components are the membership values. Each data pair is described by the degree of similarity between the two objects x and y under the partition A.

For membership strings, crisp similarity can be measured by Hamming distance, based on a bitwise-AND. The fuzzy generalization of this operation is defined once we

appropriately define the conjunction connective [23]. We adopt the product t-norm [16], which provides uniformity with respect to other models; in particular, it yields a formal equivalence with the scalar product operation between vectors, which in turn can be used to define popular distance measures. This allows some generality in the indexes studied, which would be lost with other, more specific conjunction choices.

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 (also termed *bonding relationship* in [3]) under partition A is:

$$s_{ij}^A = s^A(x_i, x_j) = \sum_{l=1}^{|A|} \mu(x_i, a_l)\mu(x_j, a_l) \tag{5}$$

As in [20], we serialize s^A and s^B and use the single-index equivalent σ_h^A and σ_h^B where h runs over the $|A|(|A|+1)/2$ unique pairs ij.

3.2 The Fuzzy Jaccard Index

Some indexes of partition similarity based on the co-association matrix are reviewed in [15] and some are experimentally compared in [13]. See also [20] for more background. Here we focus only on the fuzzy version of Jaccard's index [11], a classic measure of set similarity, and one of the most general. It is defined as the ratio of the size of the intersection of two sets to the size of their union: $J(A,B) = |A \cap B|/|A \cup B|$. In the crisp case, this pairwise index can be practically computed by counting $N_{11} = |A \cap B|$, number of points put in the same cluster by both partitions, and N_{10} and N_{01}, number of points assigned to the same cluster only by partition A or B respectively. so that $N_{10} + N_{01} + N_{11} = |A \cup B|$ and $J(A,B) = \frac{N_{11}}{N_{10}+N_{01}+N_{11}}$..

In the fuzzy case, we have a degree of coincidence. If De Morgan's law holds, as it is reasonable to assume, for the product t-norm we can define the associated disjunction operator as the *probabilistic sum* t-conorm, $\mu \,\mathrm{OR}\, \nu = \mu + \nu - \mu\nu$. Therefore:

$$A \cap B = \sum_h \sigma_h^A \sigma_h^B \qquad \text{and} \qquad A \cup B = \sum_h \left(\sigma_h^A + \sigma_h^B - \sigma_h^A \sigma_h^B\right) \tag{6}$$

The choice of the Jaccard index over other possible measures is suggested by the conclusions drawn in [21] after analyzing a set of 39 different measures. The Jaccard distance is a metric; the value 0 is attained only for disjoint sets; and the value 1 only for equal sets. For the fuzzy Jaccard index, the latter property holds only for crisp sets. Possibilistic memberships are unconstrained, and can have small values, so that self-co-associations can take on any value from 0 to 1; the asymmetric model chosen avoids the case of values larger than 1. In the fuzzy case, self-co-association gives an indication about the degree of fuzziness of a clustering.

4 Visual Stability Analysis Based on Comparing Fuzzy Clusterings

This section introduces our proposed clustering stability method, which, unlike other proposals [22,2] does not address comparison of points in a clustering, but builds on our previous work on fuzzy partition similarity indexes [20] to compare whole clusterings.

Hennig [9] proposed a method to exploit cluster-based comparisons to analyze the robustness of clusters, as measured by the "dissolution point" criterion. In our case this approach is not directly applicable. Our aim is to provide a tool to make a decision which is to some extent arbitrary, namely, model selection. The proposed approach is visual and interactive, allowing the user to perceive the effect of the variation of one or few parameters, in our case α and β in GPCM. The degree of possibility α has a clear theoretical interpretation, but its effect is difficult to perceive in practice. In addition, its effect combines with that of the width parameter β, as exemplified by an image segmentation application presented in [14]. Our proposal addresses these issues.

The procedure, exemplified for α, is as follows:

1. Perform repeated clustering by sweeping values of α
2. Build a similarity matrix by evaluating the pairwise fuzzy Jaccard similarity of clusterings obtained by each pair of values of α.
3. Visualize the similarity matrix and analyze it (see below)
4. Select the column or row index corresponding to maximum stability and confidence

Step 3 of the procedure is as follows. The similarity matrix compares every possible pair of clusterings. The matrix is symmetric, but in our possibilistic case the diagonal does not necessarily contain all ones: as noted earlier, fuzzy similarity indexes may have a low self-co-association value. In possibilistic clustering, this usually indicates that the cluster centers are not significant, i.e., that the local minimum found during training is quite bad. Therefore, we are searching for values for which the diagonal is brighter.

Stable clusterings correspond therefore to values with many nearby bright elements. These correspond to clusterings which are similar for a whole range of values of α. The optimal value corresponds to the largest, brightest square patch around the diagonal. To maximize the robustness margin, the central value of the patch should be selected. Due to the inherent ambiguity in clustering, there may be more than one possible optimal value. Especially when varying β, we may witness the phenomenon of "phase transition" [19], e.g., with splitting or collapsing clusters. This is also visible in the experimental results presented in the following section.

As a final remark, we note that this procedure is completely general and may be applied to any clustering method, being most useful in fuzzy cases. To increase confidence, multiple initializations may be used.

5 Experiments

5.1 Data and Experimental Setup

To experimentally verify the proposed method, we have chosen a dataset with a good degree of structure, but at the same time is not clearly clustered. This results in a visible instability, for instance when starting from different initialization points.

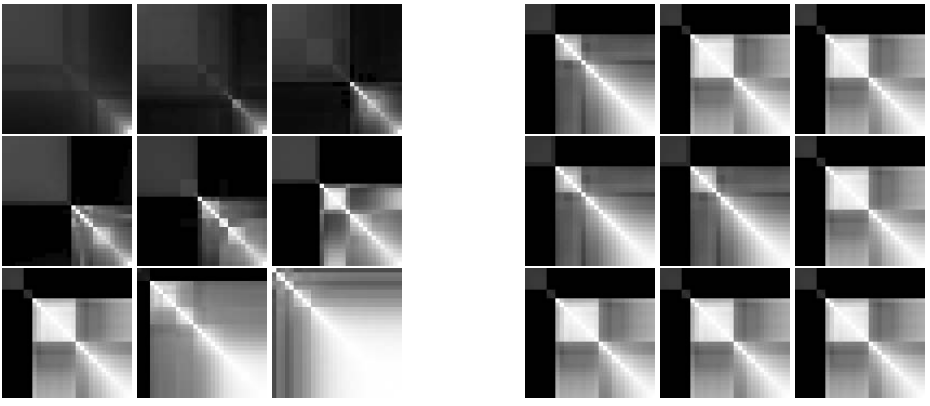


Fig. 1. Visual representation of the similarity index. Each inner square is 30 experiments with α ranging from 0.1 to 1. Across the 9 sub-squares (left to right, top to bottom): (left) single random initialization, β ranging from 0.0039 to 0.032; (right) different random initializations, $\beta = 0.011$.

The problem is provided in the base dataset package of the R language and environment (www.R-project.org) as "quakes". It consists of a subset of 1000 observations of quakes (seismic events with magnitude MB > 4.0) from a larger database of 5000 observations. These quakes occurred around Fiji, starting in 1964, and are described by three-dimensional coordinates (latitude, longitude and depth of event), plus the Richter magnitude and the number of stations that reported it, for a total of 5 variables.

Since setting a large number of cluster centroids reduces instability, we kept this number relatively small, fixing it at 7; this is a little higher than the value of 5 used by Hennig in his robustness measurement experiments.

To set the width parameter β, in the robust fuzzy clustering literature [17], various criteria are available that provide a first estimate. We only need a criterion to set a range of values, and the method itself allow us to select the optimal value. We proceed as follows: compute the average distance $\hat{d}$ between pairs of data points in the training set; fix a range by setting a minimum and a maximum, expressed as fractions of $\hat{d}$, and vary β across this range. The actual value of the β is therefore free for attribution on the user's part, but with this procedure its magnitude is related to the specific data set used and there is a search range to find the most suitable value. With the dataset used and $\beta \in [\, 0.2\hat{d}, 1.8\hat{d}\,]$, a 9-step sequence was obtained with the following values of β: $0.0039, 0.0074, 0.011, 0.014, 0.018, 0.021, 0.025, 0.028, 0.032$. In this work the number of steps is set to 9 because an odd number provides a well-defined middle value, and also for aesthetic reasons. The training was performed by 9 individual runs, each with a fixed value of β.

Each individual run consisted of one random initialization, and 30 complete optimizations, each one initialized with the output of the previous one, and α ranging from 0.1 to 1 geometrically. Another parameter is varied across the 9 individual runs of each experiment. α, starting at 0.1 and progressing up to 1, so that we obtain 30×30 similarity matrices. The stopping criterion was 100 iterations, or root-mean-square variation of centroids less than 0.001 with respect to the previous step.

5.2 Results and Comments

The visual output of the method is shown in Figure 1. These are plots of the value of the fuzzy Jaccard index. The value of the index is mapped as a lightness of each matrix entry, with black = 0 and white = 1. The experiments, as described above, are ordered from left to right and from top to bottom: in the left experiments there is a single random initialization, and β ranges from 0.0039 to 0.032; on the right, there are different random initializations with fixed $\beta = 0.011$. Lighter parts of the image correspond to pairs of similar clusterings, as already explained in Sect. 4.

The first figure shows the variation as a function of β. As we can see, the most stable, significant patch is attained in the seventh step ($\beta = 0.011$). The large, blurred patches in the last steps are due to the excessive value of the width parameter, so that all points were attributed essentially to a single, large cluster. When the width is too small, even the diagonal has low values, and only for the extreme values of α (lower right corner) data points are attributed to clusters with some confidence.

From this analysis it turns out that, in this particular instance, the best value for α is not at the possibilistic or probabilistic extremes, but settles around an intermediate value, between 0.17 and 0.24. If we now consider the experiments with fixed $\beta = 0.011$ and different random initializations (Figure 1), we can see that, despite random variations in the results, the stable patch recurs in most experiments in about the same location, confirming the selected values of $\beta = 0.011$ and $\alpha = 0.21$ as significant.

6 Remarks and Future Work

The proposed method is notably flexible, since it is agnostic with respect to cluster structure and representation. Here we have focused on centroid-based clusterings; however, other models may be treated similarly, although this might require additional cluster significance assessment. We also restricted the analysis to the Jaccard index, but a comparison between the possible choices from [20] could be performed.

Incidentally, the obtained value for α indicates that our method, using a localized representation for membership functions, provides better results than either probabilistic or possibilistic models.

For completely automated operation, it is possible to replace the visual analysis with a metaclustering step, based on the cluster similarity matrix. This would certainly sacrifice part of the flexibility and potentiality of the method, but would allow to use it in an embedded fashion, as part of a more complex system.

References

1. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell (1981)
2. Bezdek, J., Hathaway, R., Huband, J.: Visual assessment of clustering tendency for rectangular dissimilarity matrices. IEEE Transactions on Fuzzy Systems 15(5), 890–903 (2007)
3. Brouwer, R.K.: Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions. J. Intell. Inf. Syst. 32(3), 213–235 (2009)

4. Davé, R.N., Krishnapuram, R.: Robust clustering methods: a unified view. IEEE Transactions on Fuzzy Systems 5(2), 270–293 (1997)
5. Filippone, M., Masulli, F., Rovetta, S.: Stability and performances in biclustering algorithms. In: Masulli, F., Tagliaferri, R., Verkhivker, G. (eds.) CIBB 2008. LNCS, vol. 5488, pp. 91–101. Springer, Heidelberg (2009)
6. Fred, A.L.N., Jain, A.K.: Data clustering using evidence accumulation. In: International Conference on Pattern Recognition, 4 (2002), http://dx.doi.org/10.1109/ICPR.2002.1047450
7. Frigui, H., Krishnapuram, R.: A robust competitive clustering algorithm with applications in computer vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(5), 450–465 (1999)
8. Frigui, H., Krishnapuram, R.: A robust clustering algorithm based on m-estimator. In: Proceedings of the 1st International Conference on Neural, Parallel and Scientific Computations, Atlanta, USA, vol. 1, pp. 163–166 (May 1995)
9. Hennig, C.: Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis 52(1), 258–271 (2007)
10. Huber, P.J.: Robust Statistics. John Wiley and Sons, New York (1981)
11. 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)
12. Krishnapuram, R., Keller, J.M.: The possibilistic C-Means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems 4(3), 385–393 (1996)
13. 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)
14. Masulli, F., Rovetta, S.: Soft transition from probabilistic to possibilistic fuzzy clustering. IEEE Transactions on Fuzzy Systems 14(4), 516–527 (2006)
15. Meilă, M.: Comparing clusterings–an information based distance. Journal of Multivariate Analysis 98(5), 873–895 (2007)
16. Menger, K.: Statistical metrics. Proceedings of the National Academy of Sciences of the United States of America 28(12), 535–537 (1942)
17. Nasraoui, O., Krishnapuram, R.: A robust estimator based on density and scale optimizations and its application to clustering. In: FUZZ-IEEE 1996: Proceedings of the International Conference on Fuzzy Systems, pp. 1031–1035. IEEE, New Orleans (1996)
18. Rand, W.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66, 846–850 (1971)
19. Rose, K., Gurewitz, E., Fox, G.: A deterministic annealing approach to clustering. Pattern Recognition Letters 11, 589–594 (1990)
20. Rovetta, S., Masulli, F.: An experimental validation of some indexes of fuzzy clustering similarity. In: Gesù, V.D., Pal, S.K., Petrosino, A. (eds.) WILF 2009. LNCS, vol. 5571, pp. 132–139. Springer, Heidelberg (2009)
21. Shi, G.: Multivariate data analysis in palaeoecology and palaeobiogeographya review. Palaeogeography, Palaeoclimatology, Palaeoecology 105(3-4), 199–234 (1993)
22. Strehl, A., Ghosh, J.: Cluster ensembles — a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. 3, 583–617 (2003)
23. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)

Granular Data Regression with Neural Networks

Mario G.C.A. Cimino[1], Beatrice Lazzerini[1],
Francesco Marcelloni[1], and Witold Pedrycz[2]

[1] Dipartimento di Ingegneria dellInformazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, Largo Lucio Lazzarino 1 - 56122 Pisa, Italy
{m.cimino,b.lazzerini,f.marcelloni}@iet.unipi.it
[2] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada, T6G 2G7
pedrycz@ece.ualberta.ca

Abstract. Granular data offer an interesting vehicle of representing the available information in problems where uncertainty, inaccuracy, variability or, in general, subjectivity have to be taken into account. In this paper, we deal with a particular type of information granules, namely interval-valued data. We propose a multilayer perceptron (MLP) to model interval-valued input-output mappings. The proposed MLP comes with interval-valued weights and biases, and is trained using a genetic algorithm designed to fit data with different levels of granularity. The modeling capabilities of the proposed MLP are illustrated by means of its application to both synthetic and real world datasets.

Keywords: Granular computing, Information granules, Neurocomputing, Interval analysis, Symbolic data analysis, Function approximation.

1 Introduction

Human capabilities are based on the ability of processing non-numeric information clumps (*granules*) rather than individual numeric values [1]. Information granules can be regarded as collections of objects that exhibit some similarity in terms of their properties of functional appearance [2]. There are a number of formal models of information granules including sets, rough sets, fuzzy sets, and shadowed sets to name a few options. In [3] the authors claim that the implementation of granules in terms of interval-valued data is the easiest to comprehend and express by a domain expert, and the simplest to process when there is a great variability of granule sizes.

The objective of this study is to propose a neural architecture to process information granules consisting of interval-valued data. Interval-valued data arise in several practical situations, such as recording monthly interval temperatures at meteorological stations, daily interval stock prices, inaccuracy of the measurement instruments, range of variation of a variable through time. In the proposed model, each operation performed in the network is based on interval arithmetic and this allows creating mappings at different levels of granularity. Since the

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 172–179, 2011.

level of granularity is problem-oriented and user-dependent, it is a parameter of our neural architecture.

The first conceptualization of neural networks for processing granular data was introduced by Pedrycz and Vukovich in [3]. Here, several design approaches are discussed, together with a number of architectures of granular neural networks and associated training methods. Also, the authors tackle a number of fundamental issues of these networks, such as specificity of information granules, learning complexity, and generalization capabilities. Neural architectures based on interval arithmetic have been proposed in [2,4,5,8,9,10]. In particular, the model developed in [4] uses a standard multilayer perceptron (MLP) with numeric weights and biases, and a neuron transfer function able to operate with interval-valued inputs and outputs. Here, the training process uses an error function based on a weighted Euclidean distance between intervals, and a Quasi Newton method for the minimization of the error function. More robust minimization methods such as genetic algorithms and evolutionary strategies have been also proposed [5].

In its most general architecture proposed in the literature, an MLP that processes interval-valued data is characterized by weights and biases expressed in terms of intervals, and maps an interval-valued input vector to an interval-valued output. However, very often, in the design of the training methods some simplifying assumptions are made, e.g. input, weights and biases may be real numbers, or the error function between intervals is not compliant with the rules of the interval arithmetic.

This paper proposes a new genetic-based learning method for a general interval-valued neural architecture. We also show the effectiveness of this method by using three interval-valued datasets.

2 Interval Arithmetic: Some Definitions

We employ a basic implementation of granules in terms of conventional interval-valued data. An interval-valued variable $\ddot{X}$ is defined as:

$$\ddot{X} = [\underline{x}, \overline{x}] \in \mathbb{IR},\ \underline{x}, \overline{x} \in \mathbb{R}\ . \tag{1}$$

where $\mathbb{IR}$ is the set of all closed intervals in the real line, and $\underline{x}$ and $\overline{x}$ are the boundaries of the intervals. An F-dimensional granule is then represented by a vector of interval-valued variables as follows:

$$\ddot{\mathbf{X}} = [\ddot{X}_1, ..., \ddot{X}_F] \in \mathbb{IR}^F,\ \ddot{X}_i \in \mathbb{IR}\ . \tag{2}$$

Sometimes an interval variable is expressed in terms of its midpoint $\dot{x}$ and half-width $\hat{x}$, as follows [5]:

$$\widehat{\ddot{X}} = \langle \dot{x}, \hat{x} \rangle \in \mathbb{IR},\ \dot{x}, \hat{x} \in \mathbb{R},\ \dot{x} = (\overline{x} + \underline{x})/2,\ \hat{x} = (\overline{x} - \underline{x})/2\ . \tag{3}$$

Table 1 summarizes some basic operations of interval arithmetic that have been used in this study. The interested reader can find a detailed discussion in [6,7].

3 The Adopted Interval-Valued Neural Architecture

Let $f : \mathbb{IR}^F \rightarrow \mathbb{IR}$ be an F-dimensional interval-valued regression model:

$$f(\ddot{\mathbf{X}}) = f(\ddot{X}_1, ..., \ddot{X}_F) = \ddot{Y} \ . \tag{4}$$

Table 1. Some basic interval arithmetic operations used in interval-valued MLP

Operation	Implementation
Addition	$[\underline{x}, \overline{x}] + [\underline{y}, \overline{y}] = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$
Multiplication	$[\underline{x}, \overline{x}] \times [\underline{y}, \overline{y}] = [min\{\underline{x}\underline{y}; \underline{x}\overline{y}; \overline{x}\underline{y}; \overline{x}\overline{y}, max\{\underline{x}\underline{y}; \underline{x}\overline{y}; \overline{x}\underline{y}; \overline{x}\overline{y}]$
Function evaluation	$F([\underline{x}, \overline{x}]) = [F(\underline{x}), F(\overline{x})]$, F monotonically increasing
Real distance	$dist([\underline{x}, \overline{x}], [\underline{y}, \overline{y}]) = max\{\lvert\underline{x} - \underline{y}\rvert, \lvert\overline{x} - \overline{y}\rvert\}$
Absolute value	$\lvert[\underline{x}, \overline{x}]\rvert = dist([\underline{x}, \overline{x}], [0, 0]]) = max\{\lvert\underline{x}\rvert, \lvert\overline{x}\rvert\}$

Fig. 1 shows the MLP we adopt to deal with the regression problem modeled by (4). This architecture has been already proposed by some authors (for instance, in [5]). The novelty of our approach concerns the training process, which allows an effective and efficient sensitivity analysis (i.e., to quantify the effect of input variability on the outputs). The hidden layer comprises N nonlinear hidden units and the output layer consists of one linear output unit. The activation of each hidden unit j is obtained as sum between the weighted linear combination, with weights $\ddot{\Omega}_{i,j}$, $i = 1, ..., F$, $j = 1, ..., N$, of the F interval-valued inputs $\ddot{X}$ and the bias $\ddot{\Omega}_{0,j}$. Since both weights and biases are intervals, this linear combination results in a new interval. The output of each hidden unit is then obtained by transforming its activation interval using a hyperbolic tangent (sigmoid) function. Since the function is monotonic, this transformation yields a new interval [5]. Finally, the output of the network, $\ddot{Y}$, is obtained as the sum between the weighted linear combination, with weights $\ddot{\Omega}_j$, $j = 1, ..., N$, of the outputs of the hidden layer, and the bias $\ddot{\Omega}_0$. The overall processing method is based on the fundamental arithmetic operations on $\mathbb{IR}$ shown in Table 1. The resulting model can be used in two ways [4]: (i) as a granular function approximation model, whose granular weights can be adjusted through supervised learning by minimizing an error function; (ii) as an instrument to evaluate the prediction

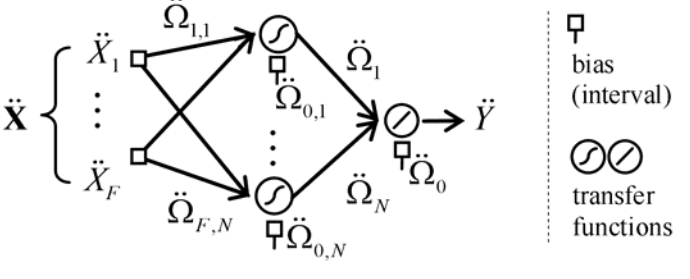


Fig. 1. The proposed architecture of MLP for interval-valued data

of a pre-adjusted MLP model subject to variable uncertainty associated with its input variables. Such input uncertainty can be characterized using interval inputs of different lengths.

4 The Training of the Interval-Valued MLP

As pointed out in [5], the width of the predicted output regions for an interval MLP is affected by the width of the weight intervals. Wide widths cause the propagation of large ranges of intermediate values through the network, thus generating wide output intervals. This is known as "bound explosion" effect. To control this effect, we adopt the following procedure. Let $\{\ddot{\mathbf{X}}, \ddot{Y}\} \in \mathbb{IR}^{F+1}$ be a set of T input-output interval-valued samples, represented as midpoint and half-width. First, we use the midpoints of T to train a conventional MLP which has the same structure as the interval-valued MLP to be developed. We adopt the Levenberg-Marquadt method. In this way, we form a reference model, which solves the regression problem for reference points of the interval-valued data. When we tackle the regression problem for the interval-valued data, we expect that the interval weights and biases contain the numerical weights and biases ($\dot{\omega}_{i,j}^{(init)}$ and $\dot{\omega}_{j}^{(init)}$) of the reference model, respectively. Further, we expect that the widths of these intervals are constrained by the level of granularity of the mapping that is determined by the problem and by the desired resolution with which the user is interested in observing the data.

The first requirement is satisfied by enforcing the following relationship:

$$\dot{\omega}_{i,j}^{(init)} \in \ddot{\Omega}_{i,j} \text{ and } \dot{\omega}_{j}^{(init)} \in \ddot{\Omega}_{j} \ \forall i,j \ . \tag{5}$$

As regards the second requirement, we enforce that the half-widths of weights and biases are bounded by an interval-valued percentage of the initial values:

$$\hat{\omega}_{i,j} \in |\dot{\omega}_{i,j}^{(init)}| \cdot \ddot{G} \text{ and } \hat{\omega}_{j} \in |\dot{\omega}_{j}^{(init)}| \cdot \ddot{G} \ \forall i,j \ . \tag{6}$$

where $\ddot{G} = [\underline{g}, \overline{g}] \in \mathbb{IR}^{+}$, with $\underline{g}, \overline{g} \in \mathbb{R}^{+}$, is a granularity interval expressed in percentages which allows to adapt the granularity of the mapping to the granularity level of the information. The choice of $\ddot{G}$ depends on the specific performance index used to assess the quality of the model. For instance, in our case we used the network error.

To learn the interval-valued weights and biases, standard error back propagation is likely to give poor results [5]. Indeed, the network prediction error surface is expected to be very nonlinear with several local minima. A global search method is much more desirable. Genetic algorithms (GAs) and evolutionary strategies are, in general, effective examples of such methods. Thus, we decided to adopt a GA. Fig. 2 shows the chromosome coding used in the GA. The initial population is randomly generated by satisfying the constraints in formulas (5) and (6). Chromosomes are selected for mating by a fitness proportional strategy. We apply the classical two-point crossover operator, with a user-defined crossover probability P_c. The mutation operator is controlled by a mutation probability defined as γ_m/L, where γ_m is a user-defined mutation coefficient and $L = 2N(F+2)+2$ is the chromosome length. We randomly choose

Fig. 2. The chromosome coding

a user specified percentage P_R of the genes that undergo mutation. Then, we replace the current values of each selected gene by randomly extracting two values in the intervals $[|\dot{\omega}_{i,j}^{(init)}| \cdot \underline{g},\ |\dot{\omega}_{i,j}^{(init)}| \cdot \overline{g}]$ and $[\dot{\omega}_{i,j}^{(init)} - \hat{\omega}_{i,j},\ \dot{\omega}_{i,j}^{(init)} + \hat{\omega}_{i,j}]$, respectively. The first interval is directly related to the definition of $\ddot{G}$. Once provided $\hat{\omega}_{i,j}$, from formulas given by (6) we derive that the maximum distance from $\dot{\omega}_{i,j}^{(init)}$ can be $\dot{\omega}_{i,j}^{(init)} \pm \hat{\omega}_{i,j}$.

As regards the fitness function, unlike the network error functions proposed in the literature that implicitly assume an isomorphism between $\mathbb{IR}$ and $\mathbb{R}^2$, we adopt the following error function directly derived from the interval arithmetic operations shown in Table 1:

$$E = \frac{1}{T}\sum_{i=1}^{T} E_i\ ,\quad E_i = dist(\ddot{Y}_i, \ddot{Y}'_i) \in \mathbb{R}^+\ . \tag{7}$$

where $\ddot{Y}_i$ and $\ddot{Y}'_i$ are the desired and network outputs. The algorithm stops if a maximum number N_G of generations is reached or if the best fitness of the population is lower than a prefixed fitness threshold τ.

5 Experimental Results

A variety of works have been developed in the field of interval-valued data. Unfortunately there is still a lack of significant benchmark datasets for interval-valued data regression. In this section, we discuss the application of our interval-valued neural architecture to one real world and two synthetic datasets. In all experiments, the data are normalized between 0 and 1 (by subtracting the minimum value and dividing the data by the difference between the maximum and the minimum values). The population of the GA consists of 20 chromosomes. The parameters P_c, γ_m, P_R, and τ have been set to 0.4, 0.7, 10%, and 0.001, respectively. We used a value of N_G equals to 500 except for the experiment in section 5.3 where we adopted a value equals to 1000.

5.1 The Salary Dataset

The Salary dataset [10] shown in Fig. 3.a consists of 30 interval-valued samples which represent the range of salaries by years of experience for American males with degree in 1989. The original data samples are not granular, and subject to significant sampling error. First, fuzzy information granules have been generated via FCM clustering. Hence, an alpha-cut of 0.05 has been applied to the resulting fuzzy partition. Finally, interval-valued data have been derived considering, for each alpha-cut, the smallest containing rectangle. We adopted a 10-fold cross-validation: for each trial, the training and the test sets consist of the randomly

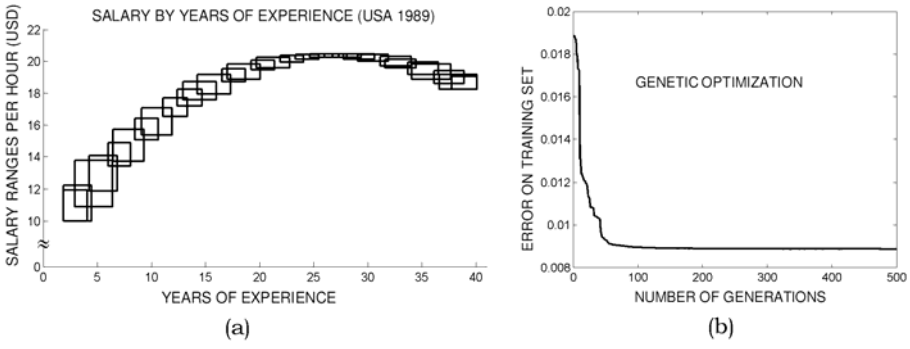


Fig. 3. (a) The Salary dataset. (b) Training error versus generations in a trial.

extracted 90% and 10% of the original data, respectively. The granularity interval used to observe the mapping is $\ddot{G} = [0, 4]\%$. The network has been equipped with 15 hidden neurons. The mean values $\pm$ the standard deviations of the error on training and test set are, respectively, 0.007 ± 0.0051 and 0.018 ± 0.028. We can observe a good balance between the values of errors for the training and test sets. This confirms sound generalization capabilities of the network.

Fig. 3b shows the error of the best chromosome of each generation versus the number of generations in a sample trial. We observe that the error gets stable around 100 generations.

5.2 The Peak Dataset

The Peak dataset shown in Fig. 4a consists of 189 synthetic interval-valued samples. Again, we adopted a 10-fold cross-validation. The granularity interval used to observe the mapping is $\ddot{G} = [0.4, 4]\%$. The network has been equipped with 30 hidden neurons. The mean values $\pm$ the standard deviations of the

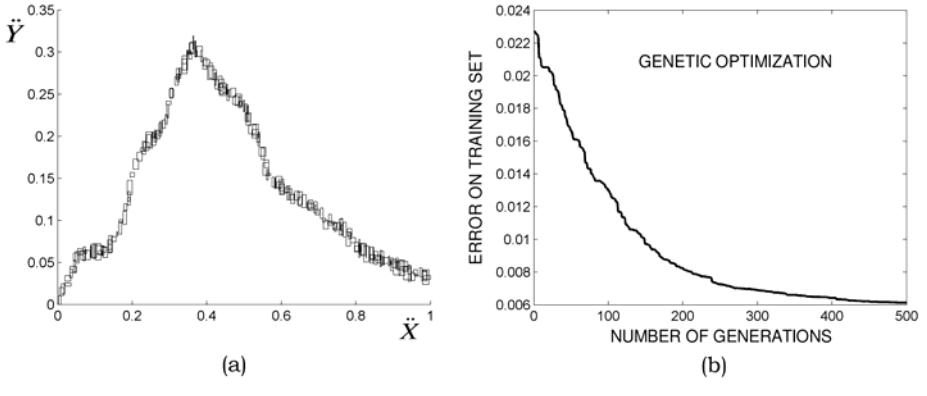


Fig. 4. (a) The Peak dataset. (b) Training error versus generations in a trial.

error on training and test set are, respectively, 0.0064 ± 0.0014 and 0.0085 ± 0.0043. We can observe that the error in the test set is very close to the error in the training set, thus pointing out the good generalization capabilities of the network. Fig. 4.b shows the error of the best chromosome of each generation versus the number of generations in a sample trial. We observe that the error gets stable around 500 generations.

5.3 The Wave Dataset

The Wave dataset shown in Fig. 5.a consists of 400 synthetic interval-valued samples in the three-dimensional space. The network has been equipped with 15 hidden neurons. We use this dataset for analyzing the differences between the mappings with different granularity intervals. Fig. 5.b shows the error of the best chromosome of each generation versus the number of generations when

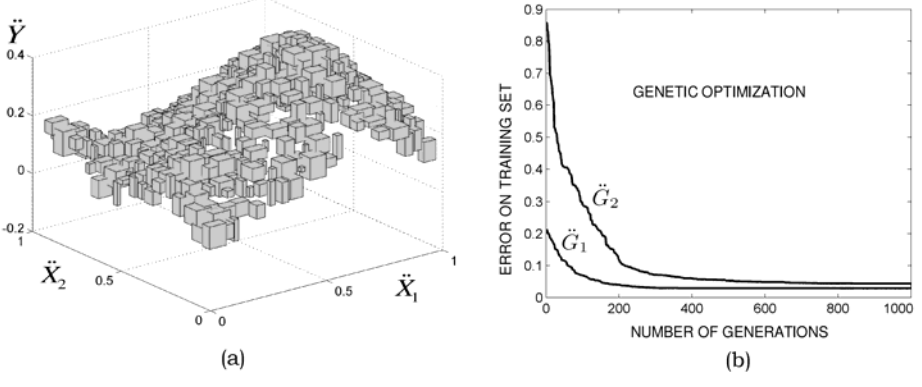


Fig. 5. (a) The Wave dataset. (b) Training error versus generations with $\ddot{G}_1 = [0.2, 10]\%$ and $\ddot{G}_2 = [10, 50]\%$.

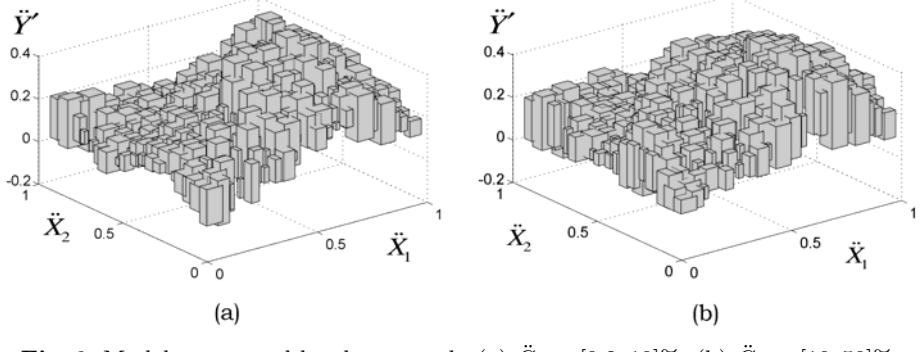


Fig. 6. Models generated by the network. (a) $\ddot{G}_1 = [0.2, 10]\%$. (b) $\ddot{G}_2 = [10, 50]\%$.

the network is trained with the granularity intervals $\ddot{G}_1 = [0.2, 10]\%$ and $\ddot{G}_2 = [10, 50]\%$.

Fig. 6.a and Fig. 6.b show the models generated by the network in the two cases, respectively. We can observe how the model generated by using $\ddot{G}_2$ is coarser than the model generated using $\ddot{G}_1$. Further, in the former, the error gets stable around 500 generations against the 300 of the latter.

6 Conclusions

We have proposed a new genetic-based learning method for a general interval-valued neural architecture. The originality of the approach concerns the training process which allows a valuable sensitivity analysis. We have quantified the effectiveness of the approach in terms of generalization capabilities and sensitivity by using three interval-valued datasets.

References

1. Pedrycz, W.: Knowledge-Based Clustering: From Data to Information Granules. John Wiley and Sons, Hoboken (2005)
2. Nandedkar, A.V., Biswas, P.K.: A Reflex Fuzzy Min Max Neural Network for Granular Data Classification. In: 18th Int. Conference on Pattern Recognition, vol. 2, pp. 650–653 (2006)
3. Pedrycz, W., Vukovich, G.: Granular Neural Networks. Neurocomp 36, 205–224 (2001)
4. Muñoz San Roque, A., Maté, C., Arroyo, J., Sarabia, Á.: iMLP: Applying Multi-Layer Perceptrons to Interval-Valued Data. Neural Processsing Letters 25, 157–169 (2007)
5. Chetwynd, D., Worden, K., Manson, G.: An application of interval-valued neural networks to a regression problem. Proceedings of the Royal Society A 462, 3097–3114 (2006)
6. Kearfott, R.B.: Interval computations-Introduction, uses, and resources. Euromath. Bull. 2, 95–112 (1996)
7. Alefeld, G., Claudio, D.: The basic properties of interval arithmetic, its software realizations and some applications. Computers & Structures 67, 3–8 (1998)
8. Ishibuchi, H., Tanaka, H., Okada, H.: Architecture of neural networks with interval weights and its application to fuzzy regression analysis. Fuzzy Sets and Systems 57, 27–39 (1993)
9. Baker, M.R.: Universal Approximation Theorem for Interval Neural Network. Reliable Computing 4, 235–239 (1998)
10. Patio-Escarcina, R.E., Callejas Bedregal, B.R., Lyra, A.: Interval Computing in Neural Networks: One Layer Interval Neural Networks. Intelligent Information Technology 3356, 68–75 (2005)
11. Murphy, K.M., Welch, F.: The Structure of Wages. Quarterly Journal of Economics 1, 285–326 (1992)

A Fuzzy Declarative Approach for Classifying Unlabeled Short Texts Using Thesauri

Francisco P. Romero[1], Pascual Julian-Iranzo[1], Andres Soto[2], Mateus Ferreira-Satler[1], and Juan Gallardo-Casero[1]

[1] Dept. of Information Systems and Technologies, University of Castilla La Mancha. Ciudad Real, Spain
{FranciscoP.Romero,Pascual.Julian,Mateus.Ferreira,Juan.Gallardo}@uclm.es
[2] Univ. Autonoma del Carmen, Ciudad del Carmen Campeche, Mexico
soto_andres@yahoo.com

Abstract. The classic approach to text categorisation is based on a learning process that requires a large number of labelled training texts to achieve an accurate performance. The most notable problem is that labelled texts are difficult to generate because categorising shorts texts as snippets or messages must be done by human developers, although unlabelled short texts could be easily collected. In this paper, we present an approach to categorising unlabelled short texts which only require, as user input, the category names defined by means of an ontology of terms modelled by a set of *proximity equations*. The proposed classification process is based on the ability of a fuzzy extension of the standard Prolog language named Bousi∼Prolog for flexible matching and knowledge representation. This declarative approach provides a text classifier which is fast and easy to build, as well as a classification process that is easy for the user to understand. The results of the experiment showed that the proposed method achieved a reasonably good performance.

Keywords: Text Categorization, Thesauri, Fuzzy Declarative Languages.

1 Introduction

Text categorisation (also known as text classification) is the task of automatically sorting a set of documents into categories from a predefined set [9]. In automatic text categorisation, the decision criterion of the text classifier is usually learned from a set of training documents, labelled for each class. One of the problems with this method is related with the exponential growth of the number of training set documents required as the precision degree of the method increases. This way, the time and effort required for collecting and preparing an adequate training set could be a restriction, and probably, prohibitive. This is an important issue when classifying short texts because although there are a lot of available short texts, the majority of them are unlabelled. Moreover, short text categorisation cannot be carried out by relying only on statistical methods; it is also necessary to exploit the semantic relationships between words.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 180–187, 2011.

This work tackles the problem of classifying short texts without using training data; that is, a collection of short texts which has been previously classified. Instead, a method that only uses prior domain knowledge is developed. Hence, the one single input of this method is the list of category names. These names are transformed into a set of concept descriptions extracted from thesauri and ontologies like WordNet [2] or ConceptNet [7] by measuring the semantic closeness between the concepts involved with each one of the categories. These descriptions are the basis of the classification method. Furthermore, in our method, the classification process is based on the ability of an extension of the standard Prolog language, named Bousi∼Prolog fto carry out flexible matching and knowledge representation. This extension offers a fuzzy unification mechanism based on closeness relations which allows a flexible search for concepts in texts. Hence, our method implements a clean separation between knowledge knowledge (refined by an ontology), logic (expressed by rules) and control (let automatic to the abstract machine supporting the underlying programming language).

The combination of these components provides a declarative approach to text classification, in which a text classifier is easier to build than it usually is. At the same time, the classification process becomes more understandable for the user, since it mainly relies on an ontology description. While most of the work in classification nowadays is founded on statistical methods, this paper takes a Semantic Web and Soft-Computing approach using thesauri as a source of domain knowledge.

The paper is organized as follows: Section 2 includes a concise summary about proximity relations between concepts and the Bousi∼Prolog language which offers the required mechanisms to implement a text categorization method using a declarative approach. Section 3 describes our method in detail including an explanatory example. Section 4 explains the experiment and the results obtained in order to verify the goodness of the solution. Finally, our conclusions and future work lines are given in Section 5.

2 Background

2.1 Proximity Relations between Concepts

Binary fuzzy relations were introduced by Zadeh in [10]. Formally, a *binary fuzzy relation* on a set U is a fuzzy subset on $U \times U$ (that is, a mapping $U \times U \longrightarrow [0, 1]$). Given a and b two elements in U, an *entry* of a fuzzy relation will be denoted as $\mathcal{R}(a, b) = \alpha$, being α its *relationship* degree.

A binary fuzzy relation $\mathcal{R}$ is said to be a *proximity relation* if it fulfills the *reflexive* property (i.e. $\mathcal{R}(x, x) = 1$ for any $x \in U$) and the *symmetric* property (i.e. $\mathcal{R}(x, y) = \mathcal{R}(y, x)$ for any $x, y \in U$). A proximity relation which in addition fulfills the *transitive* property (i.e., $\mathcal{R}(x, z) \geq \mathcal{R}(x, y) \triangle \mathcal{R}(y, z)$, for any $x, y, z \in U$) is said to be a *similarity relation*. The operator '$\triangle$' is an arbitrary t-norm. The notion of transitivity above is $\triangle$-transitive. If the operator $\triangle = \wedge$ (that is, it is the minimum of two elements), we speak of *mim*-transitive or $\wedge$-transitive. This is the standard notion of transitivity used in this paper.

According to the approach introduced in this paper, concepts with a positive closeness relation to a category name should be identified, including the degree of the relationship (a real value ranging from 0 to 1) and formalised into a fuzzy relation. For this purpose, different conceptual relations included in Concept Net and WordNet have been used to estimate the relationship degree between two concepts, i.e. to estimate how semantically similar or close they are. In the following paragraphs, we summarise the main sources for determining the conceptual relations used in this paper.

ConceptNet [7], is a freely available common sense knowledge base and natural language processing toolkit. ConceptNet provides utilities for computing the degree of conceptual neighborhood between some elements (concepts). Then, for each element b in the list of contextual neigbour concepts of a source concept a and their degree of relationship α, an entry $\mathcal{R}(a, b) = \alpha$ of a fuzzy relation is built. This degree is found by performing spreading activation from that source concept, radiating outwardly to include other concepts. The relatedness of any particular concept with some other concept is a function of the number of links and the number of paths between them, and the directionality of the edges.

WordNet [2] has been another source of knowledge used in this work. WordNet::Similarity [8] has been used here to obtain several measures of semantic relatedness between pairs of concepts (or word senses) which are based on the WordNet lexical database. For this purpose, definitions of each one of the categories in WordNet and Wikipedia has been used to estimate a proximity degree by means of the measures included in WordNet::Similarity.

Before ending this subsection, it is important to mention that all of the above detailed methods build just a partial view of a fuzzy relation (certainly, only the entries connecting a category with a set of related terms are produced). Therefore, some post-processing of that partial relation may be needed, depending on what features of the semantic relationship are required. It is necessary to build the reflexive, symmetrical closure of the partial relation to work with a proximity relation. On the other hand, if the desired relation is a similarity one, the reflexive, symmetrical and transitive closure of the partial relation should be built. Fortunately, Bousi∼Prolog gives automatic support for the generation of these kinds of closures.

2.2 Bousi∼Prolog and Flexible Search

Bousi∼Prolog [4] is a fuzzy logic programming language whose main objective is to make the query answering process flexible, and to manage the vagueness which occurs in the real world by using declarative techniques. Its design has been conceived to make a clean separation between Logic, Vague Knowledge and Control. In a Bousi∼Prolog program Logic is specified by a set of Prolog facts and rules, Vague Knowledge is mainly specified by a set of, what we call, *proximity equations*, defining a fuzzy binary relation (expressing how close two concepts are), and Control is let automatic to the system, through a "weak" SLD resolution operational mechanism. *Weak SLD resolution* is an enhancement of the SLD resolution principle where the classical syntactic unification procedure is

replaced by a fuzzy unification algorithm based on proximity relations defined on a syntactic domain. Informally, this *weak unification* algorithm states that two terms $f(t_1, ..., t_n)$ and $g(s_1, ..., s_n)$ weakly unify if the root symbols f and g are close, with a certain degree, and each of their arguments t_i and s_i weakly unify. Therefore, the weak unification algorithm does not produce a failure if there is a clash of two syntactical distinct symbols, whenever they are approximate, but a success with a certain approximation degree. Hence, Bousi∼Prolog computes substitutions as well as approximation degrees.

Bousi∼Prolog is implemented as an extension of the standard Prolog language. The Bousi∼Prolog syntax is mainly the Prolog syntax but enriched with a built-in symbol "∼" used for describing proximity relations [1] by means of what we call a "proximity equation". Although, a proximity equation represents an entry of an arbitrary fuzzy binary relation, its intuitive reading is that two constants, n-ary function symbols or n-ary predicate symbols are approximate or similar with a certain degree. That is, a proximity equation $a \sim b = \alpha$ can be understood in both directions: a is approximate/similar to b and b is approximate/similar to a with degree α. Therefore, a Bousi∼Prolog program is a sequence of Prolog facts and rules followed by a sequence of proximity equations.

On the other hand, Bousi∼Prolog implements a number of remarkable features, such as the inclusion of fuzzy sets in the core of the language [6] or the automatic support for generating reflexive, symmetric and transitive closures of a fuzzy relation [5]. The last one is intensively used in our proposal of categorization through the internal operational mechanism of Bousi∼Prolog.

3 Text Classification Proposal

Bousi∼Prolog allows us to implement a declarative approach to text categorisation using flexible matching and knowledge representation by means of an ontology of terms modelled by a set of proximity equations. The following sections show how proximity equations can be used as a fuzzy model for text categorisation when the knowledge base is selected from an ontology; that is, a structured collection of terms that formally defines the relations among them [3] inside a semantic context rather than a purely syntactic one.

The objective of any process of text categorisation is to classify each one of the documents by assigning them to one or more predetermined categories. In our approach, the availability of a set of labelled texts or a training process is not necessary; only background knowledge is used to classify the documents.

In order to describe our classification method effectively and to detail the different phases enumerated, let us consider a running example that we shall develop throughout this section. Firstly, the problem of classifying a short text (160 characters) with regard to a set of categories will be considered. Then, the results produced by the proposed method will be described.

[1] Actually, fuzzy binary relations which are automatically converted into proximity or similarity relations.

Example 1. Consider a set with tree categories `wheat`, `grain` and `ship` jointly with the following short text extracted from the Reuters Test collection[2].

```
Israel will tender overnight for 33,000 long tons of U.S. sorghum and/or 22,000 long
tons of U.S. corn for April grain shipment, private export sources said.
```

First we want to link one o several categories with the document, since this is the essence of a classification process. On the other hand, note that the "Reuters expert" classified the text with the category `grain`. A category like `ship` is also possible but the expert didn't choose them. Therefore, we would like to identify what is the knowledge that the expert used to classify the document in these categories.

The proposed method consists of the following steps or phases:

1. **Knowledge Base Building**: The first step is to define each of these categories as accurately as possible. The starting point of this definition is the concept related to the category name. Therefore, concepts like "wheat" or "ship" (which are defined as categories in the Reuters Collection), will be the source of the background knowledge. The definition of a concept is built from the set of concepts that are semantically close to it. These semantic relationships are built according to the techniques described at the end of Section 2.1 to construct a (partial) fuzzy relation. Once the fuzzy relation is established, it is represented through a set of proximity equations and then these equations are loaded into the Bousi∼Prolog system. Bousi∼Prolog automatically generates the reflexive and symmetric closures of the original (partial) relation, starting from the set of proximity equations. Optionally, a transitive closure could also be generated, producing a similarity relation. For our running example (Example 1), the knowledge is extracted from a thesaurus. More precisely, the sources are Wordnet related terms and one of the WordNet::Similarity measures, in this case, the Vector measure (the relatedness of two terms is computed as the cosine of their representative gloss vectors), leading to the following set of proximity equations:

```
wheat~bulghur=0.25    wheat~cereal=0.89  wheat~durum=0.96  wheat~grain=0.27
grain~cereal=1        grain~seed=0.32    grain~corn=0.35   grain~sorghum=0.33
ship~watercraft=0.19  ship~shipment=0.5  ship~travel=0.18  ship~transport=0.8
```

2. **Document Processing**: The input documents are processed using classical techniques of natural language processing: removing stop words, performing a stemming process based on WordNet and grouping meaningful couples of words. For our running example, the text obtained after this process is the following:

```
israel tender overnight long tons sorghum long tons corn april grain shipment
private export sources
```

3. **Flexible Search**: Bousi∼Prolog is used to search for the terms which are close to a category in the content of each text to classify, obtaining their degrees of occurrence. The Bousi∼Prolog predicate `inspect/3` reads a text, word by word, looking for words that are close (according to the proximity equations) to one of the pre-established categories. As a result, the predicate

[2] http://www.daviddlewis.com/resources/testcollections/reuters21578/

`inspect/3` returns a sequence of triples $t(X_i, N_i, D_i)$, where X_i is a term close or similar to the category, with degree D_i, occurring N_i times in a text i. In order to search for words close or similar to a given one, this predicate relies on the fuzzy unification mechanism implemented in the core of the Bousi∼Prolog language. More specifically, it uses a *weak unification operator*, also denoted by ∼, which is the fuzzy counterpart of the syntactic unification operator present in the standard Prolog language. Coming back to our running example, after inspecting the text of the example for the category `ship`, by using the predicate `inspect/3`, the result would show the number of times that the word "shipment" occurs within the text (i.e. just only once) and the degree of relation between this word and the category `ship` (a degree of 0.5).

4. **Computing Document Compatibility Degrees**: The compatibility degrees of the text with regard to a category are computed using a certain compatibility measure. A *compatibility measure* is an operation which uses the occurrence degrees D_i of the terms close to a category c to calculate a text compatibility degree CD_i^c, that is, an index of how compatible is the document with regard to the analyzed category. The predicate `compDegree/4` process the information provided by the predicate `inspect/3`. For each triple $t(T_{i_1}, N_{i_1}, D_{i_1}), ..., t(T_{i_n}, N_{i_n}, D_{i_n})$, the compatibility degree of one category c with the text i, (CD_i^c) is computed. The use of several formulae to obtain these compatibility degrees is possible. For our running example, using the compatibility measure operator `sum` (*weighted sum of the occurrence degrees*), defined in Equation 1, we obtain a 0.5 compatibility degree with the category `ship` for the corresponding document. At the end of this process, each category reaches a compatibility degree within each document.

$$CD_i^c = \sum_{k=1}^{n} (N_{i_k} * D_{i_k}) \tag{1}$$

5. **Classification Process**: The text classification procedure is very simple: the categories with the highest compatibility degree are selected as the "winners". A predicate `classify/2` takes the document compatibility list, obtained in the previous phase. This list may contain one or several categories or it may be empty. In the last case, the document is not classiffied. For our running example, the categories `wheat`, `grain` and `ship` have a positive compatibility degree, 1.57, 1.68 and 0.5 respectively. The category `grain` is the winner. The word "grain" occurs in the text and the semantic proximity between the words "sorghum" and "corn" and `grain` provides a high compatibility degree between the document and this category. The category `wheat` has a positive degree because the word "wheat" is indirectly related to "sorghum" and "corn" (through `grain`). The word "shipment" is also directly related to `ship`. This reasoning scheme would be, more or less, the procedure that the expert could have followed to classify the document based on the ontological/semantic knowledge represented by the proximity equations.

4 Experimental Results

Our experiment consists of classifying a set of short texts (news limited up to 160 characters long) selected from Reuters-21578, the most widely used test collection for text categorisation research. The test data set contains 173 pre-classified articles corresponding to 6 overlapped categories (Table 1). The classification accuracy, defined as the percentage of samples correctly classified, was used to determine the performance of our technique.

Table 1. Categories

Category	Documents	Category	Documents
earn	81	acq	58
grain	16	money	17
wheat	7	ship	7

During the analysis of the experimental results, while comparing the percentage of correct classifications with the "incorrect ones", it is important to distinguish between those produced by wrong classifications and those from undone classifications. The first case, wrong classifications, implies a contradiction with the knowledge used by the expert for classifying the texts. The undone classifications is the worst case as it means that one or more definitions are absent from the knowledge base, although they should (or could) be completed in a subsequent phase.

The classification process was carried out with each one of the semantic relations previously defined in Section 2.1, i.e., the Context Neighborhood extracted from Concept Net, and WordNet::Similarity combined with Wikipedia a WordNet. The baseline is represented by the use of the syntatic equality, a category is represented only by its name. Classification results are shown in Table 2. The number of elements not classified by the contextual proximity is not acceptable. The best results were obtained by using the combination of Wikipedia and Wordnet::Similarity, which brings a more complete concept definition of the categories. The approach proposed here is a classification method with a limited complexity and a high dependency on the knowledge base used. Despite of this, our preliminary results could be considered very encouraging because they are better than those results obtained by [1], which also used WordNet and Wikipedia in combination.

Table 2. Accuracy Results

Proximity Relation	Corrects	Wrongs	Unclass.
Wikipedia/Vector	81%	13%	6%
WordNet/Vector	73%	9%	17%
ConceptNet/Context Neighborhood	65%	3%	32%
Baseline	10%	2%	88%

5 Conclusions and Future Work

This paper proposes a fuzzy declarative text categorisation approach which does not need previously labelled documents for training. This method of categorising short texts is based exclusively on domain knowledge instead of training examples. The knowledge used for the classification process could be obtained from generic thesauri and is expressed in a way that is understandable for any non-expert user. The experiment results show that the approach is satisfactory according to the semantic relations used. Thus, the proposed approach could be considered a reasonable methodology for classifying short texts if there is good domain knowledge that allows building up the fuzzy knowledge base. Therefore, the main problem of the approach is that its performance depends on the quality of the category definitions (represented by the proximity equations). In order to solve this problem, in our future work it is necessary to define how closeness relations could be combined and whether such a combination would be helpful, or how to find the appropriate sources of background knowledge. It is also recommendable to carry out new experiments with more specific short text data collections.

Acknowledgments. This research was partially supported by the MICINN (Spain) under TIN2010-20395 and TIN2007-65749 projects and by the JCCM (Spain) under PEIC09-0196-3018, POII10-0133-3516 and PII1I09-0117-4481 projects.

References

1. Barak, L., Dagan, I., Shnarch, E.: Text categorization from category name via lexical reference. In: Proc. of the NAACL 2009, pp. 33–36. Association for Computational Linguistics, Morristown (2009)
2. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
3. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. 43(5-6), 907–928 (1995)
4. Julián, P., Rubio, C.: A Declarative Semantics for Bousi∼Prolog. In: Proc. of the PPDP 2009, pp. 149–160. ACM, New York (2009)
5. Julián, P.: A procedure for the construction of a similarity relation. In: Proc. of the IPMU 2008, pp. 489–496. U. Málaga (2008)
6. Julián, P., Rubio, C.: An efficient fuzzy unification method and its implementation into the Bousi∼Prolog system. In: Proc. of the FUZZ-IEEE 2010, pp. 658–665 (2010)
7. Liu, H., Singh, P.: Commonsense reasoning in and over natural language. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3215, pp. 293–306. Springer, Heidelberg (2004)
8. Pedersen, T., Patwardhan, S., Michelizzi, J.: Wordnet:similarity: measuring the relatedness of concepts. In: HLT-NAACL 2004. pp. 38–41 (2004)
9. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. 34(1), 1–47 (2002)
10. Zadeh, L.: Fuzzy sets. Information and Control 8, 338–353 (1965)

Subtractive Initialization of Nonnegative Matrix Factorizations for Document Clustering

Gabriella Casalino[1], Nicoletta Del Buono[2], and Corrado Mencar[1]

[1] Dipartimento di Informatica,
[2] Dipartimento di Matematica,
Università degli Studi di Bari Aldo Moro, Via E. Orabona 4, I-70125 Bari, Italy
gabriella.casalino@gmail.com, delbuono@dm.uniba.it, mencar@di.uniba.it

Abstract. Nonnegative matrix factorizations (NMF) have recently assumed an important role in several fields, such as pattern recognition, automated image exploitation, data clustering and so on. They represent a peculiar tool adopted to obtain a reduced representation of multivariate data by using additive components only, in order to learn parts-based representations of data. All algorithms for computing the NMF are iterative, therefore particular emphasis must be placed on a proper initialization of NMF because of its local convergence. The problem of selecting appropriate starting initialization matrices becomes more complex when data possess special meaning, and this is the case of document clustering. In this paper, we present a new initialization method which is based on the fuzzy subtractive scheme and used to generate initial matrices for NMF algorithms. A preliminary comparison of the proposed initialization with other commonly adopted initializations is presented by considering the application of NMF algorithms in the context of document clustering.

1 Introduction

Several applications store pertinent information in a huge matrix which is often non-negative. Examples are documents in document collections, which are stored as columns of term-by-document matrix, whose elements count the number of times (possibly weighted) a corresponding term appears in a selected document. Similarly, in image collections, each image is represented by a vector whose elements correspond to the intensity and/or the color of the image pixels. In recommender systems, the information for a purchase history of customers or ratings on a subset of items is stored in a non-negative sparse matrix.

Three common goals can be identified when mining information from non-negative matrices: to automatically cluster similar items into groups, to retrieve items most similar to a user query, to identify interpretable critical dimensions within the collection.

Taking into account the non-negativity constraint, benefits in terms of meaningful interpretations of the obtained model can be added to any data analysis process. Nevertheless, classical tools are not able to guarantee the conservation of the non-negativity.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 188–195, 2011.

Recently, non-negative matrix factorization (NMF) received an increasing attention from the data analysis community due to its capabilities of obtaining a reduced representation of data only using positive restrictions [9]. These constraints led to a part-based representation, because they allow nonnegative linear combination of a set of nonnegative "bases" that represent realistic "building blocks" for the original data. More formally, given an initial set of data expressed by a $n \times m$ matrix X, whereas each entries X_{ij} represents in a broad sense the score obtained by the entity j on the variable i, a NMF consists in approximating (generally, in terms of Frobenius norm) the matrix X with the product of two reduced rank nonnegative matrices, the $n \times r$ basis matrix W, and the $r \times m$ encoding matrix H (with rank factor $r < \min(m, n)$), so that $X \approx WH$.

In this way, the perception of the whole, being it an image or a document in a collection, becomes a combination of its parts represented by basis vectors. Particularly, in the standard vector space model context, when X represents a term-by-document matrix, the basis vectors identify a set of words denoting a particular concept or topic and each column of H contains an encoding of the linear combination of basis vectors approximating the corresponding column of X. Hence, each document is viewed as combination of basis vector and it can be categorized as belonging to a specific topic. So, nonnegative factors of NMF can be directly applied to perform partitional clustering that identifies semantic features in a document collection and groups the documents into clusters on the basis of shared semantic features [8,10,11]. Moreover, this factorization can be used to compute a low rank approximation of a large sparse matrix along with preservation of natural data non-negativity.

All algorithms for computing the NMF are iterative and require initialization of the basis and encoding matrices. Therefore, the efficiency of many NMF algorithms is affected by the selection of the starting matrices: poor initialization often results in slow convergence or lower error reduction. Furthermore, the problem of selecting appropriate initializations becomes more complicated when certain structures or constraints are imposed on the factorized matrices or when the data possess special meaning as in the context of document clustering. Different initialization mechanisms have been proposed in literature: some of them lead to rapid error reduction and faster convergence of the adopted NMF algorithm, others lead to better overall error accuracy at convergence. However, there does not exist a definitive suggestion about the best initialization strategy to be adopted for different NMF algorithms [6].

In this paper, we propose the use of the subtractive clustering [3], a fast method for estimating clusters in the data, as a basis scheme to generate the initial matrices $W^{(0)}$ and $H^{(0)}$ for any NMF algorithm. Each obtained cluster center can be directly translated into columns of the initial basis matrix $W^{(0)}$, while elements in the encoding matrix can be obtained as fuzzy membership degree of each data to each cluster. With respect to other cluster methods, such as k-means, widely used for NMF initialization, subtractive clustering could also be used to suggest the proper rank factor, when average distance between document data is estimated.

The rest of the paper is organized as follows. In the next section, we briefly review some of the NMF algorithms commonly used for document clustering together with some initialization strategies. Then, we illustrate the subtractive clustering and how it can be adopted to generate $W^{(0)}$ and $H^{(0)}$. In section 4, we report the results obtained from different NMF algorithms initialized by the proposed approach in clustering a subset of the Reuters data corpus. Comparisons with some usually adopted initialization techniques and evaluations of the obtained clusters are also reported. Finally, some conclusive remarks and guidelines for future work are sketched in section 5.

2 NMF Algorithms and Classical Initializations

A NMF of a given data matrix X can be obtained by finding a solution of a non-linear optimization problem over a specified error function. The most frequently adopted error function is the squared Euclidean distance which leads to the minimization of the functional $\|X - WH\|_F^2$ subject to the non-negativity constraints over the elements W_{ij} and H_{ij}.

The most popular approach to numerically solve the NMF optimization problem is the multiplicative update algorithm (NMFLS) proposed in [9]. It can be shown that, starting from some nonnegative initial matrices, the square Euclidean distance is non-increasing under the following iterative update rules:

$$H_{ij} \leftarrow H_{ij} \frac{(W^\top X)_{ij}}{(W^\top WH)_{ij} + \varepsilon} \qquad W_{ij} \leftarrow W_{ij} \frac{(XH^\top)_{ij}}{(WHH^\top)_{ij} + \varepsilon} \tag{1}$$

where ε is a small positive parameter used to avoid division by zero.

Algorithms following an alternating process, approximating (in the sense of mean squared error) firstly W, then H, and so on, can be also adopted to obtain a NMF of X. Particularly, starting from some nonnegative initialization of W, an elementary Alternate Least Square algorithm (ALS) [2] for minimizing the square Euclidean distance measure is:

$$\begin{array}{l} -\text{Solve matrix equation : } W^\top WH = W^\top X \text{ w.r.t } H \\ -\text{Set to 0 negative elements in } H \text{ (\textit{projection step})} \\ -\text{Solve matrix equation : } HH^\top W^\top = HX^\top \text{w.r.t } W \\ -\text{Set to 0 negative elements in } W \text{ (\textit{projection step})} \end{array} \tag{2}$$

Different modifications of the standard cost functions have been proposed to include further constraints on the factors W and/or H, such as sparsity or orthogonality. A nonnegative sparse encoding scheme (NMFSC), proposed in [7], has the peculiarity of controlling the statistical sparsity of the H matrix in order to discover parts-based representations that are qualitatively better than those given by standard NMF. Orthogonal nonnegative matrix algorithms (ONMF) attempt instead, to obtain the basis or the encoding matrix with columns as orthogonal as possible, to minimize the number of basis components required to represent the data and the redundancy between different bases [4].

2.1 Initialization Mechanisms

All algorithms for NMF are iterative and require the computation of initial matrices $W^{(0)}$ and/or $H^{(0)}$ by some numerical mechanism and then alternately update W and H until there is no further appreciable change in the objective function, yielding locally optimal solutions. The initial pair $(W^{(0)}, H^{(0)})$ plays a crucial role for the convergence speed of the iterative algorithm and to improve algorithm performance. Moreover, when NMF is applied to document clustering, initial matrices should also posses meaningful interpretations.

Initialization schemes can be classified in simple mechanisms, based on some kind of randomization, and complex schemes based on some alternative low rank factorization or clustering algorithms. The former class includes: (i) the random initialization which produces dense matrices $W^{(0)}$ and $H^{(0)}$ of dimension $n \times r$ and $r \times m$, respectively, with elements randomly generated in [0,1], (ii) several variants of random choices of columns in X used to build $W^{(0)}$ together with random or zeros initialization of $H^{(0)}$. Complex initialization strategies exploit clustering algorithms or some alternative low rank factorization scheme to construct the initial pair $(W^{(0)}, H^{(0)})$. Among this class, we can enumerate spherical k-means initialization (kmeans) [12], Fuzzy C-Means (FCM) initialization [13], and Nonnegative Double Singular Value Decomposition (NNDSVD) based on two SVD processes [1]. Generally speaking, complex initialization strategies, require a higher computational costs, but they produce a fast error reduction, a high convergence rate in NMF algorithms and reduce to the minimum or definitely do not require the use of any randomization step.

3 Initialization by Subtractive Clustering

In this section, we briefly describe the initialization scheme based on the subtractive clustering (SC) ([3]) and we illustrate how to generate the initial $(W^{(0)}, H^{(0)})$ for any NMF iterative algorithm. It should be pointed out that, all clustering methods adopted to initialize NMF algorithms need to fix the number of clusters corresponding to the rank factor r, defining the dimensionality of the subspace approximating the data. The SC, instead, is able to automatically discover the most appropriate value of r, when an estimation of distance among data is provided.

Consider the data matrix $X = [X_1, X_2, \ldots, X_n]$, where without loss of generality each column vector $X_j \in \mathbb{R}^m$ is assumed to be normalized to have unit l_2 norm.

The SC assumes each data point is a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the potential of surrounding data points as follows:

$$P_j = \sum_{k=1}^{n} \exp\left(-\frac{4}{r_a^2}\|X_j - X_k\|^2\right), \tag{3}$$

being r_a a positive constant representing a normalized radius defining a neighborhood. According to (3), high potential values correspond to a data point with

many neighborhood data points. Hence, we compute the potential of each data point and then we select the point with the highest potential as the first cluster center. Then, in order to avoid that points near to the first cluster center could be selected as another center cluster, we subtract from each data point an amount of potential proportional to its distance from the first cluster center. After the potential reduction, we select the data point with the highest remaining potential as the second cluster and we further reduce the potential of each data point according to their distance to the second cluster center. Generally, after the k-th center cluster $\tilde{X}_k$ has been obtained with potential $\tilde{P}_k$, we reduce the potential of each data point by:

$$P_j \leftarrow P_j - \tilde{P}_k \exp\left(-\frac{4}{r_b^2}\|X_j - \tilde{X}_k\|^2\right), \quad j = \ldots, n, \tag{4}$$

where r_b is a positive constant (typically chosen as $r_b = 1.25r_a$). The process of finding new cluster center and reducing potential of all data iterates until the remaining potential of all data points is bounded by some fraction of the potential $\tilde{P}_1$ of the first center cluster. The stopping criterion usually adopted is $\tilde{P}_k < 0.15\tilde{P}_1$.

After the stopping criterion is satisfied, the SC applied to a term-by-document matrix provides: the number r of clusters, the cluster centroids and their potential value $\tilde{P}_k$, $k = 1, \ldots, r$.

The initial matrices $W^{(0)}$ and $H^{(0)}$ are constructed as follows. The basis matrix collects the cluster centroid vectors $\tilde{X}_k$ ordered by decreasing values of their potential $\tilde{P}_k$, i.e., $W^{(0)} = [\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_r]$. The encoding matrix $H^{(0)}$ provides the degree to which each document is assigned to each cluster. Particularly, the elements $H^{(0)}_{kj}$, $k = 1, \ldots, r$ and $j = 1, \ldots, m$, provide the fuzzy membership value for the j-th document in the k-th cluster and are computed by

$$H^{(0)}_{kj} = \frac{\exp\left(-\frac{1}{2}\frac{\|X_j - W^0_{*k}\|^2}{\sigma^2}\right)}{\sum_{i=1}^{r}\exp\left(-\frac{1}{2}\frac{\|X_j - W^0_{*i}\|^2}{\sigma^2}\right)} \tag{5}$$

being r the total number of clusters and $\sigma^2 = \frac{{r_a}^2}{8}$. The denominator inside the previous formula represents a normalization which is needed since the reconstruction of a column X_i can be regarded as a weighted average of the centroids $W^{(0)}_{*i}$ with respect to membership values in $H^{(0)}$ (which act as weights). The sum of membership values must be equal to 1 in order to obtain a convex average value.

4 Numerical Experiments

In this section, we illustrate the performance of some NMF algorithms applied on a document clustering problem and we aim to compare the SC initialization with other complex initialization schemes. The initialization strategies have

Table 1. Performance of the NMF algorithms initialized with different strategies applied to cluster data with $k = 10$

	Effectiveness of init.		NMFLS		ALS		ONMF		NMFSC	
Initial.	Time	Init. Err.	Err.	n iter	Err.	n iter	Err.	n iter	Err.	n iter
SC	0.6977	159.73	169.51	160	162.13	59	169.86	34	165.43	500
FCM	2.5053	190.74	163.92	406	164.10	270	165.50	309	167.22	500
NNSVD	0.7167	190.74	165.49	205	164.12	68	166.54	500	167.33	280
Rand	-	$39.51e^6$	164.40	500	164.48	122	165.58	500	167.73	500
K-Means	9.8546	168.73	167.25	81	164.34	170	169.55	66	169.32	156

been compared in terms of both the effectiveness of the starting pair $(W^{(0)}, H^{(0)})$ (evaluated by $\|X - W^{(0)}H^{(0)}\|_F$) and the run-time required to compute the initial factors (evaluated in seconds). The NMF algorithms with different initializations are compared in terms of error reduction and number of iterations. All the numerical results have been obtained by Matlab 7.7 codes implemented on an Intel Core Quad CPU Q6600 2.40 GHz.

The clustering problem is related to a subset of the Reuters data, consisting in 201 documents belonging to 10 categories. The dataset has been pre-processed to remove the stop words by means of a common words dictionary and by applying a stemming algorithm. The term-by-document matrix has been composed using the standards TF-IDF weights, and possesses the 97% of sparsity degree.

Table 1 reports the effectiveness of initialization strategies, together with their run-time values and the results obtained for each NMF algorithm (combined with different initializations) at the end of the learning phase over the Reuters dataset. For a fair comparison among all the algorithms, we adopted the same stopping criteria: a maximum number of iterations (`maxiter=500`) and a fixed tolerance ($toll = 10^{-6}$) for the difference between two subsequent values of the objective function. It should be pointed out that SC shows a good trade-off between run-time values and effectiveness of the initial pair $(W^{(0)}, H^{(0)})$, when compared with other complex initialization schemes. The run-time value for the random initialization has been omitted (being negligible), however this initialization produces full initial matrices with poor accuracy. Moreover, SC determines a higher convergence rate for some NMF algorithms.

The effectiveness of the clusters provided by NMF algorithms, with different initializations, has been evaluated by the Davies-Bouldin index (DBI) [5], given by $\frac{1}{r}\sum_{i=1}^{r} \max_{j:j\neq i} \frac{S_i+S_j}{M_{ij}}$, where S_i and S_j are the inter cluster similarity of the ith and jth cluster and M_{ij} is their intra-cluster separability. Figure 1 illustrates the behavior of the DBI, when the cluster number increases. Small values of DB correspond to clusters that are compact and whose centers are far away from each other. The graphs also suggest that, for almost all initializations, the most appropriate cluster number is 13. It should be also pointed out that SC is able to automatically discover the proper number of clusters when the r_a parameter is set to the mean distance among documents.

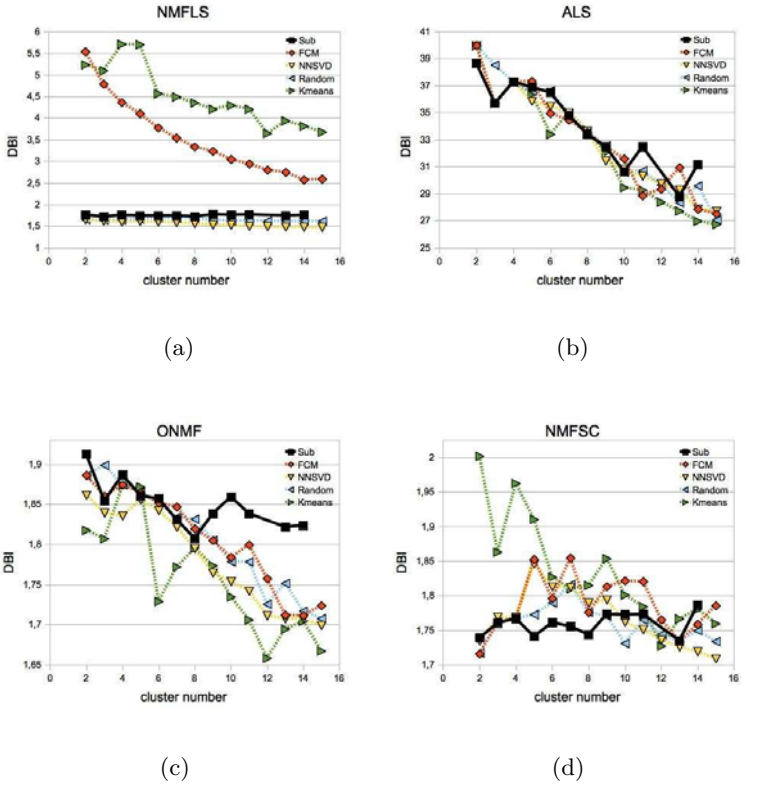


Fig. 1. DBI behavior related to the increasing number of clusters for (a) NMFLS , (b) ALS, (c) ONMF and (d) NMFSC, with different initializations

Table 2. Cluster accuracy and mutual information results for the NMF algorithms initialized with different strategies

	NMFLS		ALS		ONMF		NMFSC	
Initial.	Accuracy	MI	Accuracy	MI	Accuracy	MI	Accuracy	MI
SC	0.45	1.20	0.48	1.27	0.44	1.17	0.48	1.49
FCM	0.45	1.35	0.50	1.54	0.44	1.27	0.45	1.36
NNDSVD	0.48	1.48	0.46	1.38	0.48	1.48	0.50	1.60
Rand	0.41	1.19	0.42	1.25	0.43	1.32	0.47	1.42
kmeans	0.41	1.33	0.45	1.31	0.41	1.33	0.41	1.33

Table 2 reports a further evaluation of the obtained clusters with respect to the set of clusters defined by the original categorized documents in terms of cluster accuracy and mutual information measures. As it can be observed from the table, even if there are no appreciable differences among different initializations, SC and FCM provide the better results for almost all NMF algorithms.

5 Conclusive Remarks

In this paper, we proposed the fuzzy subtractive scheme as initialization for NMF algorithms in the context of document clustering. The proposed method is very fast in constructing initial pairs for NMF algorithms; in some cases it is able to increase the performance of NMF methods both in terms of number of iterations and accuracy of the final approximation. Moreover, differently from other clustering initialization strategies (such as K-Means), the SC method is able to predict the proper number of clusters, and consequently the rank factor for the low rank factorization, when mean distance among documents is provided. Future work can be addressed to assess the performance of NMF algorithms with SC initialization on different datasets as well as to further investigate its capability of predicting the most appropriate factor rank for data.

References

1. Boutsidis, C., Gallopoulos, E.: Svd based initialization: ahead start for nonnegative matrix factorization. Pattern Recognition 41(4), 1350–1362 (2008)
2. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.I.: Alternating Least Squares and Related Algorithms for NMF and SCA Problems. In: Nonnegative Matrix and Tensor Factorizations. John Wiley & Sons, UK (2009)
3. Chiu, S.L.: Fuzzy Model Estimation based on Cluster Estimation. J. Intelligent and Fuzzy Systems 2, 267–278 (1994)
4. Choi, S.: Algorithms for orthogonal nonnegative matrix factorization. Proc. Intern. Joint Conf. Neural Networks (2008)
5. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Machine Intell. 1(4), 224–227 (1979)
6. Del Buono, N., Lucarelli, M.: Comparative studies on initializations for nonnegative matrix factorization algorithms, Tech. Rep. 17/10, Univ. Bari, Italy (2010)
7. Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. J. Machine Learning Research 5, 1457–1469 (2004)
8. Lazar, C., Doncescu, A., Kabbaj, N.: Non Negative Matrix Factorization clustering capabilities; application on multivariate image segmentation. Int. J. of Business Intel. Data Mining 5(3), 285–296 (2010)
9. Lee, D.D., Seung, S.H.: Algorithms for non-negative matrix factorization. In: Proc. Adv. Neural Information Proc. Syst. Conf., vol. 13, pp. 556–562 (2000)
10. Shahnaz, F., Berry, M.W., Pauca, M.P., Plemmons, R.J.: Document clustering using nonnegative matrix factorization. Information Processing and Managements: Intern. J. 42(2), 373–386 (2006)
11. Xu, W., Liu, X., Gong, Y.: Document clustering based on nonnegative matrix factorization. In: Proc. SIGIR, pp. 267–273 (2003)
12. Xue, Y., Tong, C.S., Chen, Y., Chen, W.-S.: Clustering-based initialization for non-negative matrix factorization. Appl. Math. and Comp. 205, 525–536 (2008)
13. Zhenga, Z., Yang, J.: Initialization enhancer for non-negative matrix factorization. Eng. Appl. Art. Int. 20, 101–110 (2007)

Asymmetric Kernel Scaling for Imbalanced Data Classification

Antonio Maratea and Alfredo Petrosino

Department of Applied Science
University of Naples "Parthenope", Isola C4, Centro Direzionale, Napoli, Italy
(antonio.maratea,alfredo.petrosino)@uniparthenope.it

Abstract. Many critical application domains present issues related to imbalanced learning - classification from imbalanced data. Using conventional techniques produces biased results, as the over-represented class dominates the learning process and tend to naturally attract predictions. As a consequence, the false negative rate may result unacceptable and the chosen classifier unusable. We propose a classification procedure based on Support Vector Machine able to effectively cope with data imbalance. Using a first step approximate solution and then a suitable kernel transformation, we enlarge asymmetrically space around the class boundary, compensating data skewness. Results show that while in case of moderate imbalance the performances are comparable to standard SVM, in case of heavily skewed data the proposed approach outperforms its competitors.

1 Introduction

Data imbalance is a well known problem in the data mining community that severely biases the learning process and badly affects the performances of the mining algorithms. While both class imbalance and instance imbalance share many common aspects - and affect both the supervised than the unsupervised framework - the focus of our work is in instance imbalanced classification. Such problem is crucial in many critical application domains, for example: intrusion detection, fire alarm systems, gene selection, satellite data analysis, medical diagnosis and in all situations in which available instances of one class vastly outnumber instances of other classes. A good survey of the general problem framework may be found in [10]. The lack of data in minority class may be due to extremely expensive or impossibly uniform data gathering process, natural rarity, biased and/or partial information sources, errors, uneven sensor positioning etc. Nonetheless, rare cases are often the most interesting and valuable for the data analyst. In most real world critical applications the cost of misclassification of negative instances far exceeds the positive ones. For example in all aspects of security, the system designer surely prefers to have a reasonable amount of false alarms instead of just one undetected violation; in diagnosis of severe diseases, the doctor surely prefers to have a reasonable amount of patients wrongly alerted instead of some of them infected and undiagnosed. Leaving a classifier learn on an imbalanced dataset without corrections will produce a classification

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 196–203, 2011.

biased towards the over-represented class and, as a side-effect, will skew the class boundary towards the under-represented class.

The aim of our paper is to propose a modification of the SVM algorithm able to balance the skewed data distribution. Nonlinear SVM are soft margin classifiers and hence our approach can be classified in the broad area of soft computing. For sake of simplicity, we consider a two class problem, but the approach is easily extensible to n classes through the classics One-Versus-One/One-Versus-All rules, or preferably through one of the more recent SVM aggregation methods proposed in literature (for example the multi class tree structured SVM proposed in [8], that seems to help by itself with unequal data distribution). Differently from many others (see related work), our approach is not based on preprocessing the data through a more or less "smart"resampling strategy: we propose an a posteriori kernel transformation that unevenly expands distances in the feature space in proximity of the boundary region. Based on real data testing with different percentages of imbalance - once parameters are well setted - the method has shown to be very effective, especially in extreme cases.

The paper is organized as follows: the most common variations of the SVM algorithm for imbalanced learning are outlined in next section, where related work is discussed; in section 3 the solution of the SVM algorithm is briefly presented and the proposed variations described in detail; in section 4 used data and experiments setup are described, results summarized; in section 5 main conclusions are drawn and future work is outlined.

2 Related Work

Many approaches have been proposed in literature to allow SVM to better cope with imbalanced data. From a general perspective, preprocessing strategies aim to mask data imbalance to the classifier and are essentially based on oversampling the minority class, undersampling the majority class or a combination of both. Strategies varies by author, some are "smarter"than others, but all share the same problem: oversampling increases artificially computational cost as adds phantom data, while undersampling may discard useful information. A popular method is Synthetic Minority Oversampling Technique [6], which rooted many variations (i.e. [1,5]).

Algorithm modifications on their side aim to bias the learning process in order to reduce majority class domination. They act on the boundary position: dynamically during learning, changing or adding some parameters or weights, or statically after learning, following proportionality criteria.

In literature the label "asymmetric Support Vector Machine"is often used to indicate Cost Sensitive Learning strategies acting on slack variables, but may broadly (and loosely) refer to any combination of the approaches previously presented. In our method, we focus on kernel modification strategies on a two class problem. We will not use data preprocessing, nor to generate, nor to discard any data. Our purpose is to modify the Kernel K in a way equivalent to asymmetrically changing the spatial resolution around the boundary.

We will not consider here the recent family of one-class classifiers, as while learning with only one target class (or only samples from one class) may be seen as an extreme case of imbalancing, data gathering and processing is actually different from learning two or more classes with heavily skewed sample distribution. In our opinion, one class classifiers should not be used when multi class labels are available.

3 Method

The classical solution of the SVM optimization problem for binary classification is the hyperplane given in following equation (see [9]):

$$f(x) = \sum_{s \in SV} \alpha_s y_s \, K(x_s, x) + b \tag{1}$$

Instance class depends on the side of the hyperplane in which each point is located, formally by the sign of $f(x)$. *Support vectors* are by definition the x_i such that $\alpha_i > 0$. If the points are linearly separable each support vector satisfies:

$$f(x_s) = y_s = \pm 1$$

When data are not linearly separable the solution may be found bounding mutipliers α_i with the condition $\alpha_i \leq C$ for - usually big - values of a positive constant C.

It has been observed that the kernel $K(x, x')$ induces a riemannian metric in the input space S through mapping ϕ [4]. The metric tensor induced by K on $x \in S$ is

$$g_{ij}(x) = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x'_j} K(x, x') \mid_{x'=x} \tag{2}$$

where K is the internal product $K(x, x') = \Phi(x)\, \Phi(x')$ in some higher dimensional feature space H, where ϕ is a mapping from S to H.

The volume element with respect to this metric is:

$$dV = \sqrt{g(x)} dx_1 ... dx_n \tag{3}$$

where $g(x)$ is the determinant of the matrix whose i, j element is g_{ij}. The magnification factor $\sqrt{g(x)}$, expresses a local volume expansion under mapping ϕ.

Many kernels have been studied in literature [7] and the optimal choice of kernel is an active area of research. In the following we will consider the gaussian RBF kernel.

3.1 Conformal Kernel Transformations

A conformal transformation is a transformation that preserves local angles. To improve SVM discrimination power, authors in [2] proposed a (quasi) conformal

transformation on the kernel, with the purpose of increasing the separation between the two classes close to the boundary, so to enlarge resolution in this area. The general transformation form is:

$$K'(x, x') = D(x)K(x, x')D(x') \tag{4}$$

for a suitable definite positive function $D(x)$. If $D(x)$ and K are gaussian, then K' will also be gaussian. In this case K' satisfies Mercer condition. The specific transformation function considered in [2] follows:

$$D(x) = \sum_{x_i \in SV} e^{-k||x-x_i||^2} \tag{5}$$

where k is a positive constant. Support Vectors are by definition close to the boundary, so enlargement is applied in proximity of them, indirectly involving the boundary. The main problem is that they are not known in advance and so a two step approach is needed: first a standard SVM is needed to find an approximate solution (set of SV), then a second SVM is executed using the new transformed kernel. Such a choice can be very sensitive to SV distribution, enlarging more areas with high density of SV. Cited work does not explicitly address the data imbalance issue.

The paper [12] tries to overcome the problem of support vector distribution using a different adaptive (quasi) conformal transformation:

$$D(x) = \sum_{x_i \in SV} e^{-k_i||x-x_i||^2} \tag{6}$$

In this case each support vector has a different weight k_i that can be used to control its influence on space dilation. Considering a variable radius neighborhood for each SV and averaging the distance in feature space between the given x_i and all neighbor SV that have different labels, authors claim it is possible to calculate k_i so to compensate for irregular spatial distribution of SV. They also suggest that k_i can be used to compensate for class imbalance, assigning bigger values to the SV of the majority class.

From the same authors, an extension to non vector and non fixed dimension data is presented in [13]. The proposed transformation is:

$$D(x) = \frac{1}{|\chi_b^*|} \sum_{x_b \in \chi_b^*} e^{-k_b||\phi(x)-\phi(x_b)||^2} \tag{7}$$

where χ_b^* is the set of interpolated boundary instances and ϕ is the implicit transformation from input space S to feature space H. The first notable difference is in the way data imbalance is treated: observing that imbalanced data produce a separating hyperplane that is skewed towards the minority class, authors suggest as boundary an hyperplane found by interpolation between the center hyperplane and the majority SV hyperplane. To estimate boundary hyperplane shifting from

the center, a cost function balancing losses of false positives and false negatives is proposed. The second notable difference is that distances are computed in the feature space, instead of the input space, allowing comparison of data of different lengths and even non vectorial. The main problem of both methods [12,13] is their huge computational cost.

A similar, but more straightforward and efficient approach, called Kernel Scaling (KS) is proposed in [11]. After a preliminary standard SVM, a second one is executed with a transformed kernel based on the plain distance from the separating hyperplane, instead of support vectors. In this way points that are further from the boundary are shifted minimally, while points close to the boundary are subject to maximum shifting. The concentration of support vectors in an area is irrelevant. The proposed transformation function is:

$$D(x) = e^{-kf(x)^2} \tag{8}$$

where $f(x)$ is given by 1 and k is a positive constant independent from data. $D(x)$ reaches it maximum on the boundary surface, where $f(x) = 0$, and decays smoothly to e^{-k} on the margins, where $f(x) = \pm 1$. Data imbalance is not explicitly addressed.

3.2 Asymmetric Kernel Scaling (AKS)

The method proposed in [11] proved to be robust and efficient, but does not account for imbalanced data. We propose an extension of it that can effectively manage a markedly different number of training instances in the two classes. The basic idea is to enlarge differently areas on the two sides of the boundary surface, so to compensate for its skewness towards minority instances.

We first perform a standard SVM to compute an approximate boundary position, then we split the points in two sets, the negatives χ^- and the positives χ^+, according to first step prediction. In the second step, the applied kernel transformation function is:

$$D(x) = \begin{cases} e^{-k_1 f(x)^2}, & \text{if } x \in \chi^+ \\ e^{-k_2 f(x)^2}, & \text{if } x \in \chi^- \end{cases} \tag{9}$$

where $k_1 > k_2$ considering the positives as the majority class. In this way space enlargement is different on the two sides of the boundary surface, allowing actually to compensate the bias due to data imbalance. Classification is performed using the transformed kernel and a suitable value for k_1 and k_2. Concerning the problem of estimation of parameters, while it may seem reasonable to connect their value to the size of input data of each class, no evidence of an explicit relation emerged in the experiments. We used grid search and cross validation to find optimal values. The proposed solution is more flexible with respect to [11], includes it as a special case ($k_1 = k_2$), and proved to be effective, especially in extreme cases.

4 Data, Experiments and Results

Data used in the experiments are from a dataset downloadable from the UCI Machine Learning repository[1]. This dataset is composed of a range of biomedical voice measurements from 31 people, 23 of which are affected by Parkinson's disease (PD) [3]. Each column in the table is a voice measurement, and each row corresponds to one of 195 voice recording from these individuals. The most common learning task on these data is to discriminate healthy people from those with PD, according to "status" column, which is set to 0 for healthy and 1 for PD. To test method robustness, artificially varying percentages of skewness have been obtained, further removing minority class instances from the already imbalanced PD data. 4 block of tests have been performed: the first one considering the whole dataset; the second one removing 24 instances; the third one removing 36 instances; the fourth one removing 42 instances (extreme imbalancing), always and only from minority class. Effects of parameters is shown in figure 1.

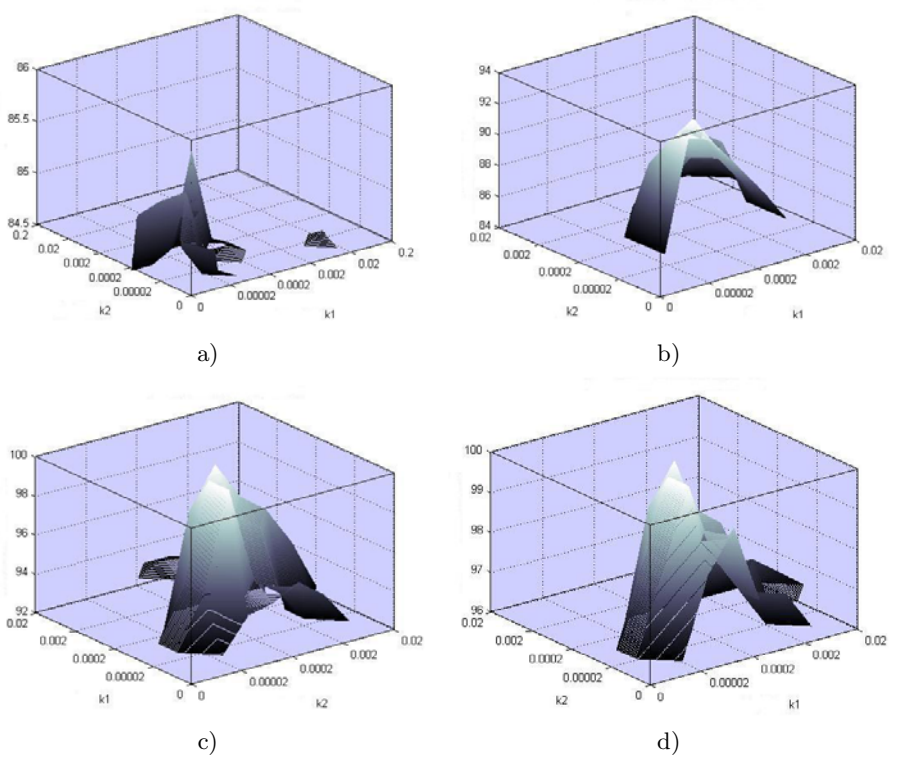


Fig. 1. Average accuracy on test sets for different sizes of the minority class a) whole dataset; b) minority class less 24 instances; c) minority class less 36 instances; d) minority class less 42 instances

[1] http://archive.ics.uci.edu/ml/datasets/Parkinsons

In the first step we perform a standard SVM classification, for which we used a gaussian kernel with base 0.5 and $C = 10$. In the second step we performed the classification with the transformed kernels. Best accuracy has been obtained with $k_1 = 2.04 \times 10^{-6}$ and $k_2 = 2 \times 10^{-5}$. These values were obtained through grid search and 5-fold cross validation.

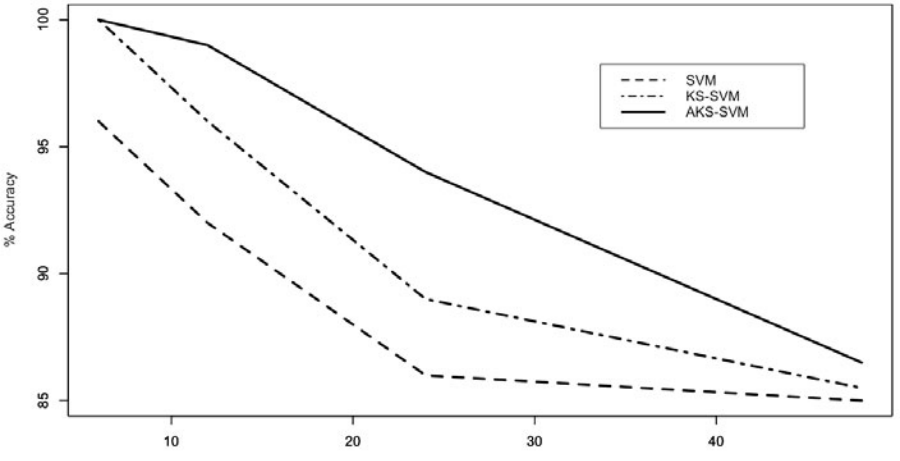


Fig. 2. X axis corresponds to the number of samples of the under-represented class, while Y axis represents the global accuracy. Classic SVM: dashed line; [11] SVM: dashed-dotted line; our algorithm: solid line.

Comparing performances with both standard SVM and the method proposed in [11], we have seen that If the data are not heavily imbalanced, there is not a remarkable advantage in using our approach, but when very few instances remain in the minority class, then our method markedly outperforms its competitors (see figure 2). We point out that the accuracy showed to be very sensitive to a good choice of parameters. Out of a narrow interval of k_1 and k_2 of effective improvement, performance tends to drop to standard SVM.

5 Conclusions

In many challenging emerging applications, like fraud detection, genetic data analysis or video classification, data are often imbalanced, and so is misclassification cost. Using conventional techniques produces biased results, as the over-represented class dominates the learning process and tend to always prevail. We presented a classification procedure based on Support Vector Machine able to effectively cope with data imbalance that is a generalization of [11]. On the basis of real medical diagnosis data, we have shown that the more the distribution is skewed, the more the proposed compensation is effective in improving the performance of the Support Vector Machine. The method has two free parameters

- whose choice critically affects performances - that have been empirically estimated. Future work is in studying a more effective estimation procedure for parameters and different kernels.

Aknowledgments. The authors thank Emanuele de Carlo for the help in code development and experiments' execution during his bachelor's degree at the University of Naples "Parthenope".

References

1. Akbani, R., Kwek, S., Japkowicz, N.: Applying support vector machines to imbalanced datasets. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 39–50. Springer, Heidelberg (2004)
2. Amari, S., Wu, S.: Improving svm classifiers by modifying kernel functions. Neural Networks 12(6), 783–789 (1999)
3. Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O.: Accurate telemonitoring of Parkinsons disease progression by non-invasive speech tests. IEEE Transactions on Biomedical Engineering (2009) (to appear)
4. Burges, C.J.C.: Geometry and Invariance in Kernel Based Methods. In: Schlkopf, B., Burges, C.J.C., Smola, A. (eds.) Advances in Kernel Methods - Support Vector Learning. MIT Press, Cambridge (1999)
5. Castro, C.L., Carvalho, M.A., Braga, A.P.: An Improved Algorithm for SVMs Classification of Imbalanced Data Sets. Engineering Applications of Neural Networks, 108–118 (2009)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Philip Kegelmeyer, W.: SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321–357 (2002)
7. Cristianini, N., Shawe-Taylor, J.: An introduction to Support Vector Machines and other Kernel based Learning Methods. Cambridge University Press, Cambridge (2001)
8. Guo, J., Takahashi, N., Hu, W.: An Efficient Algorithm for Multi-class Support Vector Machines. In: International Conference on Advanced Computer Theory and Engineering (ICACTE 2008), Phuket, December 20-22, pp. 327–331 (2008)
9. Vapnik, V.: Statistical Learning Theory. John Wiley and Sons, New York (1998)
10. Weiss, G.M.: Mining with Rarity: A Unifying Framework. ACM SIGKDD Explorations 6(1), 7–19 (2004)
11. Williams, P., Li, S., Feng, J., Wu, S.: Scaling the Kernel Function to Improve Performance of the Support Vector Machine. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) ISNN 2005. LNCS, vol. 3496, pp. 831–836. Springer, Heidelberg (2005)
12. Wu, G., Chang, E.Y.: Adaptive Feature-Space Conformal Transformation for Imbalanced-Data Learning. In: The Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, pp. 816–823 (2003)
13. Wu, G., Chang, E.Y.: KBA: Kernel Boundary Alignment considering imbalanced data distribution. IEEE Transaction on Knowledge and Data Engineering 17(6), 786–795 (2005)

Improving Expert Meta-schedulers for Grid Computing through Weighted Rules Evolution

R.P. Prado, J.E. Muñoz Expósito, and S. García-Galán

Telecommunication Engineering Department in University of Jaén,
Alfonso X el Sabio, 28 Linares, Jaén, Spain
{rperez,jemunoz,sgalan}@ujaen.es

Abstract. Grid computing is an emerging framework which has proved its effectiveness to solve large-scale computational problems in science, engineering and technology. It is founded on the sharing of distributed and heterogeneous resources capabilities of diverse domains to achieve a common goal. Given the high dynamism and uncertainty of these systems, a major issue is the workload allocation or scheduling problem which is known to be NP-hard. In this sense, recent works suggest the consideration of expert schedulers based on Fuzzy Rule-Based Systems (FRBSs) able to cope with the imprecise and changing nature of the grid system. However, the dependence of these systems with the quality of their expert knowledge makes it relevant to incorporate efficient learning strategies offering the highest accuracy. In this work, fuzzy rule-based schedulers are proposed to consider two learning stages where good quality IF-THEN rule bases acquired with a successful and well-known strategy rule learning approach, i.e., Pittsburgh, are subject to a second learning stage where the evolution of rule weights is entailed through Particle Swarm Optimization. Simulations results show that evolution of rule weights through this swarm intelligence -based strategy allows the improvement of the expert system schedules in terms of workload completion and increase the accuracy of the classical genetic learning strategy in FRBSs.

Keywords: Bio-inspired Algorithms, Knowledge Acquisition, Fuzzy Rule-Based Systems, Grid Computing.

1 Introduction

Grid computing is increasingly emerging as a promising distributed platform to solve large-scale computational problems in diverse areas of science and technology which cannot be faced by organizations local resources capabilities as independent entities [1]. A grid consists of a wide set of heterogeneous and geographically distributed resources from several administrative domains that share capabilities and cooperate to overcome current limitations of single machines. Moreover, grids are defined as *"fully dynamic environments with uncertainties"* [2] where resources actively become available, fall down or reduce its capabilities according to diverse administrative policies with time. A critical problem of

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 204–211, 2011.

these systems is how to cope with the uncertainty of resources state to achieve an efficient workload allocation in the NP-hard scheduling problem [3].

Classical scheduling approaches in grid computing such as First Come First Served (FCFS) or EASY-Backfilling (EASY-BF) [4] are based on the management of jobs regarding diverse queue-based algorithms. Nevertheless, new trends are focused on adaptive scheduling approaches which propose the consideration of the grid state in order to prevent or avoid the scheduling strategy performance. In this sense, schedule-based strategies such as Earliest Suitable Gap (ESG) [4], try to provide efficient schedules according to a certain state of the system. However, as pointed out before, the grid state is inherently dynamic and uncertain and thus, a precise characterization of resources current situation is not feasible with the consequent deterioration of traditional schedule-based strategies. Hence, alternative strategies able to tolerate imprecision in systems information and dynamism are pursued. In this regard, the role of Fuzzy Rule-Based Systems (FRBSs) is to be underlined [5]. FRBSs are expert systems that consider the application of fuzzy logic in the description of environment information through fuzzy sets and that incorporate knowledge in the form of IF-THEN rules. These systems have been extensively applied to a wide range of areas such as control of elevator systems [6] or classification in the discrimination of speech and music [7]. Furthermore, FRBSs are increasingly used for the design of scheduling systems for grid computing [8,9] mainly due to their ability to provide a good characterization of resources state and adaptability to system dynamism. However, the scheduler performance highly depends on the quality of its knowledge base and this way, with the learning process, which arises as one of the striking aspects of these systems.

In the learning of FRBSs, the importance of Genetic Algorithms (GAs) must be underlined [5]. Genetic Fuzzy Rule-Based Systems (GFRBSs) [5] suggest the encoding of whole rule bases (RBs) or rules as individuals of a population to be evolved through the application of genetic operators such as selection, crossover or mutation. One of the more successful genetic strategy for rules evolution is Pittsburgh approach [10] where a population of RBs is evolved through a competitive process. Nevertheless, once the expert scheduler has been provided with a good quality knowledge, it could be interesting to entail a fine tuning of the expert system through small modifications in the fuzzy sets that define the universe of discourse of each system feature describing a given state. However, as found in literature [11], a rule weight adjusting process is equivalent to the modification of the membership functions of both antecedent and consequent fuzzy sets. Moreover, this process requires a smaller computational effort when dealing with a relative reduced number of fuzzy rules. In this work, the performance of FRBSs used as meta-schedulers for grid computing is to be improved through the adjusting of fuzzy rule weights learned for the expert system with a successful strategy such as Pittsburgh approach. Furthermore, the utilization of the well-known bio-inspired strategy based on swarm intelligence, Particle Swarm Optimization (PSO) [12], is proposed.

The rest of the paper can be summarized as follows. Section 2 provides an overview of the scheduling problem in grids and the structure of experts systems used as fuzzy meta-schedulers. In Section 3, the learning strategy for the expert system with the utilization of PSO for rules weight optimization is introduced. Section 4 presents simulation results of the proposed fuzzy strategy in both training and validation scenarios. Finally, Section 5 concludes the paper.

2 Background

A computational grid can be defined as a collection of heterogeneous computational resources r which are allocated within several sites or resources domains RD, i.e., $RD_j = \{r_{j,1}, r_{j,2}, \ldots, r_{j,H_j}\}$, where H_j and $r_{j,k}$ represent the number of resources in site j and resource k of site j, respectively. A RD is an independent entity that contemplates its own access and sharing policies [3] and that is associated to other RD in order to cooperate and make up a global computational entity or Virtual Organization, $VO = \{RD_1, RD_2, \ldots, RD_G\}$. Furthermore, the scheduling in grid computing can be considered as a hierarchical problem of two levels. On the one hand, grid meta-schedulers are responsible for the allocation of L users jobs $J = \{J_1, J_2, \ldots, J_L\}$ to participating RDs in the VO, whereas on the other hand, local schedulers have to distribute jobs within their own RDs. FRBSs schedulers are characterized by their flexibility and tolerance of uncertainty of resources and they are increasingly emerging for scheduling in large-scale infrastructures as grids. Fig. 1 illustrates the general organization of a fuzzy meta-scheduler in a grid. Three main components can be differentiated corresponding to the structure of Mamdani fuzzy logic systems [5], i.e., *fuzzification*, *inference* and *defuzzification* systems.Also, the associated knowledge base

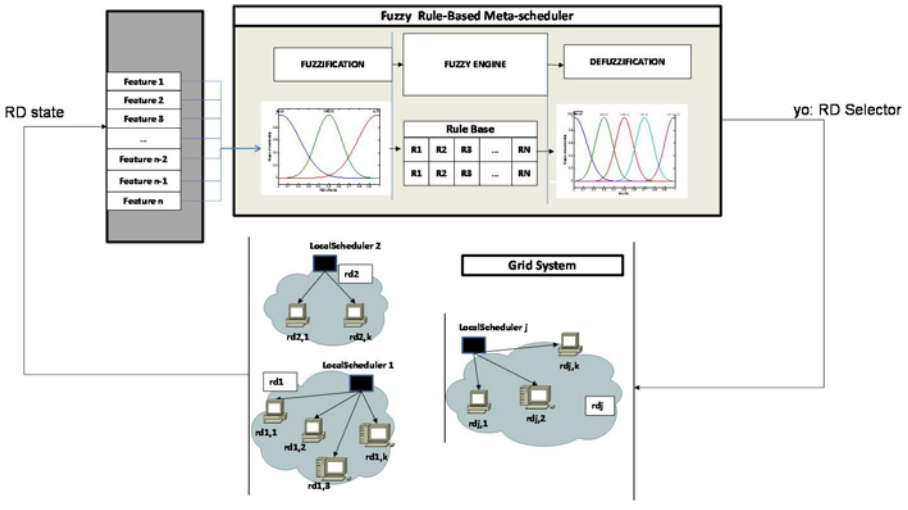


Fig. 1. Fuzzy Rule-Based Meta-scheduler structure

of the expert system can be observed. The performance of the scheduler can be summarized as follows. Initially, a *RD* state, given by a set of predefined features, is collected from the grid environment. Next, in the fuzzification stage, every feature value is associated a linguistic label corresponding to a fuzzy set and thus, the uncertainty in the grid resources description is born in mind. Then, the inference system is responsible for the application of the system knowledge in the form of rules to obtain a the fuzzy output of the expert system. This output represents the suitability of the RD to be selected in the next schedule. Finally, the defuzzification system transforms this fuzzy output into a crisp value that provides a quantifiable index or *RD* selector y_o. This process is repeated for every *RD* and the one presenting a higher suitability is selected for the next schedule. It must be highlighted that the scheduling strategy is founded on the application of fuzzy rules to a given system state and thus, the quality of this base is critical for the whole system performance.

3 Acquisition of Knowledge Bases for Meta-schedulers

As introduced before, the expert meta-scheduler is to be learned through a two step learning process. On the one hand, the acquisition of high quality RBs is entailed through the application of a success full strategy in FRBSs. On the other hand, the accuracy of the RBs is increased considering the evolution of the obtained rule weights through a bio-inspired process. First, the encoding of the fuzzy rules is presented. The expert meta-scheduler knowledge is presented in the form of Mamdani type rules [5]. Rules consists of antecedents and a consequent, that represent the activation condition and output, respectively. Thus, rules are encoded as

$$R_i = IF\, x_1 is\, A_{1l}\, and/or \ldots\, x_n is\, A_{nl}\, THEN\, x_o\, is\, B_l \tag{1}$$

where $(x_1, \ldots, x_n)$ represents the *RDs* input features, A_{nl} and B_l indicate the associated fuzzy sets for feature x_n and output for rule i, respectively, and l is bounded to the number of fuzzy sets for input and output, NF_{in} and NF_{out}.

3.1 Rule Base Acquisition with Pittsburgh Approach

Pittsburgh approach considers a complete RB as an individual of the population to be evolved through a genetic process [10]. Two main systems drive the evaluative process following a prototype organization:

1. *Performance system.* This system evaluates the performance of every individual of the population within the grid environment. Hence, it is responsible of associating a fitness value to every RB in a way that its quality can be distinguished and so, the population can be evolved towards the best suited set of rules.
2. *Rule-Base discovery system.* Three genetic operators are applied at the level of RBs to allow the learning process in every generation:

- Selection. In every generation, a collection of RBs of the population is selected in order to conform a new population for the following generation. These RBs are chosen considering the obtained fitness value in the *Performance System* showing the suitability of every individual.
- Crossover. This genetic operator is considered in order to modify the genetic code of RBs or individuals through generations by the mixing of the genome of the selected individuals. Thus, this operator considers the combination of RBs in the same way that reproduction in biological is addressed.
- Mutation. Mutation is generally introduced in GAs with the aim of keeping and introducing diversity in the population between generations and to prevent local minima.

3.2 Evolution of Weighted Fuzzy Rules

Once a RB is obtained through the classical learning process, the incorporation of rule weights is considered in a way that the encoding of rules is reformulated as

$$R_i = IF\, x_1 is\, A_{1l}\, and/or \ldots\, x_n is\, A_{nl}\, THEN\, x_o\, is\, B_l\, with\, w_i \tag{2}$$

where w_i represents the associated weight to rule i varying between 0 and 1. In order to obtain an efficient weight for every rule, an optimization strategy based on PSO is considered [12]. With this aim, a set of particles x is born in mind where a particle consists of a real number vector ranging from 0 to 1 that concerns as many components as the number of rules in the RB. In this way, particles are moved within the search space following the classical process in PSO, i.e., velocity updating of particles considering particles own inertia, the best position found by the particle and the swarm global best result. To be precise, particles x and velocity v for particle j in iteration k is formulated as

$$v_j^{(k+1)} = c_1 \cdot v_j^k + c_2 \cdot r_1 \cdot \left(Pb_j^k - x_j^k\right) + c_3 \cdot r_2 \cdot \left(Gb^k - x_j^k\right) \tag{3}$$

$$x_j^{(k+1)} = x_j^k + v_j^{k+1} \tag{4}$$

where r_1 and r_2 represent random variables following a normal distribution in the interval (0,1), c_1 is a scalar factor denoting the inertia weight and c_2 y c_3 indicate real factors. Also, Pb_j^k represents the best position reached by particle j until the iteration k and Gb^k is the best location found by the whole swarm. As a result, the particle or rule weight vector achieving the greater accuracy of the RB in the bio-inspired process is selected at the end of the process.

4 Simulations Results and Discussion

Simulations with Alea software [13] are conducted to evaluate the proposed two-stages learning approach and the fuzzy meta-scheduler performance. Alea is a

GridSim -based toolkit for the simulation of scheduling techniques within grid environments which allows the usage of real scenarios setting and workload traces. Particularly, in this work, the proposed scenario is based on the Czech National Grid Infrastructure Metacentrum [14], a computational infrastructure that cooperates in the achievement of a high performance platform by the coordination of diverse institutions resources all around the world. The Metacentrum scenario integrates a set of 210 machines, an overall collection of 806 CPUs running Linux geographically distributed among 14 RDs. On the other hand, workload characterization is retrieved from real traces corresponding to January 2009 which can be found at [14]. To be precise, 2000 jobs are considered for the training of the fuzzy meta-scheduler where *makespan* [3] is proposed as performance indicator. In addition, workload is increased by 16.84% for validation. On the other hand, RDs state is characterized by seven variables as shown in Table 1. Furthermore, these features are described by three gaussian-shaped fuzzy sets, $NF_{in} = 3$. Also, rules consequent is represented by five gaussian-shaped fuzzy sets, $NF_{out} = 5$.

Table 1. Grid system input features

Feature	Description
Number of free processing elements (FPE)	Number of free processing element within RD_i.
Previous Tardiness (PT)	Sum of tardiness of all finished jobs in RD_i.
Resource Makespan (RM)	Current makespan for RD_i.
Resource Tardiness (RT)	Current tardiness of jobs within RD_i.
Previous Score (PS)	Previous deadline score of already finished jobs in RD_i.
Resource Score (RS)	Number of non delayed jobs so far in RD_i.
Resources In Execution (RE)	Number of Resources currently executing jobs within RD_i.

The first learning stage based on Pittsburgh approach is configured with two-point crossover (i.e., two random cut points born in mind for the mixing of individuals genome [5]), decreasing exponential mutation factor (i.e., in a way that diversity is reduced at the final stage of the process), elitist selection $\lambda = 0.9$ (i.e., those individuals presenting the highest accuracy form part of the next generation population with rate λ), population size of 20 RBs and initial maximum size for RBs set to 20 rules. Simulations are conducted for 100 generations and 30 experiments. On the other hand, the second learning stage considers a swarm of 20 particles with $c1 = 0.7$, $c2 = 3$, $c3 = 3$ and rule weights ranging from 0 to 1 in a set of 200 iterations. Moreover, 30 simulations are born in mind for every stage.

Simulations results concerning both training and validation scenarios and the different stages are presented. Table 2 shows simulations results in terms of makespan of the proposed learning strategy when the second stage is faced considering an average quality RB obtained by Pittsburgh approach. As expected, it can be observed the evolution of rule weights with PSO improve the final accuracy of the first stage. Furthermore, this improvement is kept in validation scenario. To be precise, the evolution of rule weights lets the system increases its

Table 2. Learning and validation results with average Pittsburgh RB

Fitness: Makespan(s). Starting point in second learning stage: Average Pittsburgh RB		
Scenario	Training	Validation
Pittsburgh	1654505.400	2420275.128
Pittsburgh+PSO-Weighted Rules (average)	1648315.715	2369031.237
Standard Deviation	97.568	408.674
Pittsburgh+PSO-Weighted Rules (best result)	1627948.1	2416116.088

Table 3. Learning and validation results with best Pittsburgh RB

Fitness: Makespan(s). Starting point in second learning stage: Best Pittsburgh RB		
Scenario	Training	Validation
Pittsburgh	1628982.096	2415910.136
Pittsburgh+PSO-Weighted Rules (average)	1627962.711	2382142.421
Standard Deviation	38.246	371.883
Pittsburgh+PSO-Weighted Rules (best result)	1625559.100	2415424.136

final result by 2.11% on average. Also, the performance of the strategy is evaluated when the best RB in the genetic strategy is considered as starting point of the second stage. Table 3 illustrates that the fuzzy meta-scheduler is able to outperform makespan results of the best RB by 1.29% on average in validation tests through the adjusting of rule weights. Furthermore, is must be pointed out that the best acquired knowledge with the proposed strategy improves the best RB achieved with the classical learning approach in the first stage.

5 Conclusions and Future Work

Fuzzy rule-based meta-schedulers have proved their effectiveness for scheduling in grid computing. However, given the high dependence of these systems with their associated knowledge, the learning process arises as a major challenge of these strategies. In this work, the improvement of fuzzy rule-based meta-schedulers performance is entailed through the consideration of a two-stage learning process. To be precise, the proposed learning strategy suggests the evolution of RBs with a well-known genetic learning strategy, Pittsburgh approach, as to then increase the obtained RBs accuracy with a PSO-based optimization process that evolves and adjusts rule weights. Simulation results show that the suggested learning strategy improves the genetic learning strategy in terms of final result by 2.11% on average considering an average quality RB as starting point of the swarm-based strategy and 1.29% on average regarding the best RB acquired by the genetic approach in the first stage. Hence, both accuracy in final result of the learning process and simplicity of the evolution of rule weights support the use of this strategy for knowledge acquisition in fuzzy rule-based meta-schedulers. In future work, new learning strategies will be analyzed as to increase expert schedulers accuracy for grid computing.

Acknowledgments. This work has been financially supported by the Andalusian Government (Research Projects P07-TIC-02713 and P06-SEJ-01694).

References

1. Foster, I., Kesselman, C.: The Grid 2: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers Inc., San Francisco (2003)
2. Klusacek, D.: Dealing with Uncertainties in Grids through the Event-based Scheduling Approach In: Fourth Doctoral Workshop on Mathematical and Engineering Methods in Computer Science (MEMICS 2008) Vol. 1, Ing. Zdeněk Novotnỳ CSc., Ondráčkova 105, 628 00 Brno Further information, pp. 978–80 (2008)
3. Xhafa, F., Abraham, A.: Meta-heuristics for grid scheduling problems. In: Xhafa, F., Abraham, A. (eds.) Metaheuristics for Scheduling: Distributed Computing Environments. SCI. Springer, Germany (2008)
4. Klusáček, D., Rudová, H.: Alea 2: job scheduling simulator. In: Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques, SIMUTools 2010, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, pp. 61:1–61:10 (2010)
5. Cordón, O., Herrera, F., Hoffmann, F., Magdalena, L.: Genetic fuzzy systems: Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific Pub Co, Inc., Singapore (2001)
6. Jamaludin, J., Rahim, N., Hew, W.: Development of a self-tuning fuzzy logic controller for intelligent control of elevator systems. Engineering Applications of Artificial Intelligence 22(8), 1167–1178 (2009)
7. Muñoz-Expósito, J.E., García-Galán, S., Ruiz-Reyes, N., Vera-Candeas, P.: Adaptive network-based fuzzy inference system vs. other classification algorithms for warped lpc-based speech/music discrimination. Eng. Appl. Artif. Intell. 20(6), 783–793 (2007)
8. Franke, C., Hoffmann, F., Lepping, J., Schwiegelshohn, U.: Development of scheduling strategies with genetic fuzzy systems. Appl. Soft Comput. 8(1), 706–721 (2008), doi:10.1016/j.asoc.2007.05.009
9. Prado, R.P., García-Galán, S., Yuste, A.J., Muñoz-Expósito, J.E.: A fuzzy rule-based meta-scheduler with evolutionary learning for grid computing. Engineering Applications of Artificial Intelligence 23(7), 1072–1082 (2010), doi:10.1016/j.engappai.2010.07.002
10. Smith, S.F.: A learning system based on genetic adaptive algorithms, Ph.D. thesis, Pittsburgh, PA, USA (1980)
11. Nauck, D., Kruse, R.: How the learning of rule weights affects the interpretability of fuzzy systems, in: Proc. IEEE International Conference on Fuzzy Systems, Anchorage, pp. 1235–1240 (1998)
12. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks 1995, vol. 4 (1995)
13. Klusacek, D., Matyska, L., Rudova, H.: Alea - grid scheduling simulation environment. In: Wyrzykowski, R., Dongarra, J., Karczewski, K., Wasniewski, J. (eds.) PPAM 2007. LNCS, vol. 4967, pp. 1029–1038. Springer, Heidelberg (2008)
14. C. N. G. Infrastructure, Metacentrum data sets (2009), http://www.fi.muni.cz/~xklusac/index.php?page=meta2009, http://www.fi.muni.cz/~xklusac/index.php?page=meta2009

Generating Understandable and Accurate Fuzzy Rule-Based Systems in a Java Environment

J.M. Alonso and L. Magdalena

European Centre for Soft Computing, Mieres 33600, Spain
{jose.alonso,luis.magdalena}@softcomputing.es

Abstract. Looking for a good interpretability-accuracy trade-off is one of the most challenging tasks on fuzzy modelling. Indeed, interpretability is acknowledged as a distinguishing capability of linguistic fuzzy systems since the proposal of Zadeh and Mamdani's seminal ideas. Anyway, obtaining interpretable fuzzy systems is not straightforward. It becomes a matter of careful design which must cover several abstraction levels. Namely, from the design of each individual linguistic term (and its related fuzzy set) to the analysis of the cooperation among several rules, what depends on the fuzzy inference mechanism. This work gives an overview on existing tools for fuzzy system modelling. Moreover, it introduces GUAJE which is an open-source free-software java environment for building understandable and accurate fuzzy rule-based systems by means of combining several pre-existing tools.

Keywords: interpretability, fuzzy modeling, free open source software.

1 Introduction

The term Soft Computing (SC) is usually defined by its essential properties, as a family of techniques, as a complement of hard computing, and/or as a tool for coping with imprecision and uncertainty [22]. One of the main issues regarding SC techniques is their cooperative nature. Each individual technique (Fuzzy Logic, Neuro-computing, Probabilistic Reasoning, Evolutionary Computation, etc.), even each individual algorithm, has its own advantages and drawbacks. A family of several pre-existing techniques is able to work in a cooperative way yielding hybrid systems, taking profit from the main advantages of each of them, in order to solve lots of complex real-world problems for which other techniques are not well suited.

Since this work deals with modelling understandable and accurate systems, it is mainly focused on those SC techniques more suitable for dealing with the so-called humanistic systems, defined by Zadeh [30] as "*those systems whose behaviour is strongly influenced by human judgment, perception or emotions*". We will concentrate on Fuzzy Logic (FL) [28] because of its semantic expressivity close to natural language is well-known for linguistic concept modelling.

The use of linguistic variables and rules [29] favours the interpretability of fuzzy systems. Unfortunately, FL is not enough for building interpretable systems, i.e., fuzzy systems are not interpretable *per se*. Thus, the whole modelling

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 212–219, 2011.

process must be carried out carefully, paying special attention to interpretability from the beginning to the end and imposing several constraints [25]. Notwithstanding, interpretability requirements strongly depend not only on each specific problem but also on the background (experience, preferences, knowledge, etc.) of the end-user who will interact with the designed system. Notice that, looking for a good interpretability-accuracy trade-off is one of the most complex tasks on system modelling. It demands the aid of powerful software tools.

The objective of this paper is twofold. First, giving an overview on existing software tools for fuzzy system modelling. Second, introducing GUAJE. It is a java environment for building understandable and accurate fuzzy rule-based systems. Notice that, GUAJE is an enhanced version of a previous open source software (KBCT) and it integrates several algorithms provided by different open source software tools. It is worth noting that GUAJE is not merely an aggregation of different tools but there exists the possibility of automatically writing a configuration file to select the different algorithms during the modelling process, no matter the specific tool they are provided by.

The structure of this paper is as follows. The next section makes a global review on available software that implements SC techniques for system modelling. Then, section 3 presents GUAJE. Section 4 enumerates some applications. Finally, section 5 draws some conclusions and points out future works.

2 An Overview on Software for System Modelling

Most software for SC system modelling is available on the Web in the form of libraries and/or small tools which often come from academics and small research groups. In order to get wider visibility and cooperation with other researchers those tools are usually downloadable as free open source software, at least for research and education purposes. As a result, there is a huge amount of available free software what makes really easy creating new small prototypes for lots of applications without the effort of starting from scratch. However, the main drawback of such developments is their maintenance cost. Keeping a flexible and well-documented source code is a mandatory requirement in order to make feasible the cooperation of several researchers in a common development. In addition, the coordination and control of subversions is a really difficult task when several researchers, sometimes located at different parts of the world, are only working on the software development during their own free time. Of course, as alternative it is possible to resort to professional commercial tools like the Matlab toolboxes which include the well-known Fuzzy Toolbox and ANFIS (Adaptive Neuro-Fuzzy System). Nevertheless, we advocate for the use of open source software because it offers the richness of quickly incorporating new developments made by the active research community which is always working in emerging fields.

In short, some of the most famous free-software SC packages and tools are the following. In the field of evolutionary computation, JCLEC[1] (Java Class Library

[1] http://jclec.sourceforge.net/

for Evolutionary Computation) and JMetal[2] (Metaheuristic Algorithms in Java) provide two nice frameworks for both evolutionary and multi-objective optimization. JavaNNS[3] (Java version of Stuttgart Neural Network Simulator) is probably the best free suite for neural networks. Regarding fuzzy modelling, Xfuzzy[4] (a development environment for fuzzy-inference-based systems), FisPro[5] (Fuzzy Inference System Professional) and KBCT[6] (Knowledge Base Configuration Tool) represent three useful tools. In addition, regarding neuro-fuzzy algorithms we can point out, among others, to NEFCLASS[7] (Neuro-Fuzzy Classification).

There are also some interesting and successful attempts for going beyond specialized tools. For instance, FrIDA [12] is free and open source software in the form of a java-based graphical user interface (GUI) that joins several individual tools for data analysis and visualization. In this case all small programs were developed by the same researchers over the years. KEEL (Knowledge Extraction based on Evolutionary Learning) [1] is another more ambitious software tool created as part of a research project with several goals. To start with it includes a huge repository made up of hundreds of evolutionary learning algorithms developed by several authors (belonging to different research groups) as part of their own research works. Furthermore, new algorithms can be easily added. In addition, KEEL offers a user-friendly java GUI for designing experiments where different algorithms can be fairly compared with exactly the same datasets under a complete statistical analysis. Another quite famous tool putting together several machine-learning algorithms under the same interface is Weka[8] (Data Mining Software in Java) [27]. It is also developed following the open source philosophy and it counts with a lot of related projects and contributors. To do so, it applies the Linux model of releases. It focuses on automatic extraction of knowledge from data but, unfortunately, it does not take care of the interpretability of the generated models and it does not include any algorithms for fuzzy modelling. Lastly, KNIME[9] is a user-friendly and comprehensive open-source data integration, processing, analysis, and exploration platform for both industry and academia.

3 Description of the GUAJE Environment

GUAJE[10] stands for Generating Understandable and Accurate fuzzy rule-based systems in a Java Environment. GUAJE implements the Highly Interpretable Linguistic Knowledge (HILK) fuzzy modelling methodology [3,6].

[2] http://jmetal.sourceforge.net/
[3] http://www.ra.cs.uni-tuebingen.de/SNNS/
[4] https://forja.rediris.es/projects/xfuzzy/
[5] http://www.inra.fr/internet/Departements/MIA/M/fispro/
[6] http://www.mat.upm.es/projects/advocate/kbct.htm
[7] http://fuzzy.cs.uni-magdeburg.de/nefclass/
[8] http://www.cs.waikato.ac.nz/ml/weka/
[9] http://knime.org/
[10] http://www.softcomputing.es/guaje

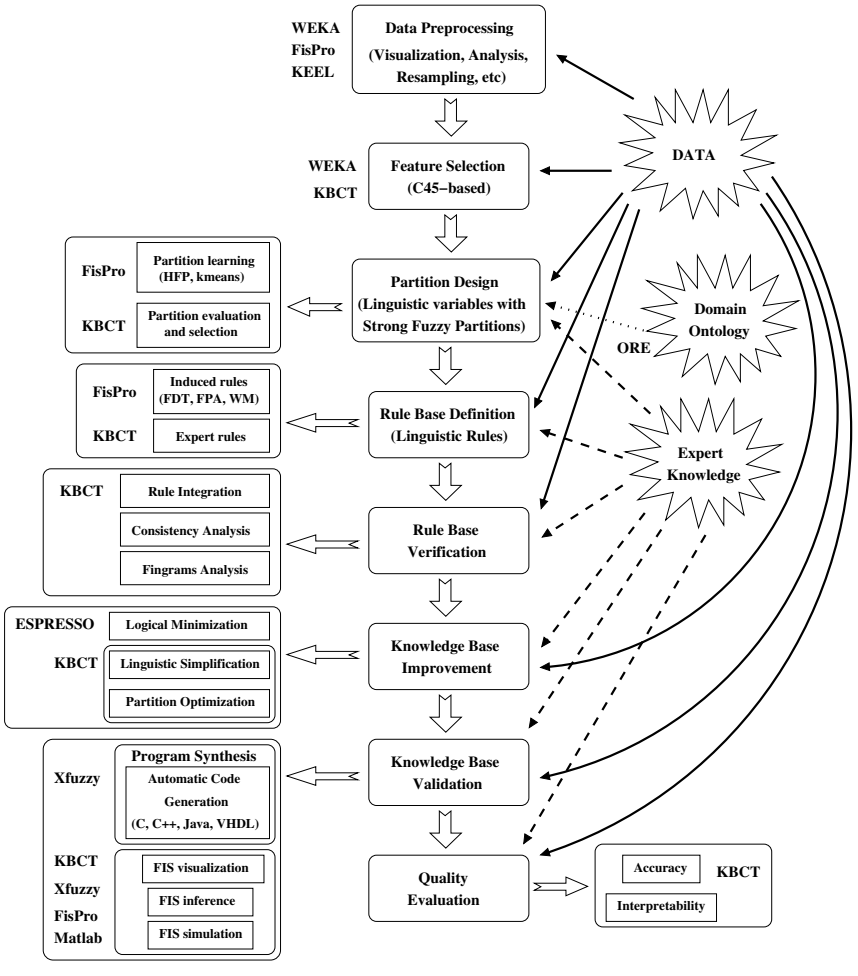


Fig. 1. Scheme of the proposed GUAJE environment

The main building blocks are sketched in Fig. 1. The core of GUAJE is the last downloadable version of KBCT [5] (version 3.0) which has been upgraded with new functionalities. It is an open source software for knowledge extraction and representation which combines expert knowledge and induced knowledge (knowledge automatically extracted from experimental data).

The whole modelling process is made up of the next steps. First of all, available experimental data must be pre-processed and translated into the format handled by GUAJE. Secondly, a feature selection process is needed in the case of dealing with complex problems involving many inputs. Thirdly, the partition design stage is based on the definition of linguistic variables characterized by strong fuzzy partitions. Both expert partitions and partitions automatically generated from experimental data are compared. The best partitions according to data

distribution and expert knowledge are selected for each input variable. Then, two sets of linguistic rules (expert and induced rules) describe the system behaviour by means of combining the previously generated linguistic variables. They are rules of form **If** *Premise* **Then** *Conclusion* where both premises and conclusions are expressed by linguistic propositions. Then, both sets of rules are integrated in a unique one after checking integrity, consistency, and so on. Then, the resultant knowledge base can be improved regarding both interpretability (minimization and simplification) and/or accuracy (optimization). Finally, after validating the final fuzzy system (knowledge base + inference mechanism) it is possible to generate native code with the aim of running the designed system in a stand-alone application.

Following the cooperative spirit of SC applications, GUAJE not only promotes the combination of several SC techniques but also the combination of several available system modelling tools with the aim of building interpretable fuzzy systems:

- **FisPro**. An open source tool for creating fuzzy inference systems (FIS) to be used for reasoning purposes, especially for simulating a physical or biological system [19]. It includes many algorithms (most of them implemented as C programs) for generating fuzzy partitions and rules directly from experimental data. In addition, it offers data and FIS visualization methods with a java-based user-friendly interface. GUAJE makes use of the following algorithms provided by FisPro: K-means [20]; Hierarchical Fuzzy Partitioning (HFP) [18]; Wang and Mendel (WM) [26]; Fast Prototyping Algorithm (FPA) [19]; and Fuzzy Decision Trees (FDT) [21].
- **ORE**[11] (Ontology Rule Editor). A java-based open source platform-independent application for defining, managing and testing inference rules on a model represented by a specific ontology [23]. GUAJE calls to libraries provided by ORE for visualizing domain ontologies with the aim of making easier the process of expert knowledge extraction [8].
- **Espresso**. Free software designed for logical minimization which implements the algorithm developed by R. Brayton [13]. GUAJE calls to Espresso as part of the module in charge of running the interpretability assessment approach based on semantic co-intension proposed by [24].
- **Graphviz**[12]. A collection of free software for viewing and manipulating abstract graphs [16]. It is used by the module of GUAJE responsible for a novel interpretability analysis at fuzzy inference level (fingrams analysis) [2].
- **JMetal**. Free software that comprises a set of Metaheuristic algorithms implemented in Java by Durillo et al. [15]. We have combined GUAJE with JMetal looking for embedding HILK into a multi-objective evolutionary framework [4,14].
- **Weka**. Open source tool for data mining. It includes the implementation of many classical algorithms like for example J48 which corresponds to the

[11] http://sourceforge.net/projects/ore/
[12] http://www.graphviz.org/

well-known C4.5 algorithm. GUAJE offers a feature selection procedure based on such algorithm.

Fuzzy knowledge bases generated with GUAJE can be exported to the format recognized by FisPro, Xfuzzy and Matlab Fuzzy Toolbox. Thus, those tools can be used just at the final modelling stages. Notwithstanding, the inverse translation is not allowed, i.e., knowledge bases modified with FisPro, Matlab, or Xfuzzy can not be imported and opened again with GUAJE. This is a restriction to preserve the interpretability of the final model because FisPro, Matlab, or Xfuzzy may violate the interpretability constraints imposed and satisfied by GUAJE. We recommend the use of Xfuzzy to generate code for stand-alone applications. On the other hand, Matlab may be useful for system simulations. Finally, KEEL offers many learning algorithms that may be incorporated to GUAJE in the near future. Anyway, they can also be used for the first data pre-processing stage.

4 Applications

KBCT, the ancestor of GUAJE, has been successfully used with the aim of designing interpretable fuzzy systems for many applications: detecting the inattentiveness level of a driver [11]; making intelligent diagnosis in robotics [7]; classifying glucose measurements in a telemedicine system [17]; localization of autonomous robots in indoor environments [9]; and human activity recognition [10].

Thanks to its new functionalities (feature selection, visual analysis and simplification by fingrams, interpretability indexes, etc.), GUAJE is expected to overcome results reported by KBCT in previous applications. Moreover, GUAJE is supposed to yield good results even in complex large-size problems where KBCT suffered from scalability problems.

5 Conclusions and Future Works

This paper has presented a new system modelling suite mainly focused on designing FRBSs with a good interpretability-accuracy trade-off by means of combining several pre-existing tools. This approach lets us saving a lot of time because we reuse many algorithms already freely available on the Web as part of other tools which are distributed as open source software. New algorithms can be added in the future with the aim of complementing the existing ones or adding new functionalities.

GUAJE is freely available (under GPL license) as open source software at:

http://www.softcomputing.es/guaje

Acknowledgment. GUAJE is an enhanced version of KBCT whose first version was developed as part of the European research project ADVOCATE II supported by the European Commission (IST-2001-34508). The initial development started from FisPro what explains why both tools are so closely linked.

The present work is partly supported by the European Centre for Soft Computing (ECSC) and the Spanish Ministry of Science and Innovation under project TIN2008-06890-C02-01.

References

1. Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M., Fernández, J.C., Herrera, F.: KEEL: A software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13(3), 307–318 (2009)
2. Alonso, J.M., Cordón, O., Quirin, A., Magdalena, L.: Analyzing interpretability of fuzzy rule-based systems by means of fuzzy inference-grams. In: World Congress on Soft Computing (2011)
3. Alonso, J.M., Magdalena, L.: HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. Soft Computing (2010), doi:10.1007/s00500-010-0628-5
4. Alonso, J.M., Magdalena, L., Cordón, O.: Embedding hilk in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers. In: IV International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS), pp. 15–20 (2010)
5. Alonso, J.M., Magdalena, L., Guillaume, S.: KBCT: A knowledge extraction and representation tool for fuzzy logic based systems. In: IEEE International Conference on Fuzzy Systems, pp. 989–994 (2004)
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(7), 761–794 (2008)
7. Alonso, J.M., Magdalena, L., Guillaume, S., Sotelo, M.A., Bergasa, L.M., Ocaña, M., Flores, R.: Knowledge-based intelligent diagnosis of ground robot collision with non detectable obstacles. Journal of Robotic & Intelligent Systems 48, 539–566 (2007)
8. Alonso, J.M., Muñoz, A., Botía, J.A., Magdalena, L., Gómez-Skarmeta, A.F.: Uso de ontologías para facilitar las tareas de extracción y representación de conocimiento en el diseño de sistemas basados en reglas borrosas. In: XIV Spanish ESTYLF Conference on Fuzzy Logic and Technologies, pp. 233–240 (2008)
9. Alonso, J.M., Ocaña, M., Sotelo, M.A., Bergasa, L.M., Magdalena, L.: WiFi localization system using fuzzy rule-based classification. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2009. LNCS, vol. 5717, pp. 383–390. Springer, Heidelberg (2009)
10. Alvarez, A., Alonso, J.M., Trivino, G., Hernandez, N., Herranz, F., Llamazares, A., Ocaña, M.: Human activity recognition applying computational intelligence techniques for fusing information related to wifi positioning and body posture. In: IEEE World Congress on Computational Intelligence, pp. 295–304 (2010)
11. Bergasa, L.M., Nuevo, J., Sotelo, M.A., Barea, R., López, M.E.: Real-time system for monitoring driver vigilance. IEEE Transactions on Intelligent Transportation Systems 7(1), 63–77 (2006)
12. Borgelt, C., González-Rodríguez, G.: FrIDA - a free intelligent data analysis toolbox. In: IEEE International Conference on Fuzzy Systems, pp. 1892–1896 (2007)
13. Brayton, R.K., Hachtel, G.D., McMullen, C., Sangiovanni-Vincentelli, A.: Logic Minimization Algorithms for VLSI Synthesis. Kluwer Academic Publishers Group, Dordrecht (1984)

14. Cannone, R., Alonso, J.M., Magdalena, L.: Multi-objective design of highly interpretable fuzzy rule-based classifiers with semantic cointension. In: V International Workshop on Genetic and Evolutionary Fuzzy Systems, GEFS (2011)
15. Durillo, J., Nebro, A.J., Alba, E.: The jMetal framework for multi-objective optimization: Design and architecture. In: IEEE World Congress on Computational Intelligence, pp. 4318–4325 (2010)
16. Gansner, E.R., North, S.C.: An open graph visualization system and its applications to software engineering. Software - Practice and Experience 30(11), 1203–1233 (1999)
17. Garcia-Saez, G., Alonso, J.M., Molero, J., Rigla, M., Martinez-Sarriegui, I., de Leiva, A., Gomez, E.J., Hernando, M.E.: Mealtime blood glucose classifier based on fuzzy logic for the diabtel telemedicine system. In: Combi, C., Shahar, Y., Abu-Hanna, A. (eds.) AIME 2009. LNCS, vol. 5651, pp. 295–304. Springer, Heidelberg (2009)
18. Guillaume, S., Charnomordic, B.: Generating an interpretable family of fuzzy partitions. IEEE Transactions on Fuzzy Systems 12(3), 324–335 (2004)
19. Guillaume, S., Charnomordic, B.: Learning interpretable fuzzy inference systems with FisPro. Information Sciences, Special Issue on Interpretable Fuzzy Systems (2011) (In press)
20. Hartigan, J.A., Wong, M.A.: A k-means clustering algorithm. Applied Statistics 28, 100–108 (1979)
21. Ichihashi, H., Shirai, T., Nagasaka, K., Miyoshi, T.: Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. Fuzzy Sets and Systems 81, 157–167 (1996)
22. Magdalena, L.: What is soft computing? Revisiting possible answers. In: 8th International FLINS Conference, 2008, pp. 3–10 (2008)
23. Muñoz, A., Vera, A., Botía, J.A., Gómez-Skarmeta, A.F.: Defining basic behaviours in ambient intelligence environments by means of rule-based programming with visual tools. In: 1st Workshop of Artificial Intelligence Techniques for Ambient Intelligence. ECAI (2006)
24. Mencar, C., Castiello, C., Cannone, R., Fanelli, A.: Interpretability assessment of fuzzy knowledge bases: a cointension based approach. International Journal of Approximate Reasoning 52(4), 501–518 (2011)
25. Mencar, C., Fanelli, A.M.: Interpretability constraints for fuzzy information granulation. Information Sciences 178, 4585–4618 (2008)
26. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man and Cybernetics 22 (6), 1414–1427 (1992)
27. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
28. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
29. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. on SMC 3, 28–44 (1973)
30. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Parts I, II, and III. Information Sciences 8, 8, 9, 199–249, 301–357, 43–80 (1975)

Serendipitous Fuzzy Item Recommendation with ProfileMatcher

Danilo Dell'Agnello, Anna Maria Fanelli,
Corrado Mencar, and Massimo Minervini

Dept. of Informatics, University of Bari, 70125 Bari, Italy
{danilodellagnello,mencar,fanelli}@di.uniba.it,
minervini.massimo@gmail.com

Abstract. In this paper an approach to serendipitous item recommendation is outlined. The model used for this task is an extension of ProfileMatcher, which is based on fuzzy metadata describing both user and items to be recommended. To address the task of recommending serendipitous resources, a priori knowledge on the relations occurring among metadata values is injected in the recommendation process. This is achieved using fuzzy graphs to model similarity relations among the elements of the fuzzy sets describing the metadata. An experimentation has been carried out on the MovieLens data set to show the impact of serendipity injection in the item recommendation process.

1 Introduction

Nowadays users are facing the availability of enormous quantities of information. This makes the selection of interesting information from a user side exceedingly difficult without any searching and filtering capability.

Recommender Systems (RSs) are capable of filtering information according to user interests and preferences [1,6,7]. Their main feature consists in suggesting information that is esteemed as appropriate according to user needs. In this way, users are not compelled to search information in the whole set of data, but only in a small fraction that is easier to manage.

RSs can use user profiles tailored so that recommended information is interesting for the user; a correct implementation of this requirement is strictly dependent on the aims of the information system and on the nature of the stored information.

An approach that uses fuzzy sets for profile representation and graduated recommendation is given by ProfileMatcher [2]. In ProfileMatcher, each profile represents the complex preferences of a user in terms of profile components, which model elementary preferences. Each profile component is defined through a set of attributes and each attribute is quantified through a fuzzy set of values.

ProfileMatcher can be used when information stored in an information system describe external resources, such as buyable items, learning objects, etc. Such type of information is often called "metadata" and is represented in terms of attribute-value couples. ProfileMatcher establishes a degree of compatibility

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 220–227, 2011.

between a resource and a profile by a fuzzy matching of its components with the resource metadata. In this way, resources can be ranked on the basis of the quantified compatibility, which can be directly related to their interestingness for the user. Like other RSs, ProfileMatcher learns user profiles from past interactions.

In some applicative contexts, a "serendipitous" recommendation could be a desirable feature. Serendipity is a propensity for making fortunate discoveries while looking for something unrelated. Serendipity occurs everywhere (even in scientific research), but it is reasonably more frequent in contexts where users are not always looking for precise resources (e.g. books, movies, papers, etc.)

Any RS capable of serendipitous recommendation will certainly give added value in all information systems where serendipitous discovers are frequent. However, the simple recommendation of random resources may not be an effective solution, because the frequency of really uninteresting yet recommended resources will annoy the user, who eventually would no more rely on the recommendation service.

In this paper we introduce an extension of ProfileMatcher for serendipitous recommendation. A fuzzy graph is defined for each attribute used for calculating the degree of compatibility between a resource and a profile. The fuzzy graph establishes a semantic similarity between attribute values, so as to increase the compatibility of resources whose description does not completely match the user profile. As a result of this knowledge injection, ProfileMatcher is capable of recommending resources that are only loosely related to resources already accessed by the user, thus favouring serendipitous discovers.

In the next section ProfileMatcher is briefly outlined and the proposed approach for serendipity injection is described along with an illustrative example. Section 3 is focused on the description of an application of the proposed model for serendipitous recommendation using the MovieLens data set. Section 4 concludes the paper with some discussions on the achieved results.

2 Profile Matcher with Serendipity

In this paper we focus on ProfileMatcher as a tool for recommending resources on the basis of fuzzy metadata. It computes the degree of compatibility between a user u and a resource r as a value

$$K(u, r) \in [0, 1]$$

To compute the compatibility degree, ProfileMatcher uses profiles to describe users. Then, for each user u, a profile $P(u)$ is defined. Profiles can either be designed manually or automatically acquired from data [2]. A user profile describes the (possibly complex) preference structure of a user. This structure is composed of one or more profile components, i.e.

$$P(u) = \{p_1(u), p_2(u), \ldots, p_n(u)\}$$

where each component $p_i(u)$ represents an elementary preference. A resource r is compatible with u if it is compatible with at least one profile component, i.e.

$$K(u,r) = \max_{i=1}^{n} K(p_i(u), r).$$

Both profile components and resources are described through fuzzy metadata. Each metadatum describes a specific attribute of a resource through a fuzzy set. More formally, an attribute is a domain of values $A = \{v_1, v_2, \ldots\}$ and is denoted by an attribute label $\mathtt{A}$, while a metadatum is a couple $\langle \mathtt{A}, \mu_A \rangle$ where $\mu_A : A \rightarrow [0,1]$ is a fuzzy set on the values of the attribute. A resource r is described through a collection of metadata, i.e.

$$D(r) = \left\{ \langle \mathtt{A}_1, \mu_{A_1} \rangle, \langle \mathtt{A}_2, \mu_{A_2} \rangle, \ldots, \langle \mathtt{A}_{m(r)}, \mu_{A_{m(r)}} \rangle \right\}$$

being $m(r)$ the number of attributes used to describe the resource r.

A constraint on the attributes is defined: for each couple $\langle A_i, \mu_{A_i} \rangle$ composing the resource descriptions, the attribute A_i is unique. In other words it is impossible that the same attribute occurs more than once in different couples of a resource description $D(r)$.

Also profile components are expressed in terms of metadata. This enables an easy computation of the compatibility degree and the transparency of the profile description. We denote a profile component as

$$p_i(u) = \left\{ \langle \mathtt{B}_1, \mu_{B_1} \rangle, \langle \mathtt{B}_2, \mu_{B_2} \rangle, \ldots, \langle \mathtt{B}_{m(i)}, \mu_{B_{m(i)}} \rangle \right\}$$

being $m(i)$ the number of attributes used to describe the profile component p_i. On attributes B_j is defined the same constraint given for attributes in $D(r)$: it is impossible that the same attribute occurs more than once in different couples of a profile component $p_i(u)$.

With such a representation, the compatibility degree between a profile component and a resource is computed in terms of metadata sharing the same attribute label as follows:

$$K(p_i(u), r) = avg_{A_j = B_k} K\left(\mu_{A_j}, \mu_{B_k}\right)$$

Averaging is more suitable for preference computation since it is less stringent than t-norms. We use standard arithmentic mean for averaging compatibility values, however weighted average could also be used if some attributes are more relevant than others.

The last point required to complete the definition of compatibility degree concerns the compatibility between two fuzzy sets defined on the same attribute. In the current version of ProfileMatcher, this compatibility is computed in terms of possibility measure between fuzzy sets. This definition is a consequence of the assumption that two fuzzy sets are compatible if they share at least one element, i.e.

$$K\left(\mu_{A_j}, \mu_{B_k}\right) = \max_{v \in A_j} \min\left\{\mu_{A_j}(v), \mu_{B_k}(v)\right\} \quad (1)$$

Eq. (1) is relevant in this work because it is at the basis of serendipity injection in ProfileMatcher.

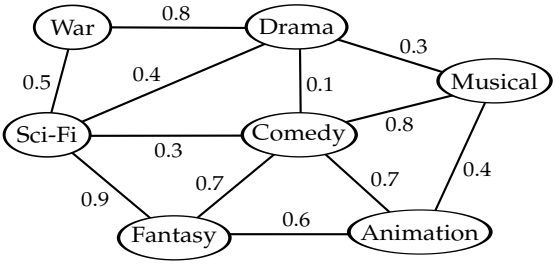


Fig. 1. An example of fuzzy graph for an attribute describing the genre of a movie

2.1 Serendipity Injection

A good serendipity module in a recommender system should be able to recommend unexpected yet potentially interesting resources exploiting some loose relationships between known preferred resources and new resources.

The task of providing non-obvious recommendations has been addressed by various authors in the recent years ([3,4,5]). In this work we use fuzzy graphs to define such relationships. A fuzzy graph is a pair $\langle A, E\rangle$ where A is an attribute and

$$E : A \times A \rightarrow [0, 1]$$

defines a fuzzy similarity relationship between attribute values. Given a pair of attribute values a', a'', $E(a', a'') = 1$ means that values a', a'' are equivalent (or indistinguishable), $E(a', a'') = 0$ refers to completely distinct attributes, while any intermediate value of $E(a', a'')$ represents partial similarity. It is assumed that $E(a, a) = 1$ for each $a \in A$.

Once a fuzzy graph is defined for an attribute, it can be used to inject serendipity in ProfileMatcher by extending eq. (1) with a formalization of the following property: two fuzzy sets are compatible if there exist two similar values, one belonging to the first fuzzy set and the other belonging to the second one. Formally, this translates into:

$$K'\left(\mu_{A_j}, \mu_{B_k}\right) = \max_{v', v'' \in A_j} \min\left\{\mu_{A_j}(v'), \mu_{B_k}(v''), E(v', v'')\right\} \tag{2}$$

By using eq. (2) in place of eq. (1) in ProfileMatcher, it is possible to recommend a resource even if the latter is described by metadata that do not share values with any profile component, provided that a similarity relationship exists between the values of the resource metadata and those of a profile component.

The definition of a fuzzy graph on an attribute is essential for serendipitous recommendations. It can be defined manually if the cardinality of an attribute domain is limited or if there exists a priori knowledge on the similarity of attribute values. Otherwise, it could be more convenient to derive the similarity degrees automatically from available data.

2.2 Illustrative Example

To provide a better explanation of the proposed methodology, the following illustrative example is considered, concerning a user interacting with a collection of movies, each of them described by its genre attribute.

A user and three resources described with the formalism of ProfileMatcher are considered:

$$P(u) = \{\{\langle \texttt{Genre}, \{1/\texttt{War}, 0.6/\texttt{Drama}, 0.1/\texttt{Comedy}\}\rangle\}\}$$
$$D(r_1) = \{\langle \texttt{Genre}, \{0.4/\texttt{War}, 1/\texttt{Drama}, 0.6/\texttt{Comedy}\}\rangle\}$$
$$D(r_2) = \{\langle \texttt{Genre}, \{0.7/\texttt{War}, 0.2/\texttt{Drama}, 1/\texttt{Comedy}\}\rangle\}$$
$$D(r_3) = \{\langle \texttt{Genre}, \{1/\texttt{Sci}-\texttt{fi}, 0.5/\texttt{Animation}\}\rangle\}$$

This models the fact that the user u, described by means of a single profile component, is highly interested in war movies, partially in dramatic movies and only a little in comedy movies. The genre of the considered resources is mainly drama (r_1), comedy (r_2) and science fiction (r_3).

If there is no injection of serendipity in the model, the compatibility values obtained applying (1) are $K(u, r_1) = 0.6$, $K(u, r_2) = 0.7$ and $K(u, r_3) = 0$, due to the high membership degrees that both the profile and the component description have for the common element "drama" in the first case and "war" in the second case, while in the third case there are no common elements.

If the knowledge described in the fuzzy graph shown in fig. 1 is injected in the model,[1] by means of (2), there is now a different compatibility estimation, that is $K(u, r_1) = 0.8$, $K(u, r_2) = 0.7$ and $K(u, r_3) = 0.5$.

The use of the fuzzy similarity relation allows to obtain a matching at a semantic level that the previous model was unable to detect, thus leading to serendipitous recommendations. The order of the recommended resources changes, since r_2 appears to be more compatible than r_1. Besides r_3 has an increased compatibility value because of the similarity between science fiction and war. The latter result is extremely important since it shows that a resource description, not sharing any value for the attribute "genre" with the user profile, is evaluated as compatible. Thus, the proposed extension allows the user to make unexpected discoveries on new movies genres.

3 Experimental Results

3.1 The Dataset

To evaluate the effectiveness of the proposed approach the MovieLens data set has been used. It stores the ratings of 943 users for 1682 movies, both of them

[1] The network models the fact that reasonably a user interested in "war" movies possibly is also highly interested in movies whose genre is "drama" and less in "comedy" movies.

identified with a numerical ID. The data set also stores the genre, the release date, the title of the movie and its IMDB[2] url. To describe the user some basic information are provided, such as the age, the gender, the occupation and the ZIP code. In this contribution we are interested only in the "genre" attribute of the resources, which is described by a binary array of 19 elements, each of them having a 1 value if the movie belongs to that genre and a 0 value otherwise.

3.2 Definition of the Fuzzy Graph

In our experimentation, we used only the attribute "genre" since we were interested in analyzing the effects of serendipity injection in a simplified setting. A more complex experimental context would have made the understanding of the effects of the proposed extension more difficult.

The fuzzy graph used for defining the similarity between genres has been acquired from the available dataset, in two steps. In the first step, we represented each genre as a crisp set defined on the domain of movies, i.e. a genre is represented by all the movies of that genre. Specifically, given a movie m in the set M of movies, the genre set of m is defined as $G(m) = \{g_1, g_2, \ldots\} \subseteq G$, being G the set of all genres. In a dual viewpoint, given a genre $g \in G$, we considered all the movies of that genre as $M(g) = \{m \in M : g \in G(m)\}$. We assumed that if two genres share the same movies, and the set of movies is considerably high, then they are similar. Thus we used the Jaccard index to evaluate the similarity between genres:

$$S(g', g'') = \frac{|M(g') \cap M(g'')|}{|M(g') \cup M(g'')|}$$

The values of the Jaccard index could be very small in a real experimental setting. However we did not desire the serendipity injection to be too weak, since we would not appreciate its effect. As a consequence, in the second step we used a linguistic hedge to define the fuzzy graph, by defining the edge set E as the fuzzy set of the genres "somewhat similar". This led to the definition of the edge set as

$$E(g', g'') = S(g', g'')^{\frac{1}{3}}$$

whose values suit better to the application, even preserving their own semantics.

3.3 Experimental Setting

The experimentation was aimed to show how the use of the fuzzy graph influences the rating of the resources to be recommended.

A set of 30 user profiles has been learnt using $N = 9$ resources. Subsequently the sequence $TS_H(u)$ of the following $H = 10$ resources accessed after the learning phase has been considered, in order to evaluate the ability of the model to provide good recommendations in the short term. The average $\overline{K}^{(u)}_{mode} = mean_{m \in TS_H(u)} K(u, m)$ is an indicator of this goodness and *mode* is

[2] Internet Movies Data Base, available at http://www.imdb.com

a placeholder indicating the mode used to calculate the compatibility by ProfileMatcher. Specifically, $mode = NoSN$ refers to the previous version of ProfileMatcher, i.e. without serendipity injection, and $mode = SN$ refers to the proposed approach, i.e. with serendipity injection.

3.4 Experimental Results

The results obtained for the first phase described in the previous section are shown in fig. 2.

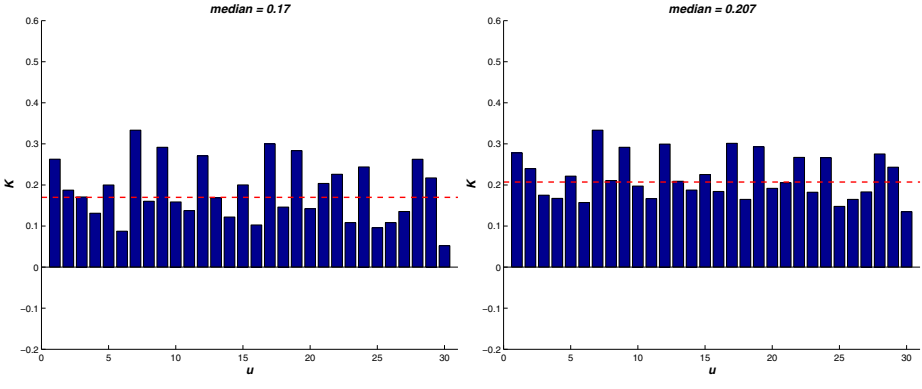


Fig. 2. Compatibility with hidden resources, with and without any serendipity injection

According to the expectations, the newly available knowledge conveyed by the fuzzy graph causes a slight increase in the median value of $\overline{K}^{(u)}_{SN}$ w.r.t. $\overline{K}^{(u)}_{NoSN}$. This can be explained by considering that in the previous stage only the syntactical matching was considered, whilst by using the fuzzy graph a non-zero compatibility value can be now obtained between elements not matching at a syntactical level. At a higher level this can be explained by the larger amount of knowledge available to the system, leading to an increase in the average compatibility between users and resources.

4 Conclusions

In this paper we have outlined an approach to serendipitous item recommendation. This has been achieved using fuzzy graphs to model the fuzzy similarity relation among the elements in fuzzy sets describing metadata in ProfileMatcher. The main motivation was to achieve unexpected recommendations in order to give to ProfileMatcher the ability of recommending non-obvious yet interesting resources. A strategy to derive such a fuzzy similarity relation from data has also been described.

The use of the fuzzy graphs has proved to be helpful, since it provides the model with the ability to match elements not only syntactically. This leads to the recommendation of resources even though their description does not share elements with the user profile, as depicted in Section 2.2, thus introducing serendipity in the recommendation process.

The experimental results are promising, since there is an increase of the average compatibility value of the resources with users and this is mainly due to the larger amount of avalaible knowledge. Nonetheless a deeper investigation on experimental results will be needed. In fact one of the major issues is the difficulty in the off-line evaluation of the serendipity in RSs. This is mainly due to the lack of any metric that allows to quantify how serendipitous a recommendation is. It is reasonable to suppose that an on-line evaluation, involving a user feedback on the goodness of the recommendation, will fix this issue and this problem will be addressed as a future work. In fact a direct investigation on the user satisfaction about a serendipitous recommendation would allow a direct evaluation of the goodness of the serendipity module, without the need to create any ad hoc metric.

Another further step will be the extension of the present study to a case involving all the attributes in the data set, to better evaluate the influence of the serendipity injection for the item recommendation in a real context.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17, 734–749 (2005)
2. Castellano, G., Castiello, C., Dell'Agnello, D., Fanelli, A.M., Mencar, C., Torsello, M.A.: Learning fuzzy user profiles for resource recommendation. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 18(4), 389–410 (2010)
3. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys 2010, pp. 257–260. ACM, New York (2010)
4. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. 22, 5–53 (2004)
5. Onuma, K., Tong, H., Faloutsos, C.: Tangent: a novel, 'surprise me', recommendation algorithm. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, pp. 657–666. ACM, New York (2009)
6. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007), http://portal.acm.org/citation.cfm?id=1768197.1768209
7. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007), http://portal.acm.org/citation.cfm?id=1768197.1768208

Fuzzy Models for Fingerprint Description

Rosario Arjona[1], Andrés Gersnoviez[2], and Iluminada Baturone[1]

[1]Dept. Electrónica y Electromagnetismo (University of Seville), and Microelectronics Institute of Seville (IMSE-CNM-CSIC), Isla de la Cartuja, 41092 Seville, Spain
{arjona,lumi}@imse-cnm.csic.es
[2]Dept. de Arquitectura de Computadores, Electrónica y Tecnología Electrónica (University of Cordoba), Cordoba, Spain
andresgm@uco.es

Abstract. Fuzzy models, traditionally used in the control field to model controllers or plants behavior, are used in this work to describe fingerprint images. The textures, in this case the directions of the fingerprint ridges, are described for the whole image by fuzzy if-then rules whose antecedents consider a part of the image and the consequent is the associated dominant texture. This low-level fuzzy model allows extracting higher-level information about the fingerprint, such as the existence of singular points and their fuzzy position within the image. This is exploited in two applications: to provide comprehensive information for users of unattended automatic recognition systems and to extract linguistic patterns to classify fingerprints.

Keywords: Fuzzy modeling, automatic fingerprint identification systems, fingerprint classification, linguistic interpretability.

1 Introduction

A fingerprint image (Figure 1(a)) is a grayscale representation based on the reproduction of *ridges* (dark) and *valleys* (bright). *Ridges* and *valleys* are structural features originated by the exterior appearance of the epidermis [2]. Since fingerprints of a person are unique (even for twins), they have been widely used for recognition purposes. The first stage of an automatic fingerprint recognition system is to acquire a digital image from the finger through a sensor. Then, the image is processed to extract several features. Those features are stored as the template of that fingerprint in the enrollment stage, or are compared (matched) with the template in the recognition stage. The features extracted from the fingerprint can be global (if they provide information of the whole fingerprint) or local. An example of global features is the *directional image* (also called orientation image, field, or map, or directional field or map). It gives the local directions of ridges for each pixel or each block of pixels. Figure 1(b) shows the block directional image corresponding to the fingerprint in Figure 1(a). Examples of local features are minutiae. While local features are employed for final identification or authentication applications, global features are used as an aid at different stages such as enhancement and segmentation of fingerprint images, singular points

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 228–235, 2011.

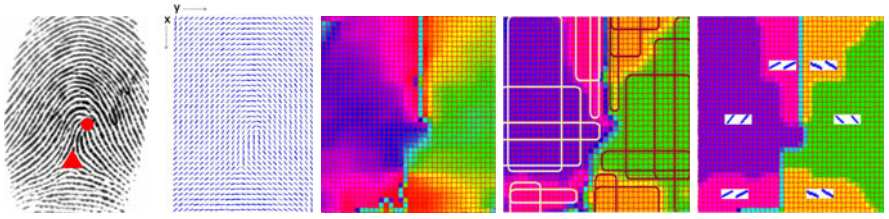


Fig. 1. (a) Fingerprint from FVC database [1] with singular points depicted (*core* as a circle and *delta* as a triangle). (b) Directional image associated. Outputs provided by *xfplot* for fuzzy models with: (c) 625 rules, (d) 61 rules, and (e) 16 rules.

extraction (*core* and *delta*, as shown in Figure 1(a)), alignment, matching, and fingerprint classification.

Fuzzy logic has been used for different purposes in fingerprint recognition systems such as fingerprint image enhancement [3], minutiae extraction and verification [4]-[5], matching [6]-[7], and classification [8]. The capability of fuzzy logic to cope with vagueness and uncertainty is typically exploited in these applications, mainly considering local features. This paper explores two other characteristics of fuzzy logic: capability of modeling and linguistic interpretability. These characteristics are employed to describe fingerprints as a whole. Hence, global features such as directional images are considered.

The paper is structured as follows. Section 2 introduces how to describe a fingerprint image by a fuzzy model. This process can be automated by using *Matlab* and the CAD tools of the design environment *Xfuzzy 3* [9]. The model can be coarser or finer depending on the application. Section 3 illustrates the use of a coarse model to provide linguistically comprehensive feedback to the user of the fingerprint-based recognition system. Section 4 shows the applicability of fuzzy modeling to facilitate fingerprint classification. Finally, Section 5 gives some conclusions and summarizes future work.

2 Fuzzy Modeling

Since fuzzy systems are universal approximators, they have been used to model the behavior of systems from a set of representative numerical data that contain outputs associated to inputs. Fuzzy modeling has been typically applied to model controllers or plants to control, whose inputs and outputs take numerical values [10]. In this sense, the first step to model an image is to select which are the inputs and the outputs considered by the model. Recently, there is a wide research on applying fuzzy logic to images, because images contain imprecise and ambiguous information [11]. In particular, fuzzy models of images have been done considering low-level information, such as textures or grey or color values [12]. This is the approach taken herein: the directional image will provide the texture information about the image to be modeled. The fuzzy models pursued

will have two inputs: the horizontal and vertical coordinates of the pixel, and one output: the local direction of ridges at that pixel. The advantage of the fuzzy model will be its capability to approximate the pixel directional image with a low number of linguistically interpretable 'if-then' rules, that is, its capability of summarizing a large number of numerical data (directions of many pixels) into a small number of rules.

The design flow to obtain these fuzzy models has been automated by using several CAD tools. Firstly, an algorithm developed by [13] in *Matlab* creates gradient-based directional images from fingerprints. The pixel directional image has as many directions as pixels in the fingerprint image. For example, the fingerprint in Figure 1(a) has 374x276 pixels, which means a complete directional image with 374x276 directions (Figure 1(b) shows directional image sampled in blocks of 9x9 directions). All or a subset of these directions and their horizontal and vertical positions are analyzed by the tool *xfdm* of *Xfuzzy 3* to extract fuzzy rules with two inputs (x and y positions) and one output (*direction* between 0° and 180°). This tool allows users to choose a grid- or clustering-based algorithm to extract fuzzy rules, the number and types of membership functions to cover inputs, and the conjunction and defuzzification operators used in the generated rule base. Since our objective are rule bases with linguistic interpretability, the grid-based *Wang-Mendel* algorithm has been selected with *product* as conjunction operator, and *Max Label* for defuzzification (which selects the consequent of the most activated rule as the output of the system) [12]. Initially, the model is finer, for example using 25 *Gaussian* membership functions for each input. In this case, the generated system is composed by 625 rules (25x25 combinations of antecedents), each one with its corresponding consequent. Since the membership functions cover the inputs uniformly, a group of 15x11 pixels is approximately described by a rule, which is something similar to compact the pixel directional image (with 374x276 directions) by a block directional image (with 625 blocks of 15x11 directions). The resulting model is shown in Figure 1(c) by the tool *xfplot* of *Xfuzzy 3*, which displays the output of the system (*direction*) depending on x and y positions. Although the model with 25x25 rules reduces information by more than two orders of magnitude, the number of rules can be further reduced if non relevant information is eliminated. This can be done if the values of the directions (the 625 consequents of the rules) are described by a given set of representative directions (let us say 3, 4, etc., depending on the coarseness desired for the model: finer models as more prototypes selected). A way to find prototypes among a set of values is to apply clustering algorithms on those values. The tool *xfsp* of *Xfuzzy 3* allows applying *Hard C-means* clustering on the consequents, with the number of clusters chosen by the user. Fuzzy models that use such clustering produce a segmented directional image, for example, the 4 clusters of directions employed in the models shown in Figures 1(d) and 1(e) correspond to the following fuzzy sets: *almost horizontal if rotating clockwise*, *almost horizontal if rotating counter-clockwise*, *almost vertical if rotating clockwise*, and *almost vertical if rotating counter-clockwise*.

The rules whose consequents are the same and antecedents are represented by neighbor membership functions can be merged into a rule covering more

situations. This merging can be also automated by the tool *xfsp* [14]. For example, when consequents are clustered into 4 groups, the 625 rules (Figure 1(c)) can be merged into 61 rules (Figure 1(d)). If the model is intended to be linguistically interpretable by a human, the number of rules should be small. A way to achieve this objective is to take the most significant rules for each consequent prototype (or exclude the least significant ones) as proposed in [12]. The idea is to select those rules obtained from merging the highest number of atomic rules. Such simplification allows reducing also the number of fuzzy sets employed to describe the x and y positions. For example, considering the model with 61 rules (Figure 1(d)), a new model with 16 rules could be achieved (Figure 1(e)). The 16 rectangles depicted in Figure 1(d) show the portions of the image that are merged to be described by 16 rules. The model with 16 rules contains an incomplete rule base, because there are combinations of input fuzzy sets that are not considered by any rule. Anyway, the pixels placed at those positions will have the directions assigned by the closest defined rule with higher activation degree, as a result of using fuzzy sets that describe the pixel positions. The resulting model illustrated in Figure 1(e) provides a simple and intuitive linguistic description because the fuzzy sets considered for antecedents and consequents are only a few and have clear linguistic meaning. An example of a rule in this model is: `IF the pixels are placed 'at the bottom' AND 'on the right' THEN direction is 'almost horizontal if rotating counter-clockwise'`.

Finally, another issue to be considered when modeling is the approximation achieved. The different models provide the following *Root Mean Square Errors* (RMSE): 0.060 (625 rules), 0.067 (61 rules), and 0.112 (16 rules).

3 Fuzzy Feedback to Users of Recognition Systems

Many fails of automatic fingerprint-based recognition systems are due to a poor acquisition of user fingerprints. For example, the user may apply too few or too high pressure with the finger on the sensor, or the finger may not be correctly placed for the sensor to capture a good image. This issue is already considered by many sensors that inform the user with a red led if acquisition goes bad or a green led if everything goes well. Such feedback is too crisp and does not guide the user to improve acquisition. Fuzzy feedback with more linguistically interpretable information would be very useful for fully automatic unattended fingerprint-based recognition systems. If fingerprint acquisition is improved, complexity of the automatic recognition system can be reduced drastically. For example, fingerprint classification can be done with a simple fuzzy inference system, as described in the following section.

Our proposal for fuzzy feedback is based on detection or not of singular points (*convex core*, *concave core* and *delta*) and their location within the fingerprint acquired. Depending on this, users have to move horizontally or vertically their fingers. The fuzzy models extracted as commented above have been employed for this purpose. The number of clusters for the rules' consequents and the partition of antecedents are fixed so as to obtain models which allow detecting fuzzy

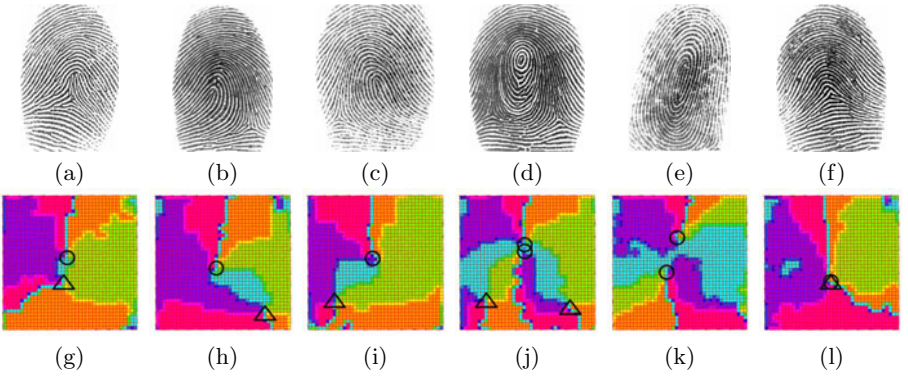


Fig. 2. (a) *Tended arch*, (b) *left loop*, (c) *right loop*, (d) and (e) *whorl*, and (f) *arch*. (g) to (l) Models based on 5 clusters for the fingerprints in (a) to (f) with fuzzy singular points detected: (g)-(i) one *convex core* and one *delta*, (j) one *convex-concave core* and two *deltas*, (k) one *convex core* and one *concave core*, and (l) one *core-delta*

singular points in different groups of fingerprints. Fuzzy singular points are the points in these models where the regions of clustered directions intersect with certain order. They may be *convex core*, *concave core*, *delta*, *convex-concave core* or *core-delta*. Let us consider the fingerprint classification proposed by Henry [2] that discriminates five fingerprints: *tended arch*, *left loop*, *right loop*, *whorl* and *arch*. Figures 2(a)-(f) display examples of fingerprints from FVC database [1] and Figures 2(g)-(l) are fuzzy models with 5 clusters and a 25x25 partition (this selection offers a good trade-off between coarseness of the models and simplicity of the results). The corresponding fuzzy singular points in the models can be viewed in Figures 2(g)-(l). The fuzzy singular points *convex-concave core* (in Figure 2(j)) and *core-delta* (in Figure 2(l)) are the result of fuzzy merging a *convex core* and a *concave core*, and a *convex core* and a *delta*, respectively. Detection of these fuzzy singular points is carried out by analyzing the direction values (consequents) of neighbor rules of the model, as shown in Figure 3. Since fingerprints can be acquired with a certain inclination, the detection of fuzzy singular points considers possible rotations (to a certain extent so as to distinguish, for example, between *convex* and *concave core*).

Location of these fuzzy singular points within the image gives information to know how acquired fingerprint is. Let us consider, for example, the fingerprints depicted in Figures 4(a)-(f) and their fuzzy models in Figures 4(g)-(l). These fingerprints are not well captured because fuzzy singular points are too shifted or they do not appear. Fuzzy feedback consists in communicating to the user what type of movement should be applied to introduce the finger in a correct way. The detection of fuzzy singular points is followed by a fuzzy inference engine based on rules that generates a message for the user. This fuzzy inference engine employs the following fuzzy sets for the positions: *very much on the left* and *very much on the right* for the horizontal position within the image (Figure 5(a)), and *very much at the bottom* and *very much at the top* for the vertical position

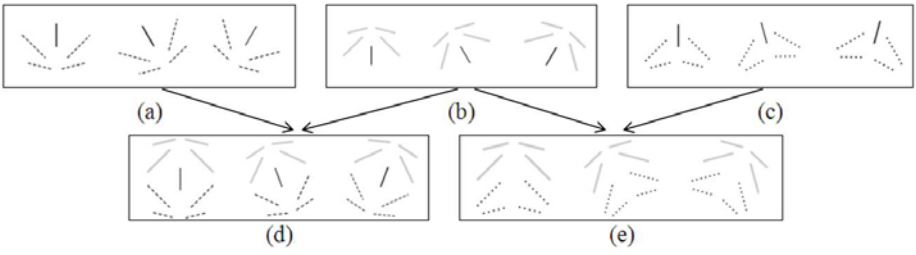


Fig. 3. Patterns for detection of fuzzy singular points based on directions: (a) *concave core*, (b) *convex core*, (c) *delta*, (d) *convex-concave core*, and (e) *core-delta*

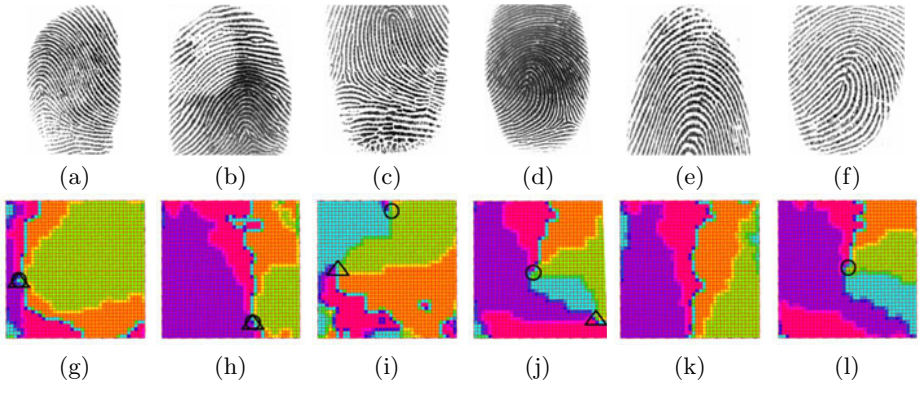


Fig. 4. Fingerprints captured with faults: (a) *core-delta* is *very much on the left*, (b) *core-delta* is *very much at the bottom*, (c) *convex core* is *very much at the top*, (d) *delta* is *very much on the right*, (e) there is not any fuzzy singular point, and (f) there is not *delta*. (g) - (l) Fuzzy models associated with 5 clusters

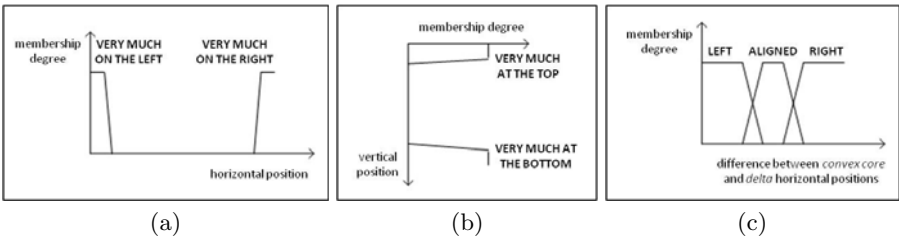


Fig. 5. Fuzzy sets employed in: (a)-(b) feedback, and (c) classification inference engines

(Figure 5(b)). The rules included in this fuzzy inference engine apply that 'if any of the fuzzy singular points is placed in one or two of these fuzzy positions, the user is required to move the finger in the opposite direction'. For example, in Figure 4(g) the *core-delta* is placed *very much on the left*, hence the system requires that the user moves the finger *on the right*. In Figure 4(e) there is not

any fuzzy singular point, therefore the fuzzy inference engine requires 'the user puts the finger on the sensor again'. In any other case, the engine decides that 'the finger is well placed'.

4 Fuzzy Models for Fingerprint Classification

Fuzzy models can be also exploited for classification purposes. Automated fingerprint identification systems have to match a capture with all the fingerprints stored in a database. A large number of comparisons are necessary. To accelerate this process, fingerprint classification gives information to split databases and search only in the selected group of fingerprints. Classification of fingerprints has been done using several techniques such as crisp (non-fuzzy) rule-based, syntactic, structural, statistical, neural network, and combined techniques [2]. The technique proposed herein is to employ fuzzy inference using fuzzy singular points (as commented in the previous sections) to distinguish groups of fingerprints by applying again a subsequent fuzzy inference engine. Each group is characterized by the types, locations and number of fuzzy singular points. For example, in Figures 2(a)-(c) and 2(g)-(i), *tended arch*, *left loop* and *right loop* have one *convex core* and one *delta*. However, they are distinguished by *delta* position with respect to *convex core*: *aligned*, *on the right* and *on the left*, respectively. The *whorl* group has two *deltas*, one *convex core* and one *concave core* (Figures 2(d) and 2(j)). The detection of such *cores* is enough to classify a fingerprint as a *whorl* (Figures 2(e) and 2(k)). Finally, *arch* case only needs to detect the *core-delta* point. Rule base for this fingerprint classification system is shown in Table 1. Figure 5(c) illustrates the fuzzy sets for describing the position of *delta* with respect to *convex core* in *tended arch*, *left* or *right loop* fingerprints. If the fingerprint is acquired correctly (following the feedback commented in the previous section) the percentages of right classification can be 100%, as has been proven with fingerprints from FVC 2002 [1].

Table 1. Rule base for the fingerprint classification system

IF	**Fuzzy singular points**	**Delta position with respect to convex**	**THEN**
1)	(*core-delta*)	-	*arch*
2)	(*convex* AND *concave*) OR (*convex-concave*)	-	*whorl*
3)	(*convex* AND *delta*)	*aligned*	*tended arch*
4)	(*convex* AND *delta*)	*right*	*left loop*
5)	(*convex* AND *delta*)	*left*	*right loop*

5 Conclusions

Coarse fuzzy models of fingerprints can be analyzed by subsequent fuzzy inference engines to provide useful feedback information for users of automatic recognition systems and to understand classification results. A future work is to analyze if finer models could be exploited for identification or authentication purposes.

Acknowledgments. This work was partially supported by Andalusian Consejería de Economía, Innovación y Ciencia under the Project P08-TIC-03674 (with support from the PO FEDER-FSE de Andalucía 2007-2013), and by Spanish Ministerio de Ciencia e Innovación under the Project TEC2008-04920.

References

1. FVC database, http://bias.csr.unibo.it/fvc2002/
2. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition, 2nd edn. Springer, Heidelberg (2009)
3. Bansal, R., Arora, P., Gaur, M., Sehgal, P., Bedi, P.: Fingerprint Image Enhancement using Type-2 Fuzzy Sets. In: Proc. Sixth Int. Conf. on Fuzzy Systems and Knowledge Discovery, Tianji, pp. 412–417 (2009)
4. Sagar, V.K., Alex, K.J.B.: Hybrid Fuzzy Logic and Neural Network Model for Fingerprint Minutiae Extraction. In: International Joint Conference on Neural Networks, Washington, pp. 3255–3259 (1999)
5. Vert, G., Nasser, S.: Minutia Verification using Fuzzy Logic and Geographic Points of Reference. In: The 14th IEEE Int. Conf. on Fuzzy Systems, Reno, pp. 839–844 (2005)
6. Yuan, L.: Application of Fuzzy Pattern Recognition on Fingerprint Processing. In: Second International Conference on Intelligent Computation Technology and Automation, Changsha, pp. 660–663 (2009)
7. Chen, X., Tian, J., Yang, X.: A New Algorithm for Distorted Fingerprints Matching Based on Normalized Fuzzy Similarity Measure. IEEE Trans. on Image Processing, 767–776 (2006)
8. Conti, V., Perconti, D., Romano, S., Tona, G., Vitabile, S., Ganglio, S., Sorbello, F.: Fast Fingerprint Classification Only Using the Directional Image. Knowledge-Based Intelligent Information and Engineering Systems. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part I. LNCS (LNAI), vol. 4692, pp. 34–41. Springer, Heidelberg (2007)
9. Xfuzzy Web page, http://www.imse-cnm.csic.es/Xfuzzy/
10. Kosko, B.: Fuzzy Systems as Universal Approximators. IEEE Transactions on Computers 43(11), 1329–1333 (1994)
11. Nachtegael, M., Van der Weken, D., Kerre, E.E.: Soft Computing Image Processing: Recent Advances. Series Studies in Fuzziness and Soft Computing, vol. 210. Springer, Heidelberg (2007)
12. Baturone, I., Gersnoviez, A.: Automatic Extraction of Linguistic Models for Image Description. In: IEEE International Conference on Fuzzy Systems. IEEE Computational Intelligence Society, Barcelona (2010)
13. Code for directional image, http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html
14. Baturone, I., Moreno-Velo, F.J., Gersnoviez, A.: A CAD Approach to Simplify Fuzzy System Descriptions. In: International Conference on Fuzzy Systems, Vancouver, pp. 2392–2399 (2006)

A Fuzzy Set Approach for Shape-Based Image Annotation

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 shape labeling approach for automatic image annotation. A fuzzy clustering process is applied to shapes represented by Fourier descriptors in order to derive a set of shape prototypes. Then, prototypes are manually annotated by textual labels corresponding to semantic categories. Based on the labeled prototypes, a new shape is automatically labeled by associating a fuzzy set that provides membership degrees of the shape to all semantic classes. Preliminary results show the suitability of the proposed approach to image annotation by encouraging its application in wider application contexts.

Keywords: Fuzzy image annotation, semantic annotation, shape clustering, shape labeling, shape representation.

1 Introduction

The rapid diffusion of large-scale digital image collections has brought about the need for efficient Content-Based Image Retrieval (CBIR) systems that enable users to find relevant images by exploiting automatically-derived low-level features related to the visual content of images, such as color, texture or shapes [11], [8]. The extraction of low-level features from images is a quite easy and fairly direct task. However, people mainly recognize images and express their content on the basis of high-level concepts. To formulate search queries, users typically employ concepts strictly related to the semantic content of images. This has led to the well-known *"semantic gap"* problem in CBIR that refers to the lack of concordance that often occurs between low-level information automatically extracted from images and high-level information [16]. One of the main remedies proposed in literature to bridge the semantic gap is represented by automatic image annotation that focuses on automatically assigning metadata in the form of textual labels to digital images that can be useful in retrieval systems in order to efficiently organize and retrieve images in large collections. Thus, a general annotation process requires the extraction of low-level features from images and the association of such features to text labels.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 236–243, 2011.

Among low-level features commonly employed to index images, shape is one of the most attractive due to its strict link to human perception. Users tend to perceive images as composed of individual objects identified by their shapes and they typically search for images based on shapes of objects contained in images. In this case, image annotation requires the association of labels to shapes.

Many approaches have been proposed in literature for shape annotation [5], [15]. For most of them the process of attaching labels to shapes is crisp, i.e. a shape is exactly classified into one semantic class [13], [18]. However, due to the presence of noise and ambiguity in shape features, it is very difficult or even impossible to classify shapes into precisely one class among the large variety of classes. Labeling shapes with multiple labels rather than with a single one can help to capture the uncertainty underlying the process of annotation. This leads to a fuzzy annotation process that assigns multiple labels to a shape together with values representing the membership degree of shapes to each semantic class [7], [6], [2].

In this paper, we propose an approach to image annotation based on a fuzzy labeling of shapes. Unlike other fuzzy labeling approaches that assume the existence of a number of semantic classes for a given image domain, in our work we automatically define the classes by means of a fuzzy clustering process that allows to derive a set of shape prototypes representative of several semantic categories. Specifically, the approach first creates a database of unlabeled object shapes represented by Fourier descriptors. Then, the well-known FCM clustering is applied to group unlabeled objects according to shape similarity. For each cluster a prototypical shape is derived. Finally, each shape prototype is manually associated to a textual label corresponding to a semantic class. To label a new shape, its visual descriptors are matched with visual descriptors of all prototypes and the similarity values are used to create a fuzzy set expressing the membership degrees of the shape to each semantic class.

2 The Proposed Approach

Three main tasks are carried out in our approach to annotate shapes, namely representation, clustering and labeling. We suppose that a collection of shapes expressed in the form of coordinates of boundaries is available. As a first step, all the shape boundaries are represented by Fourier descriptors that, among the different approaches proposed for describing shape information, are well-recognized nowadays to provide robustness and invariance, obtaining good effectiveness in shape-based indexing and retrieval [3]. Fourier descriptors represent the outside contour of a shape by means of a limited number of coefficients in the frequency domain. Since such coefficients also carry information about the size, the orientation, and the position of the shape, they have to be properly normalized in order to achieve invariance properties with respect to transformations. In addition, to obtain a compact description of a shape, a common approach is to retain only a subset of Fourier coefficients, corresponding to those with frequency closer to zero [14], [20]. The choice of a suitable number M of coefficients to be used has to

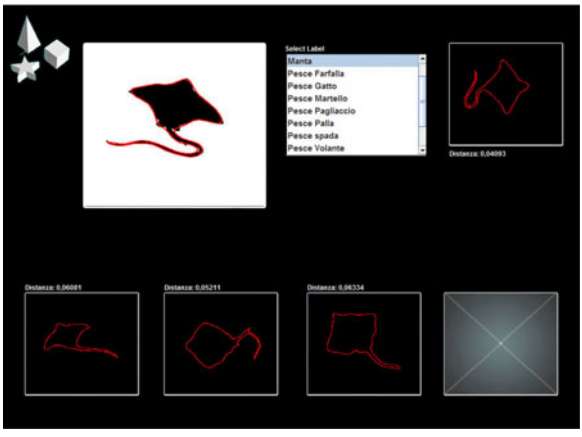


Fig. 1. Interface for manual annotation of prototypes

trade off the accuracy in representing the original boundary with the compactness and simplicity of the representation. Readers can find main details about this issue in [3]. In this work, we fixed empirically the suitable number of Fourier coefficients (see Section 3). We indicate by $\mathbf{s} = (s_1, s_2, ..., s_M)$ the representation of an object shape by means of its M Fourier descriptors.

Once all shapes have been represented by Fourier descriptors, a set of shape prototypes is automatically defined by a clustering process so as to group similar shapes into a number of classes and represent each class by means of a prototypical shape. In this work, we employ the well-known Fuzzy C-Means (FCM) clustering algorithm [4] to group similar shapes into overlapping clusters representing several semantic classes.

Given a set of N shapes represented by Fourier descriptors $S = \{\mathbf{s}_j\}_{j=1}^{N}$ and given a number of clusters K, the FCM provides a fuzzy partition matrix $\mathbf{U} = [u_{jk}]_{j=1...N}^{k=1...K}$ representing membership degrees of shapes to clusters. These are used to derive a prototypical shape for each cluster. Namely, for each cluster, the shape with maximal membership degree is selected as prototype. We denote by $\mathbf{p}_k$ the prototypical shape of cluster k.

Once shape prototypes have been derived, these are manually annotated by a domain expert according to a set of C semantic categories. Precisely, each derived shape prototype is associated to a unique textual label corresponding to a semantic class represented by the prototype. Fig. 1 shows the user interface that we developed for prototype annotation. Of course, different prototypical shapes may convey the same semantic content (i.e., several different shapes may convey the same class of objects), i.e. $K \geq C$. We consider such prototypes to belong to the same semantic class: thus, such prototypes will have attached the same class label. As a result, we may have different shape prototypes with attached the same textual description.

Every time a new shape is added to the database, its Fourier descriptors $\mathbf{s}_{j*}$ are matched against Fourier descriptors of all prototypes $\mathbf{p}_k$ by computing

the Euclidean distance $\|\mathbf{s}_{j*} - \mathbf{p}_k\|$. Then, membership degrees of the shape to clusters are calculated according to a Gaussian membership function as follows:

$$\mu_{j*k} = e^{\frac{-\|\mathbf{s}_{j*} - \mathbf{p}_k\|}{2\sigma}} \tag{1}$$

where $\sigma = \frac{\|\mathbf{p}_k - \mathbf{p}_h\|}{r}$ is calculated by using the *first-nearest-neighbor* heuristic where $\mathbf{p}_h$ is the cluster prototype nearest to $\mathbf{p}_k$ and r is an overlap parameter ranging in $[1.0, 2.0]$.

Thus, each object shape $\mathbf{s}_{j*}$ is associated with a fuzzy set of labels L_{j*} defined as follows:

$$L_{j*} = \{\mu_{j*1}, \mu_{j*2}, \ldots, \mu_{j*C}\} \tag{2}$$

where μ_{j*i} represents the membership degree of the $j*$-th shape to the i-th semantic class that are obtained by computing the maximum value of membership degrees of the shape with respect to all prototypes $\mathbf{p}_k$ representative of the i-th semantic class, namely:

$$\mu_{j*i} = max_{\mathbf{p}_k \in C_i} \{\mu_{j*k}\} \tag{3}$$

As a result, an object shape usually belongs to multiple semantic classes with different degrees of membership.

When the matching among a new shape and visual prototypes of all semantic classes provides membership values lower than a given threshold, a new class is added. Precisely, Fourier descriptors of the new shape are considered as the visual prototype of the new semantic class (i.e. $\mathbf{p}_{K+1} = \mathbf{s}_{j*}$). Then, the newly created prototype is manually labeled by the domain expert with a keyword that describes the new added semantic class.

The use of shape prototypes, which represent an intermediate level of visual signatures, facilitates the annotation process, since only a reduced number of shapes (the prototypical ones) need to be manually annotated. Secondly, the use of prototypes simplifies the search process in a retrieval system. Indeed, since any single user query is likely to match with high degree only a small number of objects, a large number of unnecessary comparisons is avoided during search by performing matching with shape prototypes rather than with specific shapes. In other words, prototypes act as a filter that reduces the search space quickly while discriminating the objects.

3 Preliminary Results

To test the effectiveness of the proposed approach, we used the Surrey Fish data set provided by [1], consisting of 1,100 text files. Each file contains the coordinates of boundary points of an object representing a marine animal. According to [3], images were manually classified into 10 semantic classes, as follows: "Seahorses" (5), "Seamoths" (6), "Sharks" (58), "Soles" (52), "Tonguefishes" (19), "Crustaceans" (4), "Eels" (26), "U-Eels" (20), "Pipefishes" (16), and "Rays" (41). Images not belonging to any category were assigned to a default class

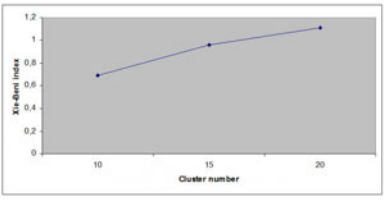


Fig. 2. Values obtained for the Xie-Beni index

(853). In these preliminary experiments, we employed a portion of the Surrey Fish data set including only classes with an higher number of images. Thus, classes such as "Crustaceans", "Seahorses" and "Seamoths" were not considered obtaining a data set composed of 232 images. Table 1 shows some images in the considered data set portion, along with their respective class.

Table 1. Sample images of the considered data set portion

Shark	U-Eel	Eel	Sole	Ray	Pipefish	Tonguefish

The suitable number of Fourier coefficients to represent each shape was empirically fixed to 32, according to [3].

The collection of the obtained 232 shape descriptors was divided into two parts: the 90% of shape descriptors represented the training set and the remaining 10% represented the test set. A 10-fold cross-validation procedure was executed in order to achieve more reliable results. The FCM was applied to the training set in order to derive clusters of similar shapes. We performed several runs by setting, in each trial, different cluster number ($K = 20, 15, 10$). Moreover, to establish the most suitable number of clusters, after each run of FCM, we calculated the Xie-Beni index [19] being one of the most employed indexes in literature to evaluate the goodness of the obtained partitions. Fig. 2 shows the values obtained for the Xie-Beni index. Here, it can be seen that the best partition of shapes (characterized by the smallest value of the cluster validity index) was obtained in correspondence of $K = 10$. As a result of the shape clustering task, for each cluster, a shape prototype was derived as explained in section 2. Therefore, 10 different shape prototypes (one for each derived cluster) were determined together with the matrix containing the membership degrees of each shape to each derived cluster. Successively, shape prototypes were manually annotated by associating to each of them a label related to the corresponding semantic class. Table 2 shows some information about the obtained clusters in

Table 2. The resulting clusters with the corresponding shape prototypes and the associated class labels

Cluster number	Cardinality	Class label	Prototype
1	22	Sharks	
2	20	Soles	
3	28	Sharks	
4	10	U-Eels	
5	34	Rays	
6	26	Soles	
7	8	U-Eels	
8	25	Eels	
9	14	Soles	
10	24	Tonguefishes	

one of the performed trials. In each row, the cardinality of each cluster with the derived shape prototype and the associated label (corresponding to the semantic class) are indicated. It can be observed that for some semantic classes, such as "Soles" and "Sharks", more than one prototype was derived. In effect, such classes are formed by very dissimilar shapes that, therefore, were categorized in different clusters. This shows that different shape prototypes can be representative of the same semantic class so that different shapes convey the same semantic concept. Moreover, FCM has recognized pipefishes very similar to other shapes such as sharks or eels. As a consequence, no prototype was derived for the semantic class "Pipefishes". Of course, this will strongly affect the effectiveness of the annotation process for images including shapes of the unidentified category. Another consideration concerns the choice of the label associated to the last shape prototype. In effect, the last cluster includes tonguefishes with some sharks that the algorithm recognized as very similar shapes. Consequently, although the derived prototype is a shark, it was labeled as a tonguefish reflecting the nature of the cluster. This has decreased the performance of the annotation process for the class "Tonguefishes", as it can be observed in table 3.

As a final step, the effective phase of shape annotation was performed. Specifically, each shape included in the testing set was matched with the derived shape prototypes and, then, it was annotated by assigning a fuzzy set derived by computing the membership degrees according to the Eq. 1. In our test, the value for the overlap parameter r was fixed to 1.50. The overall annotation process was evaluated using Precision and Recall measures on the testing set. In particular, first, we associated to each test shape the label corresponding to the semantic class with the higher membership value. Then, for each label associated to semantic classes, recall and precision were evaluated based on the true annotations

(manual annotations made by the domain expert) and the automatic annotations. More precisely, let s_j be a test shape, t_j its true annotation and a_j its automatic annotation. For each single class label w, precision and recall were computed as follows:

$$P(w) = \frac{|\{s_j | w \in t_j \wedge w \in a_j\}|}{|\{s_j | w \in a_j\}|} \; R(w) = \frac{|\{s_j | w \in t_j \wedge w \in a_j\}|}{|\{s_j | w \in t_j\}|}$$

Table 3 shows the values for Precision and Recall measures related to the considered semantic classes. It can be observed that good results are obtained for almost all classes. Only in correspondence of the class "Pipefishes", we obtained null values for Precision and Recall. This is because pipefishes actually appear very similar to sharks or eels, thus FCM fails in recognizing them as a separate class.

Table 3. Precision and Recall values for the annotation process

Class	Precision	Recall
U-Eels	0.50	1.00
Sharks	0.60	0.60
Tonguefishes	0.20	0.50
Rays	1.00	0.50
Soles	1.00	0.60
Eels	1.00	0.40
Pipefishes	0.00	0.00

4 Conclusions

In this paper, we proposed an approach to fuzzy labeling of object shapes for automatic image annotation. Preliminary results encourage the application of the proposed approach to wider contexts. However, the effectiveness of the annotation process could be improved by applying more suitable clustering methods for automatic prototype derivation. A future development could concern the use of fuzzy clustering techniques equipped with partial supervision mechanisms. By applying this kind of techniques, a low number of labeled shapes given in input to the algorithm together with unlabeled shapes allows to obtain better clustering results in terms of more significant prototypes that better represent the several semantic classes.

References

1. Abbasi, S., Mokhtarian, F., Kittler, J.: SQUID Demo Dataset 1,500 (1997), http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html
2. Athanasiadis, T., Mylonas, P., Avrithis, Y., Kollias, S.: Semantic Image segmentation and object labeling. IEEE Transaction on Circuits and Systems for Video Technology 17(3), 298–312 (2007)

3. Bartolini, I., Ciaccia, P., Patella, M.: WARP: Accurate retrieval of shapes using phase of Fourier descriptors and Time warping distance. IEEE Transaction on Pattern Analysis and machine Intelligence 27(1), 142–147 (2005)
4. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York (1981)
5. Borras, A., Llados, J.: Object Image Retrieval by Shape Content in Complex Scenes Using Geometric Constraints. Patt. Recognition and Image Analysis, 325–332 (2005)
6. Chander, V., Tapaswi, S.: Shape Based Automatic Annotation and Fuzzy Indexing of Video Sequences. In: Proc. of the 2010 IEEE/ACIS 9th Int. Conf. on Computer and Information Science (ICIS 2010), Washington, DC, USA, pp. 222–227 (2010), http://dx.doi.org/10.1109/ICIS.2010.16, doi:10.1109/ICIS.2010.16
7. Chen, Y., Wang, J.Z.: A region-based fuzzy feature matching approach to content-based image retrieval. IEEE Transaction on Pattern Analysis and Machine Intelligence 40(9), 1252–1267 (2002)
8. Datta, R., Dhiraj, J., Jia, L., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys 40, 1–60 (2008)
9. Monay, F., Gatica-Perez, D.: On image auto-annotation with latent space models. In: Proc. of the Eleventh ACM Int. Conf. on Multimedia (MULTIMEDIA 2003), pp. 275–278. ACM Press, New York (2003)
10. Yavlinsky, A., Schofield, E., Rger, S.: Automated image annotation using global features and robust nonparametric density estimation. In: Proc. of the Int. Conf. on Image and Video Retrieval, Singapore, pp. 507–517 (2005)
11. Lew, M., Sebe, N., Djeraba, L.F., Ramesh, J.: Content-based Multimedia Information Retrieval: State of the Art and Challenges. ACM Transaction on Multimedia Computing, Communications, and Applications, 1–19 (2006)
12. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175 (2001)
13. Platt, J., Cristianini, N., Shawe-Taylor, J.: Large margin DAGs for multiclass classification. Advances in Neural Information Processing Systems 12, 547–553 (2000)
14. Rafiei, D., Mendelzon, A.O.: Efficient Retrieval of Similar Shapes. The Very Large Data Bases Journal 11(1), 17–27 (2002)
15. Symonova, O., Dao, M.S., Ucelli, G., De Amicis, R.: Ontology Based Shape Annotation and Retrieval. In: Proc. of the ECAI 2006 Int. Workshop on Contexts and Ontologies: Theory, Practice and Applications, Riva del Garda, Italy (2006)
16. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence 22, 1349–1380 (2000)
17. Sudderth, A.B., Torralba, A., Freeman, W.T., Willsky, A.S.: Learning hierarchical models of scenes, objects and parts. In: Proc. of Int. Conf. on Computer Vision, Beijing, China, vol. 2, pp. 1331–1338 (2005)
18. Wu, G., Chang, E., Li, C.: SVM binary classifier ensembles for image classification. In: Proc. of ACM Conf. on Information and knowledge Management, pp. 395–402 (2001)
19. Xie, X.L., Beni, G.: A Validity Measure for Fuzzy Clustering. IEEE Transaction on Pattern Analysis and Machine Intelligence 13, 841–847 (1991)
20. Zhang, D., Lu, G.: A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval. In: Proc. of Fifth Asian Conf. on Computer Vision, pp. 646–651 (2002)

Stable Automatic Unsupervised Segmentation of Retinal Vessels Using Self-Organizing Maps and a Modified Fuzzy C-Means Clustering

Carmen Alina Lupaşcu and Domenico Tegolo

Dipartimento di Matematica e Informatica
Università degli Studi di Palermo
Palermo, Italy
lupascu@math.unipa.it, domenico.tegolo@unipa.it

Abstract. In this paper an automatic unsupervised method for the segmentation of retinal vessels is proposed. Three features are extracted from the tested image. The features are scaled down by a factor of 2 and mapped into a Self-Organizing Map. A modified Fuzzy C-Means clustering algorithm is used to divide the neuron units of the map in 2 classes. The entire image is again input for the Self-Organizing Map and the class of each pixel will be the class of its best matching unit in the Self-Organizing Map. Finally, the vessel network is post-processed using a hill climbing strategy on the connected components of the segmented image.

The experimental evaluation on the DRIVE database shows accurate extraction of vessels network and a good agreement between our segmentation and the ground truth. The mean accuracy, 0.9482 with a standard deviation of 0.0075, is outperforming the manual segmentation rates obtained by other widely used unsupervised methods. A good kappa value of 0.6565 is comparable with state-of-the-art supervised or unsupervised approaches.

Keywords: Retinal Vessels, Self-Organizing Map, Fuzzy C-Means.

1 Introduction

Automatically generated vessel maps are important in the diagnosis of many eye pathologies. The vessel tree extracted from the retinal images is used to guide the identification of retinal landmarks like optic disc or fovea. It is useful also in the registration of retinal images. Branching and crossover points in the vasculature structure are used as landmarks for image registration.

The development of an efficient computer based system for automatically segment the blood vessels in retinal images allows eye care specialists to screen larger populations for vessel abnormalities.

1.1 Related Work

Many different approaches for automated vessel segmentation have been proposed and they can be divided in two groups: supervised and unsupervised.

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 244–252, 2011.

Supervised methods require a feature vector for each pixel and manually labeled images for training. They use the neural networks [15], the K-nearest neighbor classifier [14], primitive-based methods [18], a Gaussian mixture model classifier [16], or an AdaBoost classifier [12] in order to distinguish vessel from non-vessel pixels. Unsupervised methods in the literature comprise the matched filter responses [2], adaptive thresholding [6], vessel tracking [3], morphology based methods [21] and neural network learning techniques [10].

Supervised learning assumes that the training samples are classified by an expert as either vessel or non-vessel and this is a tedious process. In the same time the supervised training process is time consuming. The lack of experts and the time consuming processes involved in the automatic supervised methods described in the literature determined us to search for an automatic unsupervised method for classification of the pixels from a retinal image as vessel or non-vessel. We found that using Self-Organizing Maps and a modified Fuzzy C-Means for clustering map units gives good results in clustering pixels and it is also very fast as it is possible to notice from the analysis of the experimental results and from the analysis of the execution time. This method is attractive also because training is performed on resized features extracted from the same image we want to segment, hence there is no need to develop a separate training set like in other supervised or unsupervised methods.

2 Methodology

In [12] we described a new supervised method for retinal vessel segmentation called FABC. The method was based on computing feature vectors for every pixel in the image and training an AdaBoost classifier with manually labeled images.

2.1 Pixel Features

The feature vector consisted essentially of the output of filters plus vesselness and ridgeness measures based on eigen-decomposition of the Hessian computed at each image pixel and a two-dimensional Gabor wavelet transform response taken at multiple scales. The components of the feature vector were: the Gaussian and its derivatives up to order 2, the green channel intensity of each pixel, a multiscale matched filter for vessels using a Gaussian vessel profile [17], Frangi's vesselness measure [4], Lindeberg's ridge strengths [9], Staal's ridges [18], a two-dimensional Gabor wavelet transform response taken at multiple scales [16] and values of the principal curvatures, of the mean curvature, of the principal directions and of the gradient of the image (the computation method of each component is described better in [12]).

The scales used in order that vessels with various dimensions could be detected were $4{:}\sqrt{2}$, 2, $2 * \sqrt{2}$ and 4, hence the total number of features is 41.

In [11] we described five feature selection heuristics designed to evaluate the usefulness of features through feature subsets. Experiments showed that the features that seemed to play the most important discriminatory role, i.e., the ones that were selected by *all* the heuristics, were the 2nd-order derivative of the Gaussian in the y direction at scale $2\sqrt{2}$, the maximum response of a multiscale matched filter using a Gaussian vessel profile, and the feature containing information about Staal's ridges. Consequently,

we used only these 3 features for the feature vectors used for clustering with the Self-Organizing Map and the modified Fuzzy C-Means algorithm.

In [10] we presented an automatic unsupervised method for the segmentation of retinal vessels based on a Self-Organizing Map (that was trained on a part of the same tested image) and K-means clustering algorithm (used to divide the map units in 2 classes). The training sample was chosen as half of the FOV (field of view) pixels, selected randomly. Randomness of the pixel samples gave us different results when we run the algorithm several times, hence we were constrained to produce first a soft classification by summing the segmented images resulted in various runs and secondly a hard classification was obtained by thresholding at half of the maximum gray level.

In order to cope with the randomness of the pixel samples used for training the Self-Organizing Map that gave different results when running several times, in this new method we resized the pixel features by a scale factor of 2 and then we used all the resized features for training the Self-Organizing Map.

2.2 Self-Organizing Maps

A Self-Organizing Map (SOM) is a neural network that is trained, using unsupervised learning, to build a map of the input space of the training samples. A new input vector will be automatically classified using the map built in the training phase. SOM was developed by Teuvo Kohonen in 1980 [7]. It consists of m neurons organized on a regular low-dimensional grid. Each neuron i is a d-dimensional weight vector $(w_{1i}, ..., w_{di})$ called *prototype vector or codebook vector*, where d is equal to the dimension of the input vectors. Usually, before the training phase, the prototype vectors are linearly initialized. It has been suggested by Kohonen et al. [8] to use rectangular (but non quadratic) maps and the number of neurons of the map is computed as 5 times the square root of the number of training samples.

After the number of map units has been determined, the map size is determined by setting the ratio between column number and row number of map units equal to the ratio of two biggest eigenvalues of the training data. The product of the column and row numbers must be as close to the number of map units as possible [19]. The size of the images used for testing our method is 584x565, hence the number of pixels in the image is 329960. The number of the FOV pixels in the image is 219848, hence scaling down the features by a factor of 2 we get 54165 pixels. The number of map units is about $5\sqrt{54165} = 1164$ neurons. Following the above rules, a 70×17 map has been used for training.

Training the Self-Organizing Map. The SOM training algorithm is based on competitive learning which is a particular case of neural network unsupervised learning. At each training iteration, a sample vector x is randomly chosen from the training set. Euclidean distances between x and all the prototype vectors are computed, in order to find the best matching neuron unit (BMU). The BMU is selected as the unit that is the nearest to the input vector at an iteration t, using $\|x(t) - w_c(t)\| = \min_i \|x(t) - w_i(t)\|$, where w_c is the weight of the winner neuron. After finding the BMU, the prototype vectors of the BMU and its neighbors are moved closer to the input vector using the following update rule for a neuron i: $w_i(t+1) = w_i(t) + \alpha(t)[x(t) - w_i(t)]$, for $i \in N_c$,

while $w_i(t+1) = w_i(t)$, for $i \notin N_c$, where $\alpha(t)$ is the learning rate and N_c is the neighborhood of the winning neuron.

Clustering the Self-Organizing Map using a modified Fuzzy C-Means. After training the map, in order to cluster the SOM units in two classes, instead of using K-Means like we have done in previous work, in this novel method we use a modified Fuzzy C-Means with the aim to increase sensitivity. By applying Fuzzy C-Means one map unit can belong to both classes at the same time but with different degrees.

The classical Fuzzy C-Means is based on minimization of the following objective function: $J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m ||x_i - c_j||^2, \quad 1 \leq m < \infty$ where m is the fuzziness factor, u_{ij} is the degree of membership of x_i in the class j and c_j is the center of the cluster. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of the degree of membership u_{ij} by $u_{ij} = 1/\sum_{k=1}^{C}\left(\frac{||x_i - c_j||}{||x_i - c_k||}\right)^{\frac{2}{m-1}}$ and the update of the cluster centers c_j by $c_j = \left(\sum_{i=1}^{N} u_{ij}^m \cdot x_i\right) / \left(\sum_{i=1}^{N} u_{ij}^m\right)$, where C is the number of classes (in our case $C = 2$) and N is the number of map units.

The iteration will stop when $max_{ij}||u_{ij}^{(k+1)} - u_{ij}^{(k)}|| < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This process converges to a local minimum of J_m.

As the classical Fuzzy C-Means delivers different results when running several times, we modify the classical Fuzzy C-Means by initializing the cluster centers at the first step of the algorithm using $C = 2$ prototype vectors having the maximum Euclidean distance between them. In this way, at every run we will get the same $C = 2$ prototype vectors as initial cluster centers, hence same results when running several times. Moreover, we choose the $C = 2$ prototype vectors having the maximum Euclidean distance between them as we want to have the best discrimination between clusters [5].

After the neurons from the SOM were classified, the class of a new input pixel will be the class of its BMU.

2.3 Post-processing the Segmented Images

We post-process the segmented images, trying to eliminate small connected components in order to remove noisy pixels and to improve in this way the segmentation accuracy and the agreement between our segmentation and the ground truth. A hill climbing strategy was used in order to determine the connected components to be removed from the segmentation. If $CC = \{c_1, ..., c_n\}$ is the set of the n image's connected components ordered in ascending order by the number of pixels in the connected component, the algorithm starts with the set $Toberemoved = \{c_1, c_2\}$. The mean of the cardinalities of the connected components included in the set $Toberemoved$ is computed, as well as their standard deviation. Connected component c_3 is added to the set $Toberemoved$ if $|c_3| < mean(Toberemoved) + 3std(Toberemoved)$. The algorithm stops when a

successive connected component from the set CC can not be added to the set of connected components to be removed.

3 Experimental Results

3.1 Evaluation Measures

The method was tested on the DRIVE (Digital Retinal Images for Vessel Extraction) publicly available database. The data set includes 40 584x565 fundus images. We use only the 20 images from the test set for testing our methodology. All images are available for download at `http://www.isi.uu.nl/Research/Databases/DRIVE/download.php` (the web site of Image Sciences Institute).

Performance is given mainly as accuracy and kappa value.

The accuracy (ACC) for one image is the fraction of pixels correctly classified $ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+FN+FP+TN}$, where TP is the number of pixels correctly classified as vessel pixels, TN the number of pixels correctly classified as non-vessel pixels, FN the number of pixels incorrectly classified as non-vessel pixels and FP is the number of pixels incorrectly classified as vessel pixels.

The sensitivity (SE) is computed by dividing TP by the total number of vessel pixels in the gold standard segmentation: $sensitivity = \frac{TP}{P} = \frac{TP}{TP+FN}$.

The specificity (SP) is computed as TN divided by the total number of non-vessel pixels in the gold standard: $specificity = \frac{TN}{N} = \frac{TN}{FP+TN}$.

We compute also the kappa value (a measure for observer agreement, where the two observers are the gold standard and the segmentation method): $kappa = \frac{P(A)-P(E)}{1-P(E)}$, where $P(A) = \frac{TP+TN}{P+N}$ is the proportion of times the 2 observers agree, while $P(E) = \frac{TP+FP}{P+N} * \frac{TP+FN}{P+N} + (1 - \frac{TP+FP}{P+N})(1 - \frac{TP+FN}{P+N})$ is the proportion of times the 2 observers are expected to agree by chance alone.

3.2 Results

In Table 1 we may see the segmentation results for the 20 test images from the DRIVE database after post-processing.

Overall the 20 test images, after the post-processing, the mean accuracy is 0.9482 with a standard deviation of 0.0075. The mean Kappa value is 0.6565. As shown in [20] and

Table 1. Results for the 20 test images from the DRIVE database, after post-processing. SE indicates sensitivity, SP indicates specificity, ACC indicates the accuracy and $Kappa$ indicates the kappa value

Image	1	2	3	4	5	6	7	8	9	10
SE	74.44	54.78	60.83	51.42	61.09	58.46	49.81	55.61	55.00	67.62
SP	97.19	99.60	97.47	99.66	98.31	98.24	99.47	96.66	98.32	97.66
ACC	0.9517	0.9506	0.9383	0.9526	0.9484	0.9439	0.9496	0.9317	0.9482	0.9520
Kappa	0.7057	0.6672	0.6282	0.6410	0.6607	0.6388	0.6174	0.5436	0.6046	0.6715

Image	11	12	13	14	15	16	17	18	19	20
SE	53.80	76.08	64.07	80.56	65.52	77.74	71.34	80.54	78.21	65.55
SP	99.49	96.20	97.46	94.75	98.86	96.61	96.59	96.41	97.02	97.91
ACC	0.9541	0.9447	0.9421	0.9361	0.9648	0.9491	0.9446	0.9515	0.9547	0.9554
Kappa	0.6537	0.6731	0.6510	0.6347	0.7078	0.7053	0.6541	0.6978	0.7156	0.6592

Table 2. Overview of the performance of different methods. $Kappa$ indicates the kappa value and ACC indicates the accuracy.

Segmentation method	Drive set	
	$Kappa$	ACC
FABC (training and test confined to the dedicated sets from the database) [12]	0.7200	0.9597 (0.0054)
FABC (leave-one-out tests) [12]	-	0.9575
Human observer	0.7589	0.9473 (0.0048)
Soares et al. [16]	-	0.9466
SOM and modified Fuzzy C-Means	0.6565	0.9482(0.0075)
SOM and K-means [10]	0.6562	0.9459(0.0094)

Segmentation method	Drive set	
	$Kappa$	ACC
Staal et al. [18]	0.7345	0.9442 (0.0065)
Niemeijer et al. [14]	0.7145	0.9416 (0.0065)
Zana et al. [21]	0.6971	0.9377 (0.0077)
Al-Diri et al. [1]	0.6716	0.9258 (0.0126)
Jiang et al. [6]	0.6399	0.9212 (0.0076)
Martinez et al. [13]	0.6389	0.9181 (0.0240)
Chaudhuri et al. [2]	0.3357	0.8773 (0.0232)
All background	0	0.8727 (0.0123)

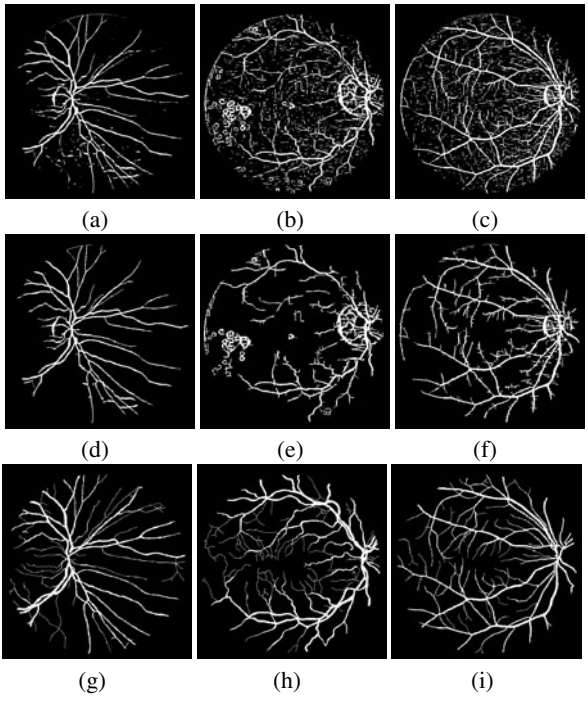


Fig. 1. Best segmentation in terms of ACC (Image $15\_test.tif$ from DRIVE database. a), d) our segmentation before and after post-processing, g) ground truth). Worst segmentation in terms of ACC and Kappa value (Image $08\_test.tif$ from DRIVE database. b), e) our segmentation before and after post-processing, h) ground truth). Best segmentation in terms of Kappa value (Image $19\_test.tif$ from DRIVE database. c), f) our segmentation before and after post-processing, i) ground truth).

in Table 2, the mean ACC of our proposed method outperforms the mean ACC of any of unsupervised methods used for comparison. As in [20] the proposed method enhances all region-of-interest, i.e. both vessel network and pathological findings in the soft classification (see Figure 1). This effect is desired in computer-aided diagnosis tools.

4 Conclusions

We have presented an automatic unsupervised method for retinal vessel segmentation based on Self-Organizing Maps and Fuzzy C-Means clustering. The choice to use Self-Organizing Maps instead of a simple Fuzzy C-Means is based upon the fact that a SOM provides a direct mean to visualize relations among different clusters (represented by the prototype vectors in the input space and by the map's neurons in the output space).

One important novelty of the method described in this paper is its **stability**. As the classical Fuzzy C-Means delivers different results when running several times, we modified the classical Fuzzy C-Means by initializing the cluster centers at the first step of the algorithm using $C = 2$ prototype vectors having the maximum Euclidean distance between them and this implies same results when running several times, hence reproducibility of the results which is very useful for practical purposes. In the same time, resizing the pixel features by a scale factor of 2 and then using all the resized features for training the Self-Organizing Map, leads also to stability of the segmentation results.

In a previous work [10], we have used K-means instead of the modified Fuzzy C-Means for clustering the SOM units. In this paper, we propose a novel method based on a revised version of Fuzzy C-Means for clustering the SOM units as it increases the sensitivity apart from stabilizing the segmentation results. In fact, using a modified K-Means, before post-processing, we get a mean sensitivity of 0.5030, while using a modified Fuzzy C-Means, before post-processing, we get a mean sensitivity of 0.6965, hence an improvement of 27.78%.

The method we proposed has the advantage that it uses knowledge about the vessel network morphology like the most accurate supervised methods, but is completely unsupervised as we do not have any a priori knowledge about the labels of the pixels we want to classify as vessel or non-vessel. Another advantage of the proposed method is its fast computational time, compared to supervised methods which are computationally more expensive.

Although, in most cases, the vessel network produced before the post-processing was acceptable, as we may see from Figure 1, post-processing methods removed efficiently false positives and improved the mean accuracy from 0.9324 (see Table 3) to 0.9482 (see Table 1), hence an improvement of 1.69%.

Table 3. Segmentation results for the 20 test images from the DRIVE database, before post-processing. ACC indicates the accuracy.

Image	1	2	3	4	5	6	7	8	9	10
ACC	0.9317	0.9511	0.9107	0.9526	0.9301	0.9282	0.9508	0.9120	0.9365	0.9284
Image	11	12	13	14	15	16	17	18	19	20
ACC	0.9542	0.9184	0.9229	0.9113	0.9625	0.9308	0.9257	0.9253	0.9293	0.9362

In the future we would like to study the influence on the results of the choice of some parameters of the SOM map (like the number of iterations, the size of the initial radius of the neighborhood and the choice of distance measure) and of the choice of the post-processing technique. In the same time, we would like also to study which is the best strategy for reducing the training samples (by scaling down the features as we are

now doing, or by scaling down the image by different factors and computing each time the features).

References

1. Al-Diri, B., Hunter, A., Steel, D.: An active contour model for segmenting and measuring retinal vessels. IEEE Transactions on Medical Imaging 28(9), 1488–1497 (2009)
2. Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M.: Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Transactions on Medical Imaging 8(3), 263–269 (1989)
3. Chutatape, O., Zheng, L., Krishnan, S.M.: Retinal blood vessel detection and tracking by matched gaussian and kalman filters. Proceeding of IEEE Int. Conf. Emg. and Bio. Society 20(6), 3144–3149 (1998)
4. Frangi, A., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
5. Giancarlo, R., Lo Bosco, G., Pinello, L.: Distance functions, clustering algorithms and microarray data analysis. In: Blum, C., Battiti, R. (eds.) LION 4. LNCS, vol. 6073, pp. 125–138. Springer, Heidelberg (2010)
6. Jiang, X., Mojon, D.: Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(1), 131–137 (2003)
7. Kohonen, T.: Self-organization and associative memory. Springer, Berlin (1989)
8. Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J.: Som pak: The self-organizing map program package. Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science (January 1996)
9. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. Int. J. Comp. Vis. 30, 117–156 (1998)
10. Lupaşcu, C.A., Tegolo, D.: Automatic unsupervised segmentation of retinal vessels using self-organizing maps and k-means clustering. In: Proc. of CIBB 2010 Conference, Palermo, Italy (September 2010)
11. Lupaşcu, C.A., Tegolo, D., Trucco, E.: A comparative study on feature selection for retinal vessel segmentation using fabc. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 655–662. Springer, Heidelberg (2009)
12. Lupaşcu, C.A., Trucco, E., Tegolo, D.: Fabc: Retinal vessel segmentation using adaboost. IEEE Transactions on Information Technology in Biomedicine 14(5), 1267–1274 (2010)
13. Martínez-Pérez, M., Hughes, A., Stanton, A., Thom, S., Bharath, A., Parker, K.: Scale-space analysis for the characterisation of retinal blood vessels. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 90–97. Springer, Heidelberg (1999)
14. Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., Abrámoff, M.D.: Comparative study of retinal vessel segmentation methods on a new publicly available database. SPIE Medical Imaging 5370, 648–656 (2004)
15. Sinthanayothin, C., Boyce, F.J., Cook, L.H., Williamson, H.T.: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. Br. J. Ophthalmol. 83, 902–910 (1999)
16. Soares, V.B.J., Leandro, J.G.J., Cesar, R.M.J., Jelinek, F.H., Cree, M.J.: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. IEEE Transactions on Medical Imaging 25(9), 1214–1222 (2006)

17. Sofka, M., Stewart, C.V.: Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures. IEEE Transactions on Medical Imaging 25(12), 1531–1546 (2006)
18. Staal, J., Abrámoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging 23(4), 501–509 (2004)
19. Vesanto, J.: Som-based data visualization methods. Intelligent Data Analysis journal (1999)
20. Vlachos, M., Dermatas, E.: Multi-scale retinal vessel segmentation using line tracking. Computerized Medical Imaging and Graphics 34, 213–227 (2010)
21. Zana, F., Klein, J.C.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. IEEE Transactions on Image Processing 10(7), 1010–1019 (2001)

Cytoplasm Image Segmentation by Spatial Fuzzy Clustering

Laura Caponetti, Giovanna Castellano, Vito Corsini, and Teresa M.A. Basile

Università degli Studi di Bari - Dipartimento di Informatica
via E. Orabona, 4 - 70125 Bari, Italy
{laura,castellano,basile}@di.uniba.it, corsini@poliba.it

Abstract. This work presents an approach based on image texture analysis to obtain a description of oocyte cytoplasm which could aid the clinicians in the selection of oocytes to be used in the assisted insemination process. More specifically, we address the problem of providing a description of the oocyte cytoplasm in terms of regular patterns of granularity which are related to oocyte quality. To this aim, we perform a texture analysis on the cytoplasm region and apply a spatial fuzzy clustering to segment the cytoplasm into different granular regions. Preliminary experimental results on a collection of light microscope images of oocytes are reported to show the effectiveness of the proposed approach.

1 Introduction

Image segmentation is a fundamental step toward high level vision, necessary in many medical applications based on image analysis and computer vision such as object/pattern recognition and tracking, image retrieval, and so on. Image segmentation is the process of partitioning an image into non-overlapped regions which are homogeneous with respect to some characteristics such as intensity, color or texture. In past years, many methods for the segmentation of medical images have been presented [6,16].

Among these methods, clustering-based approaches received a great interest, especially for image segmentation in medical domains. Both crisp and fuzzy clustering schemes have been proposed, but fuzzy techniques have revealed more robust than crisp algorithms in case of images characterized by some form of ambiguity, such as poor contrast, noise and intensity inhomogeneities. Among fuzzy clustering algorithms, the most used one is the well-known Fuzzy C-Means (FCM) algorithm. Despite its widespread use, FCM does not always provide good segmentation results, and this is mainly due to the fact that it does not incorporate any information concerning the spatial context, which is fundamental because the obtained regions are likely to be disjoint, irregular and noisy. In order to achieve more effective segmentation results, many works have been proposed aiming at incorporating the local spatial information into clustering schemes based on the conventional FCM algorithm [14], [1], [4], [15].

Here an approach based on a spatial FCM algorithm for the texture segmentation of the cytoplasm in oocyte images is presented. More specifically, we address

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 253–260, 2011.

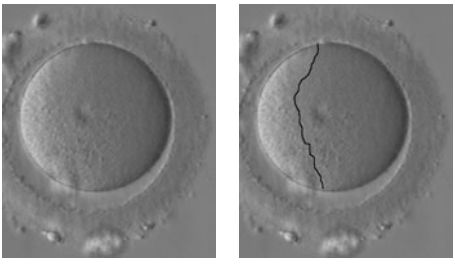


Fig. 1. A light microscope image of a human oocyte: the presence of a clear half-moon like zone (also called halo effect) in one pole of the cytoplasm can be appreciated (depicted in image on the right)

the problem of automatically analyzing the cytoplasm of human oocytes, in order to evaluate their quality during an assisted fertilization process. The oocyte quality assessment is a fundamental issue as, due to ethical and medical reasons, only a restricted number of embryos can be transferred in woman's uterus. However, in some countries law does not allow embryo selection, but only oocytes or pronuclear stage oocytes (i.e. 1 day fertilized oocytes) selection. For these reasons, a method able to support the clinicians in appropriately selecting the most promising (pronuclear stage) oocytes would be desirable.

Generally, selection criteria are based on cell morphology and mostly on the presence/absence of morphological irregular patterns in the cytoplasm that can influence negatively/positively the assisted fertilization process. Indeed, it has been suggested that regular patterns of granularity are related to oocyte quality [10,8,13,9] and in [7] a pronuclear stage oocyte score based on cytoplasmic substructures has been addressed by considering abnormalities such as excessive granularity or cytoplasmic inclusions. If present, granulation may be observed within the cytoplasm and it may be either homogeneously or centrally localized. Moreover, polarization seems to be an important factor for oocyte development and implantation [12]. An example of such effect is given in fig.1.

The ultimate goal of our work is to support the clinicians in the oocyte scoring by means of a system capable to derive a description of the oocyte cytoplasm in terms of different granular regions located in the cytoplasm. In [2] we proposed a multiresolution texture analysis approach that evaluates some texture descriptors (statistical measures in the wavelet transform domain) of a region centred inside the cytoplasmic area and employs the derived features in a clustering phase in order to automatically find a classification of unseen oocytes according to their level of central granularity. A further step forward in the quality assessment of the oocyte maturity is represented by the analysis of the whole cytoplasmic area in order to discover regions with different level of granularity inside the cytoplasm of a single oocyte. This analysis might highlight the presence or not of a particular zone known as polarization or halo effect that is another important factor already studied in literature [12].

For this reason, while in [2] we cluster the texture of different oocytes in order to categorize them, in the present work we use clustering to detect different

granular regions inside the whole cytoplasmic area of each single oocyte. To this aim, we segment the cytoplasm image into different regions according to the texture features by using a spatial fuzzy clustering algorithm [5] that is applied to texture feature vectors rather than to pixels. In details, it works on the representation of a number of blocks in which the image is subdivided, where each block is represented by first order statistics descriptors extracted from the Haar wavelet representation of the block.

The paper is organized as follows. In Section 2 we report the preliminary step of the whole process, i.e. the extraction of the cytoplasm region starting from the oocyte image. Then, Section 3 gives a formalization of the adopted spatial fuzzy clustering algorithm applied after texture analysis. Finally, in Section 4 some preliminary experimental results are provided to show the effectiveness of the proposed approach and some conclusions are drawn in Section 5.

2 Detection of the Cytoplasm Region

This phase is devoted to extract the circular region that corresponds to the cytoplasm inside the oocyte to be used for the segmentation. Indeed, different parts of the oocyte are visible in the image, such as the zona pellucida and the perivitelline space that enclose the cytoplasmic area. Hence, we need a pre-processing step that aims at isolating the region of interest for further segmentation steps. The problem can be solved by combining a priori knowledge about structure and properties of objects with information contained in the image.

Following this approach, we process the oocyte image by considering that the shape of the cytoplasm can be approximated by a circumference. Our method consists of the following steps: (1) obtain the gradient image; (2) use the gradient image to obtain points possibly belong to circumferences of different radius; (3) select the circumference that better approximates the cytoplasm boundary.

In step (2) we use the Hough transform to obtain image points belonging to different circles. The Hough transform [11] is a powerful technique which can be used to isolate features of a particular shape in an image: it is most commonly used for the detection of regular curves such as lines, circles, etc. The main advantage of the Hough transform is that it is tolerant of gaps in curve descriptions and is relatively unaffected by image noise. Its parameter space is three-dimensional. Indeed, a circle with radius R and center (a, b) can be described by the following parametric equations:

$$x = a + R\cos(\theta), \qquad y = b + R\sin(\theta)$$

When the angle θ sweeps through the full 360 degree range, the points (x, y) trace a circle of fixed radius R. The three-dimensionality of the parameter space makes a direct implementation of the Hough transform very expensive in terms of computer memory and time. Anyway, if the radius of circles is known, then the parameter space is reduced and thus the problem is reduced to search for the coordinates (a, b) of the centres. Fig. 2(a) shows the result of applying the Hough transform to the image in fig. 1, where the circle that best fits the cytoplasm boundary is overlapped to the original image.

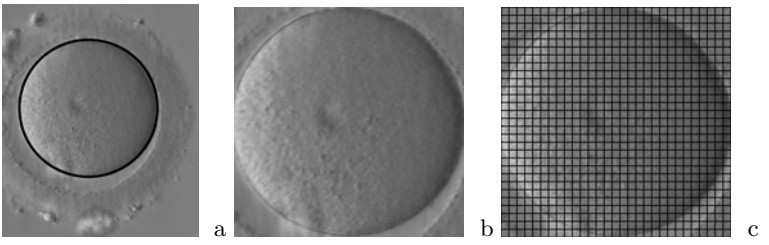


Fig. 2. Result of the cytoplasm boundary detection process. (a) Hough transform of an oocyte image: Best fitting circle - plotted in black - overlapped on the oocyte image. (b) Extracted ROI (c) Square blocks partitioning of the extracted ROI.

Once the cytoplasm boundary is detected, we automatically select the region of interest (ROI) by considering a bounding rectangular region centred in the point of coordinates (a, b) (see fig. 2(b)). This ROI is successively partitioned into square blocks as reported in fig. 2(c) and segmented by spatial fuzzy clustering, as described in the next section.

3 Segmentation by Spatial Fuzzy Clustering

A common strategy of employing a clustering algorithm for image segmentation is to divide the image into a number of blocks and to extract a number of features for each block. Successively, the clustering algorithm is applied to these features and a predefined number of clusters is obtained. Accordingly, in our approach we divide the cytoplasm ROI image into blocks of **wxw** pixels. For each block j we derive a feature vector $\mathbf{x}_j$ by describing its texture using first order statistics applied to a multiresolution decomposition of the block. More specifically, we characterize the texture of each block with the following statistical properties of its multiscale representation:

- mean and variance;
- a measure of relative smoothness;
- third moment, that is a measure of the symmetry of the histogram;
- a measure of uniformity, maximum for an image with all gray levels equal;
- a measure of average entropy, that equals to 0 for a constant image.

For each block of the ROI, we compute these features on the histograms of the subbands obtained by the Haar wavelet decomposition. The derived features are used for segmentation via a clustering process.

A common method to perform clustering is the Fuzzy C-Means (FCM) algorithm [3]. FCM is a partitional clustering method based on the minimization of the following objective function:

$$J = \sum_{j=1}^{N} \sum_{i=1}^{K} (u_{ij})^m \, d\left(\mathbf{x}_j, \mathbf{c}_i\right)^2, \quad 1 < m < \infty, \tag{1}$$

where $d(\mathbf{x}, \mathbf{c})$ is the distance between the observation $\mathbf{x}$ and the cluster centroid $\mathbf{c}$, m is the fuzziness parameter, K is the number of clusters, N is the number of observations, u_{ij} is the membership degree of observation $\mathbf{x}_j$ belonging to the cluster i, calculated as following:

$$u_{ij} = \frac{1}{\sum_{l=1}^{K} \left(\frac{d(\mathbf{x}_j, \mathbf{c}_i)}{d(\mathbf{x}_j, \mathbf{c}_l)} \right)^{\frac{2}{m-1}}}. \tag{2}$$

Using the fuzzy membership matrix $U = [u_{ij}]$, a new position of the i-th centroid is calculated as:

$$\mathbf{c}_i = \frac{\sum_{j=1}^{N} (u_{ij})^m \mathbf{x}_j}{\sum_{j=1}^{N} (u_{ij})^m}. \tag{3}$$

Given the initial parameters (number of clusters K and fuzziness parameter m), FCM iteratively computes the matrix U according to eq. (2), and updates the centroids positions, as in eq. (3). The algorithm terminates after a fixed number of iterations, or if the improvement expressed by J is substantially small. From eq. (1) it can be observed that FCM does not incorporate any spatial dependencies between observations. This may degrade the overall clustering result in case of image segmentation, because neighbouring regions may be highly correlated and thus they should belong to the same cluster.

To take into account the spatial information in the segmentation, a variant of the FCM known as spatial FCM (sFCM) has been proposed in [5]. It uses a spatial function which is defined as:

$$h_{ij} = \sum_{k \in NB(x_j)} u_{ik} \tag{4}$$

where $NB(x_j)$ represents a neighbor of the pixel x_j in the spatial domain. Just like the membership function, the spatial function h_{ij} represents the membership degree of pixel x_j belonging to ith cluster. The spatial function of a pixel for a cluster is large if the majority of its neighbourhood belongs to the same clusters. The spatial function modifies the membership function of a pixel according to the membership statistics of its neighbourhood as follows:

$$u_{ij} = \frac{u_{ij}^p h_{ij}^q}{\sum_{k=1}^{K} u_{kj}^p h_{kj}^q} \tag{5}$$

where p and q are parameters to control the relative importance of both functions. Each iteration of the sFCM includes two steps. The first one is the same as that in standard FCM to calculate the membership function in the feature domain. In the second step, the membership information of each pixel is mapped to the spatial domain, and the spatial function is computed from that. The FCM iteration proceeds with the new membership that is incorporated with the spatial function.

In our approach, to take into account both spatial and texture information of a cytoplasm ROI, the sFCM is applied to texture feature vectors rather than to pixels. Namely, we firstly calculate the membership function in the feature domain as in (2). Then, we update these values by computing the spatial function as follows:

$$h_{ij} = \sum_{k \in NB(b_j)} u_{ik} \tag{6}$$

where $NB(b_j)$ is the set of neighbours of the j-th block that is made up of the 8x8 surrounding blocks. Thus, in our case, the spatial function h_{ij} represents the membership degree of block b_j belonging to ith cluster.

The algorithm terminates when the improvement expressed by J is substantially small, namely the maximum difference between two cluster centres at two successive iterations is less than a threshold. After the convergence, a defuzzification is applied to assign each block to the cluster for which the membership is maximal. As a result, our approach provides a segmented ROI, in which the regions are made up of the clusters obtained in the feature space.

4 Experimental Results

The proposed approach was tested on 60 light microscope images of human oocytes, provided by the *Dipartimento di Endocrinologia ed Oncologia Molecolare e Clinica* of the University "Federico II" of Naples, Italy. For each image the clinicians indicated the presence or not of the polarization in the oocytes. Each image was large 1280×960 pixels.

Firstly, each image was processed using the Hough transform, so as to detect the best circle fitting the real shape of the oocyte cytoplasm. This was done by searching circles of known radius R ranging from 230 to 250 pixels. Once the circular region of the cytoplasm was identified, a squared ROI surrounding the circular region was extracted. All steps involved in this phase were implemented using *ImageJ*, a public domain Java image processing tool.

Then, the extracted ROI was splitted into blocks and the texture analysis was applied on each block in order to derive a feature vector as described in Section 3. To implement texture analysis methods, some Java plugins were developed and properly integrated into the ImageJ environment. Next, the sFCM algorithm was applied on the feature vectors derived from the image blocks in order to segment the ROI into three regions (clusters).

Several runs were carried out varying the block dimensionality in the splitting step, namely 8×8, 12×12, 24×24, 32×32, and the p and q parameters in the clustering step. As in [5], the Xie-Beni index was exploited in order to evaluate the performance of the clustering algorithm. In most cases, the index was better for 8×8 blocks and for the sFCM than the conventional FCM. Furthermore, the $\text{sFCM}_{1,1}$ (sFCM applied with $p = 1$, $q = 1$) showed the best results.

Figure 3 shows the segmentation of a sample ROI of two oocyte images obtained by using the standard FCM algorithm and our variant of the sFCM applied with parameters ($p = 1$, $q = 1$) and ($p = 1$, $q = 2$), respectively. The

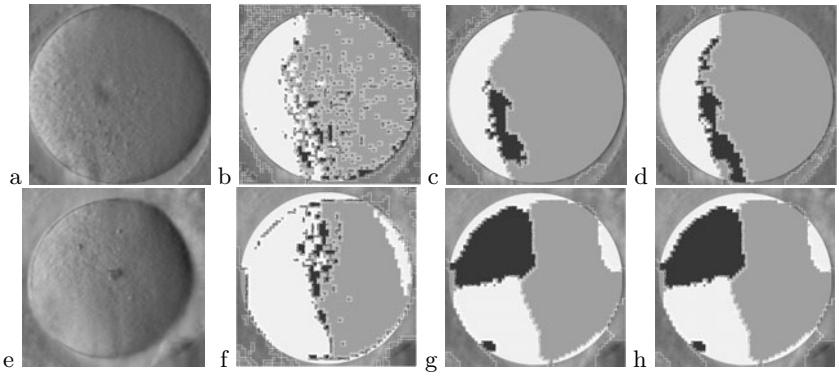


Fig. 3. Segmented images of the extracted cytoplasm (a-e) using FCM (b-f); $sFCM_{1,1}$ (c-g); $sFCM_{1,2}$ (d-h) in oocytes in which, according to the clinicians, polarization is present (a) or not (e)

reported images are two instances of oocytes with and without polarization according to the indication of the clinicians. It can be seen that, in case of oocytes with polarization, the conventional FCM can segment the ROI image into three clusters but spurious blobs of one cluster appear inside other clusters. At the contrary, both sFCM techniques drastically reduce the number of spurious blobs, and the resulting regions are more homogeneous and very quite to the clinicians indication. On the other hand, in case of oocytes without polarization, the proposed approach, differently form classical FCM, correctly fails in finding the half-moon like zone in the cytoplasmic area.

5 Conclusions

An approach for segmenting human oocyte cytoplasm images using fuzzy clustering with spatial information has been presented. Preliminary experimental results on real oocyte images show the effectiveness of the approach and encourage its application in the process of assessing the quality of human oocytes. In particular, the work presented in this paper is intended to provide an approach useful to aid clinicians in making decision about evaluation and selection of the best quality oocytes. The approach could provide a second opinion diagnostic tool for the scoring of human oocytes by embedding it in a system able to analyze microscope images of human oocytes and classify them according to their quality during some kind of assisted fertilization cycles. Currently, further work is in progress to complete the development of such diagnostic tool.

Acknowledgment. This work is partially supported by the italian project "`MIUR-FAR 08-09`: The Molecular Biodiversity LABoratory Initiative (`MBLab-DM19410`)."

References

1. Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., Moriarty, T.: A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans. on Medical Images 21(3), 193–199 (2002)
2. Basile, T.M.A., Caponetti, L., Castellano, G., Sforza, G.: A texture-based image processing approach for the description of human oocyte cytoplasm. IEEE Trans. on Instrumentation and Measurement 59(10s), 2591–2601 (2010)
3. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell (1981)
4. Chen, S., Zhang, D.: Robust image segmentation using FCM with spatial constrained based on new kernel-induced distance measure. IEEE Trans. on Systems, Man and Cybernetics 34(4), 1907–1916 (2004)
5. Chuang, K., Tzenget, H.L., Chen, S., Wu, J., Chen, T.J.: Fuzzy c-means clustering with spatial information for image segmentation. Computerized Medical Imaging and Graphics 30(1), 9–15 (2006)
6. Garbay, C.: Image structure representation and processing discussion of some segmentation methods in cytology. IEEE Trans. Pattern Anal. Mach. Intell. 8(2), 140–146 (1986)
7. Kahraman, S., Yakin, K., Donmez, E., Samli, H., Bahce, M., Cengiz, G., Sertyel, S., Samli, M., Imirzalioglu, N.: Relationship between granular cytoplasm of oocytes and pregnancy outcome following intracytoplasmic sperm injection. Human Reproduction 15(11), 2390–2393 (2000)
8. Loutradis, D., Drakakis, P., Kallianidis, K., Milingos, S., Dendrinos, S., Michalas, S.: Oocyte morphology correlates with embryo quality and pregnancy rate after intracytoplasmic sperm injection. Fertility and Sterility 72(2), 240–244 (1999)
9. Plachot, M., Selva, J., Wolf, J., Bastit, P., de Mouzon, J.: Consequences of oocyte dysmorphy on the fertilization rate and embryo development after intracytoplasmic sperm injection. a prospective multicenter study. Gynecology Obstetrique & Fertilite 30(10), 772–779 (2002)
10. Serhal, P.F., Ranieri, D.M., Kinis, A., Marchant, S., Davies, M., Khadum, I.M.: Oocyte morphology predicts outcome of intracytoplasmic sperm injection. Human Reproducion 12(6), 1267–1270 (1997)
11. Sklansky, J.: On the hough technique for curve detection. IEEE Trans. Comput. 27(10), 923–926 (1978)
12. Stalf, T., Herrero, J., Mehnert, C., Manolopoulos, K., Lenhard, A., Gips, H.: Influence of polarization effects in ooplasma and pronuclei on embryo quality and implantation in an IVF program. Journal of Assisted Reproduction and Genetics 19(8), 355–362 (2002)
13. Suppinyopong, S., Choavaratana, R., Karavakul, C.: Correlation of oocyte morphology with fertilization rate and embryo quality after intracytoplasmic sperm injection. Journal of The Medical Association of Thailand (83), 627–632 (2000)
14. Tolias, Y., Panas, S.: Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions. IEEE Trans. on Systems, Man and Cybernetics, Part A 28(3), 359–369 (1998)
15. Wang, X.Y., Bua, J.: A fast and robust image segmentation using FCM with spatial information. Digital Signal Processing 20, 1173–1182 (2010)
16. Wu, H.S., Barba, J., Gil, J.: Iterative thresholding for segmentation of cells from noisy images. Journal of Microscopy 197(3), 296–304 (2000)

A Memetic Island Model for Discrete Tomography Reconstruction

Marco Cipolla, Giosuè Lo Bosco, Filippo Millonzi, and Cesare Valenti

Dipartimento di Matematica e Informatica
Università di Palermo, via Archirafi 34, 90123, Italy
{cipolla,lobosco,millonzi,valenti}@unipa.it

Abstract. Soft computing is a term indicating a coalition of methodologies, and its basic dogma is that, in general, better results can be obtained through the use of constituent methodologies in combination, rather than in a stand alone mode. Evolutionary computing belongs to this coalition, and thus memetic algorithms. Here, we present a combination of several instances of a recently proposed memetic algorithm for *discrete tomography reconstruction*, based on the *island model* parallel implementation. The combination is motivated by the fact that, even though the results of the recently proposed approach are finally better and more robust compared to other approaches, we advised that its major drawback was the computational time. The underlying combination strategy consists in separated populations of agents evolving by means of different processes which share some individuals, from time to time. Experiments were performed to test the benefits of this paradigm in terms of computational time and correctness of the solutions.

1 Introduction

The goal of *computerized tomography* [1] is the recovery of three-dimensional objects from their projections. In the general case, in order to obtain high resolution slices, several hundreds of projections are required. *Discrete tomography* [2] is the particular reconstruction case of objects which have a few density values. Under certain conditions it is possible to reduce the number of required projections.

For instance, it can be proved in polynomial time whether there exists any object compatible with just a pair of projections [3,4]. Unfortunately, the reconstruction task is quite complex when many projections [5] or a small set of them [6] are used. Anyhow, custom reconstruction algorithms were designed for particular classes of images (e.g *periodic images* [7], which have repetitions of pixels along some directions, and *hv-convex polyominoes*, which are connected sets with 4-connected rows and columns). Besides deterministic methods [4,8,9,10], heuristic [11] and evolutionary approaches [12,13,14] shown their reconstruction ability, too. Network flows were applied [9,13].

Genetic algorithms (GAs) are often considered as global optimization methods and, since local optima can be reached in consideration of their random nature, then their convergence to a global optimum is guaranteed in a weak probabilistic

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 261–268, 2011.

sense. *Memetic algorithms* [15] (MAs) belong to the same class of GAs as they explore the solutions space through cooperative and competitive operators on proper *agents* [16]. With respect to GAs, these latter algorithms induce the evolution not only between consecutive generations, but also among the agents within the same generation.

Parallel genetic algorithms (PGAs) are parallel implementations of their corresponding GAs and usually provide better performances [19]. Moreover, PGAs fully respect the soft computing dogma since they can be described as combination of several instances of GA following some combination strategies. PGAs simultaneously span different subspaces, thus avoiding traps due to low-quality solutions, share the memory to solve the storage population problem and use more than one processing unit to evaluate the fitness. A taxonomy of PGAs is reported in [17]. In this paper, we present a faster parallel version of the memetic reconstruction algorithm we already introduced in [18] based on multi-population model.

2 Memetic Reconstruction

Let us represent a binary image I with $n \times m$ pixels by a matrix $A = \{a_{ij}\}$, whose entries are equal to 0 if the corresponding pixels in I are black (i.e. they belong to the background), or equal to 1 if they are white (i.e. they belong to the object). The *projection line* passing through a_{ij} with direction $\mathbf{v} \equiv (r, s)$, where $r, s \in \mathbb{Z}$ and $|r| + |s| \neq 0$, is the subset of A:

$$\ell_{\mathbf{v}}(i, j) = \{a_{i'j'} \in A \ : \ i' = i + zs, \ j' = j - zr \text{ with } z \in \mathbb{Z}\}.$$

Let $t(\mathbf{v})$ be the number of distinct projection lines parallel to $\mathbf{v}$ and $\mathcal{L}_k^{\mathbf{v}}(A)$ be one of these lines that intersect A, with $k = 1, \ ..., \ t(\mathbf{v})$. Denoting with $p_k^{\mathbf{v}}$ the number of 1's on $\mathcal{L}_k^{\mathbf{v}}(A)$ along $\mathbf{v}$, the *projection* is:

$$P_{\mathbf{v}} = (p_1^{\mathbf{v}}, \ p_2^{\mathbf{v}}, \ ..., \ p_{t(\mathbf{v})}^{\mathbf{v}}) \quad \text{where} \quad p_k^{\mathbf{v}} = \sum_{a_{ij} \in \mathcal{L}_k^{\mathbf{v}}(A)} a_{ij}.$$

Actually, $P_{\mathbf{v}}$ and $p_k^{\mathbf{v}}$ should be denoted by $P_{\mathbf{v}}(A)$ and $p_k^{\mathbf{v}}(A)$, but we will use this latter notation only in the case of ambiguity, when more than one matrix is considered. Here, we considered $\mathbf{v}_1 \equiv (1, 0)$, $\mathbf{v}_2 \equiv (0, 1)$, $\mathbf{v}_3 \equiv (1, 1)$, $\mathbf{v}_4 \equiv (1, -1)$.

Recently, we described a memetic algorithm (indicated as *serial memetic tomography reconstructor* or *SMTR*) for discrete tomography reconstruction [18]. Given a small set of projections $\hat{P}_{\mathbf{v}_i}$, it creates an initial population of agents by using their corresponding network flows, computed through $\mathbf{v}_1$ and $\mathbf{v}_2$. The method is shown in figure 1 and the used operators are described in the following.

2.1 Initial Population

A convenient representation of I, introduced in [9], is given by the network flow G with one source S, one sink T and two layers of nodes between S and T: the

former, named *row-nodes* $\{R_1, R_2, \ ..., \ R_m\}$, are related to the row projections while the latter, named *column-nodes* $\{C_1, C_2, \ ..., \ C_n\}$, are related to the column projections. To compute G, we consider only $\mathbf{v}_1$ and $\mathbf{v}_2$ and it is noteworthy that the maximal flow corresponds to a binary image that satisfies $\hat{P}_{\mathbf{v}_1}$ and $\hat{P}_{\mathbf{v}_2}$. Each arc $\widehat{XY}$ of G has capacity:

$$c_{\widehat{SR_i}} = \hat{p}_i^{\mathbf{v}_1}, \ c_{\widehat{R_iC_j}} = 1, \ c_{\widehat{C_jT}} = \hat{p}_j^{\mathbf{v}_2}$$

and flow:

$$f_{\widehat{SR_i}} = p_i^{\mathbf{v}_1}, \ f_{\widehat{R_iC_j}} = a_{ij}, \ f_{\widehat{C_jT}} = p_j^{\mathbf{v}_2}$$

where $i = 1, \ ..., \ m$ and $j = 1, \ ..., \ n$. This population is created outside $SMTR$ and it is passed as input to the reconstruction algorithm.

2.2 Fitness Function

The fitness function of an agent, represented by the matrix A, is defined as:

$$\mathcal{F}(A) = \sum_{i=1}^{4} \sum_{k=1}^{t(\mathbf{v}_i)} \left| p_k^{\mathbf{v}_i} - \hat{p}_k^{\mathbf{v}_i} \right|$$

where $P_{\mathbf{v}_i}$ is the actual projection of A and $\hat{P}_{\mathbf{v}_i}$ is the input projection, both taken along direction $\mathbf{v}_i$. Starting from $\hat{P}_{\mathbf{v}}$, the goal of the algorithm is to reconstruct a binary image with projections $P_{\mathbf{v}}$ equal to $\hat{P}_{\mathbf{v}}$. Therefore, $\mathcal{F}$ must be minimized.

2.3 Crossover Operator

We define here the vertical crossover, while the analogous horizontal version just operates on the transposed matrix. By swapping homologous columns of two parents A_1 and A_2, it creates two offspring B_1 and B_2. Formally, columns $A_1^{(j)}$ and $A_2^{(j)}$, with $j = 1, ..., n$, are located by a mask $M = (M_1, M_2, ..., M_n)$ of random binary values:

$$B_1^{(j)} = \begin{cases} A_1^{(j)} \text{ if } M_j = 1 \\ \\ A_2^{(j)} \text{ if } M_j = 0 \end{cases} \qquad B_2^{(j)} = \begin{cases} A_2^{(j)} \text{ if } M_j = 1 \\ \\ A_1^{(j)} \text{ if } M_j = 0 \end{cases}$$

2.4 Mutation Operator

Our mutation operator modifies no more than 5% of the pixels. This threshold was set experimentally to obtain better fitness values. In particular, this genetic operator swaps $\rho = min\{\lfloor 0.05 \times m \times n \rfloor, m \times n - w, w\}$ white and black pixels.

2.5 Switch Operator

Images that satisfy the same set of two projections can be transformed among themselves by a finite sequence of elementary switches [20]. This operator swaps a_{ij} with a_{ik} and a_{hj} with a_{hk}, where:

$$\begin{cases} a_{ij} = a_{hk} = 1 \\ a_{ik} = a_{hj} = 0 \end{cases} \quad \text{or} \quad \begin{cases} a_{ij} = a_{hk} = 0 \\ a_{ik} = a_{hj} = 1 \end{cases}$$

Usually, $P_{\mathbf{v}_3}$ and $P_{\mathbf{v}_4}$ change their values, while $P_{\mathbf{v}_1}$ and $P_{\mathbf{v}_2}$ are maintained.

2.6 Compactness Operator

Some *isolated* pixels (i.e. surrounded by 8 pixels with opposite color) are usually present and worsen the final image. This operator eliminates as many as possible isolated pixels. This approach does not guarantee a better fitness value, but nonetheless tends to let the image satisfy the input projections.

3 Parallel Implementation

We call $PMTR$ the combination of several instances of $SMTR$ by following the *island model* strategy [17]. This scheme requires the distribution of the entire population of agents into sub-populations called *demes*. Demes are well separated and their agents compete only within the same deme. Let us consider each deme of equal size na/np, where na is the total population size and np is the total number of processes. A deme d_i is assigned to the $i-th$ process, which locally evolves d_i by using the algorithm $SMTR$.

Every mi generations, some agents migrate into different demes. Such a migration is subjected to the *migration rate* (mr) that controls how many agents actually move. This event is considered general since it involves all processes. Each deme i broadcasts its mr fittest agents to all other demes d_j (i.e. $i \neq j$). Subsequently, in order to maintain the size of each population, only mr best agents among all received ones are mixed together with the retained agents. This approach selects only the fittest agents during the whole process, since demes coming from different initial populations share their best agents. A sketch of $PMTR$ is shown in figure 2.

Of course, this parallel implementation is faster than the corresponding serial reconstruction algorithm, since it does not require too many communications among processes and takes advantage of the internal structure of new multicore CPUs, which further reduce the communication time. Indeed, we use a few islands of agents (that is, just a few processing units) while these architectures do not rely on custom communication hardware (e.g. InfiniBand) to connect their cores in a very efficient fashion. According to our benchmarks, the computational time decreases almost linearly when we increment the number of processes. Needless to say, no more than one real process should run within each core to avoid excessive overload.

```
program SMTR(P, ng)
  repeat
    choose pairs of random agents;
    for all chosen pairs
      apply both vertical crossover and horizontal crossover;
      if the fitness was improved then
        replace parents with fittest offspring;
    for all agents
      apply a random elementary switch;
      if the fitness was improved then
        replace the agent with its switched version;
    for all agents
      apply mutation;
      if the fitness was improved then
        replace the agent with its mutated version;
    for all agents
      apply a random elementary switch;
      if the fitness was improved then
        replace the agent with its switched version;
    for all agents
      apply compactness;
      if the fitness was improved then
        replace the agent with its compacted version;
    for all agents
      apply a random elementary switch;
      if the fitness was improved then
        replace the agent with its switched version;
  until a solution was found or ng generations were considered;
  return last population P.
```

Fig. 1. Our serial memetic method ($SMTR$)

```
program PMTR(P, ng, na, np, mr, mi)
  split P into np sub-populations P_i with na/np elements;
  repeat
    for all P_i = SMTR(P_i, mi)
      move mr fittest agents of P_i into P_{j≠i};
      mix mr best received agents with the retained ones;
  until a solution was found or ng generations were considered;
  return last population P = ⋃ P_i.
```

Fig. 2. Our new parallel memetic method ($PMTR$)

4 Results

In real cases only the projections of the input image I will be available. We tested the algorithm on 50 hv-convex polyominoes and on 50 generic images that do not present any particular geometric property [18], with 100×100 pixels in both cases (examples are in figure 3). To evaluate the reconstruction quality of our algorithm

Fig. 3. An hv-convex polyomino and a generic image (*left*) of our database. Their reconstructed versions (*right*) have errors equal to $\varepsilon = 0.36\%$ and $\varepsilon = 4.87\%$, respectively.

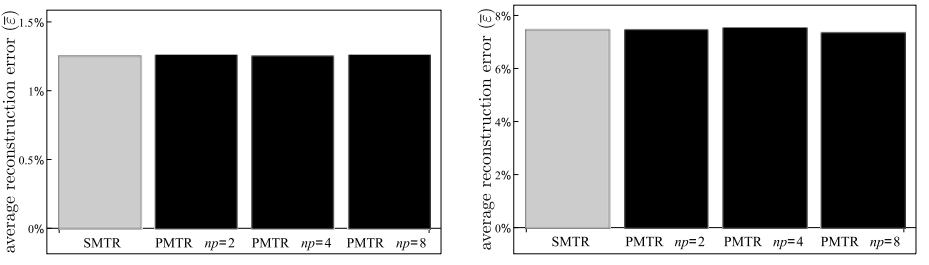


Fig. 4. Average reconstruction errors for $SMTR$ and $PMTR$ ($np = 2, 4, 8$) on hv-convex polyominoes (*left*) and generic images (*right*)

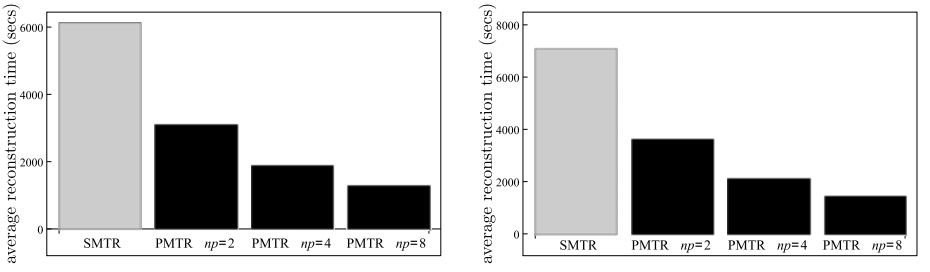


Fig. 5. Average elapsed time (in seconds) for $SMTR$ and $PMTR$ ($np = 2, 4, 8$) on hv-convex polyominoes (*left*) and generic images (*right*)

we computed the value ϵ defined as the L_1 distance between I and its reconstruction I', normalized by the size of I. The average error $\bar{\varepsilon}$ was computed on the overall database. In the experimental sessions we used $mi = \{25, 50, 75, 100\}$ and $mr = \{\frac{na}{np}, \frac{2 \cdot na}{np}\}$, and results show that the $PMTR$ is invariant to such parameters (see figure 6 for the case $np = 4$ and $na = 1200$). Figures 4–5 summarize reconstruction results obtained by both $SMTR$ ($ng = 500$, $na = 1000$) and $PMTR$ with $mi = 25$, $mr = \frac{na}{np}$, $ng = 500$, $np = 2, 4, 8$ and $na = 1024, 1200, 1600$, respectively. In the case of both hv-convex polyominoes and generic images, the island strategy does not enhance the final reconstruction, since $PMTR$ reports an average reconstruction error similar to $SMTR$ (see figure 4). Nevertheless, the new approach is significantly faster than the serial version (see figure 5). Note that reported computational times include also the process of generating the initial population $\mathcal{P}$ (see section 2.1).

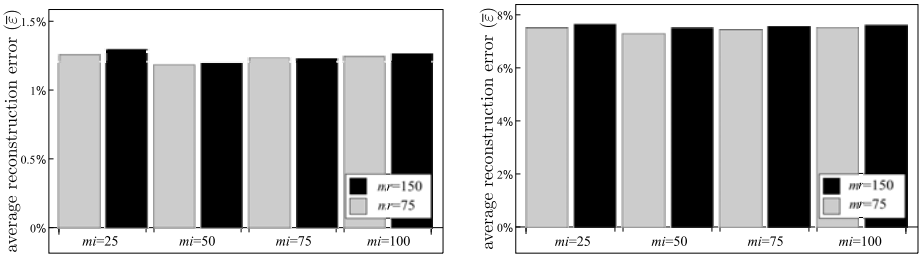


Fig. 6. Average reconstruction errors for $PMTR$ ($np = 4$, with different mr and mi values) on hv-convex polyominoes (*left*) and generic images (*right*)

The algorithm was implemented in MATLAB® and the parallelization employs the Parallel Computing Toolbox®. Experiments were carried out on just one blade of an high performance system, equipped with two quad-core Xeon E5420 (2.5 GHz, 12 MB L2 cache), 16 GB of RAM and the Linux OS. We used MATLAB co-distributed arrays to compute the min, max and average values of the fitness on the distributed populations.

5 Conclusions and Further Works

Discrete tomography is a hard task to solve and requires a lot of time to reconstruct images if no model is assumed a priori. We introduced here a combination of several instances of a memetic algorithm for discrete tomography reconstruction, already presented by us in [18], which respects the so called island model strategy. Experiments shown that $PMTR$ is effective as well as $SMTR$ but the benefit of this combination paradigm is mainly in terms of computational time. We developed the whole algorithm in the interpreted language MATLAB, but faster implementations in low level languages (e.g. compiled languages MPI and CUDA) should be able to exploit better the architectures of new common multicore CPUs. The underlying idea consists in confining all communications needed by our method within a single die, due to the fact that a single core handles an entire island of agents. Vice versa, a different MATLAB session is needed for each island, thus taking up a significant amount of memory. We studied our method on hv-convex polyominoes and generic images, but we aim to generalize it to take into account also specific models of images.

References

1. Kak, A.C., Slaney, M.: Principles of Computerized Tomography Imaging. Society for Industrial Mathematics (2001)
2. Herman, G.T., Kuba, A. (eds.): Discrete Tomography: Foundations, Algorithms, and Applications. Birkhäuser, Basel (1999)
3. Gale, D.: A theorem on flows in networks. Pacific Journal of Mathematics 7, 1073–1082 (1957)

4. Ryser, H.J.: Combinatorial properties of matrices of zeros and ones. Canadian Journal of Mathematics 9, 371–377 (1957)
5. Gardner, R.J., Gritzmann, P., Prangenberg, D.: On the computational complexity of reconstructing lattice sets from their X-rays. Discrete Mathematics 202, 45–71 (1999)
6. Wang, B., Zhang, F.: On the precise number of (0,1)-matrices in U(R,S). Discrete Mathematics 187, 211–220 (1998)
7. Frosini, A., Nivat, M., Vuillon, L.: An introductive analysis of periodical discrete sets from a tomographical point of view. Theoretical Computer Science 347(1–2), 370–392 (2005)
8. Svalbe, I., van der Spek, D.: Reconstruction of tomographic images using analog projections and the digital Radon transform. Linear Algebra and its Applications 339, 125–145 (2001)
9. Anstee, R.P.: The network flows approach for matrices with given row and column sums. Discrete Mathematics 44, 125–138 (1983)
10. Balázs, P., Balogh, E., Kuba, A.: Reconstruction of 8-connected but not 4-connected hv-convex discrete sets. Discrete Applied Mathematics 147, 149–168 (2005)
11. Herman, G.T., Kuba, A.: Discrete Tomography: Foundations, Algorithms, and Applications. In: Binary Tomography Using Gibbs Priors, pp. 191–212. Birkhäuser, Basel (1999)
12. Valenti, C.: A genetic algorithm for discrete tomography reconstruction. Genetic Programming and Evolvable Machines 9, 85–96 (2008)
13. Batenburg, J.K.: An evolutionary algorithm for discrete tomography. Discrete Applied Mathematics 151, 36–54 (2005)
14. Balázs, P., Gara, M.: An Evolutionary Approach for Object-Based Image Reconstruction Using Learnt Priors. In: Salberg, A.-B., Hardeberg, J.Y., Jenssen, R. (eds.) SCIA 2009. LNCS, vol. 5575, pp. 520–529. Springer, Heidelberg (2009)
15. Moscato, P.: On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Caltech Concurrent Computation Program, C3P Report 826 (1989)
16. Corne, D., Dorigo, M., Glover, F.: New ideas in optimization. McGraw-Hill, New York (1999)
17. Eklund, S.E.: A massively parallel architecture for distributed genetic algorithms. Parallel Computing 30(5–6), 647–676 (2004)
18. Di Gesù, V., Lo Bosco, G., Millonzi, F., Valenti, C.: A memetic approach to discrete tomography from noisy projections. Pattern Recognition 43(9), 3073–3082 (2010)
19. Isgró, F., Tegolo, D.: A distributed genetic algorithm for restoration of vertical line scratches. Parallel Computing 34(12), 727–734 (2008)
20. Ryser, H.J.: Combinatorial mathematics. The carus mathematical monographs. Ch.6, (14). MAA (1963)

An Intelligent Model for Self-compensation and Self-validation of Sensor Measurements

Javier E. Reyes Sanchez, Marley M.B.R. Vellasco, and Ricardo Tanscheit

Dept. of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro
R. Marques de Sao Vicente, 225 – 22.451-900 Rio de Janeiro, RJ – Brazil
{jreyes,marley,ricardo}@ele.puc-rio.br

Abstract. This article presents a hybrid system for self-compensation and self-validation of intelligent industrial instruments that combines a Neuro-Fuzzy model, based on the ANFIS architecture, capable of compensating errors caused by non-calibrated instruments, and a validation model based on Fuzzy Logic that provides the level of confidence of measurements. The proposed system indicates to the specialist when a new calibration must be performed. The hybrid system is tested with a differential pressure instrument, used in mining for level and pressure controls.

Keywords: Sensors, Self-calibration, Self-validation, Fuzzy Inference System, Neuro-Fuzzy systems.

1 Introduction

In several industrial processes, measurements provided by sensors are continuously monitored to obtain an adequate result. Therefore, verifying and validating the measuring process is essential for performance optimization and cost-reduction.

Regardless of its quality, an instrument degrades with time, so that the signal provided for monitoring may present other values than the expected ones. Wrong measurements in industrial processes produce economic losses and, in many cases, accidents due to control based on inaccurate feedback signals. In most cases, it is necessary to remove the measuring instrument from the field and then perform the calibration in a laboratory.

Despite the growing demand for high precision, low cost, compact size sensors, many investigations are yet to be made. The current trend towards intelligent instruments [1] is to integrate: (1) a sensor element that can be developed in a standard process; (2) electronic circuits that may periodically compensate the signal obtained; and (3) circuits to generate compatibility with the data bus, providing the status of the instrument.

The purpose of this study is to develop a model that is capable of monitoring instruments in industrial processes and produce correct values whenever they are assessed as decalibrated. The model presented here is easy to develop and

A. Petrosino et al. (Eds.): WILF 2011, LNAI 6857, pp. 269–276, 2011.

provides results that are consistent with large data sets. It consists of a Self-compensation model developed on an ANFIS (Adaptive Network based Fuzzy Inference System) [2] architecture, to compensate values measured by decalibrated sensors, and of a Self-validation model, implemented as Fuzzy Inference System, to assess the degree of reliability of the results provided by the ANFIS model and to indicate when a new calibration must be performed.

2 Self-compensation Model

Most applications for compensation of measurements are based on microprocessors incorporated to instruments, where compensation is usually performed by a software implementation. In the model proposed here (Fig. 1), compensation is performed by a neuro-fuzzy model (NFM). It is a nonlinear type of compensation and aims at correcting measurements that change due to sensor wear.

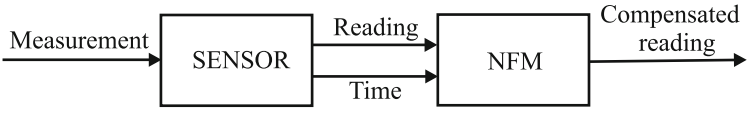


Fig. 1. Self-compensation scheme

The use of this model allows the sensor to remain active in the process for a longer period, avoiding its removal for maintenance. The NFM receives measurements from a field sensor and an information on its time in operation, generating an output that is equivalent to that of a calibrated sensor [3]. The ANFIS architecture was selected here because of its ease of implementation and fast training time, as opposed to that of a Multilayer Perceptron used previously. There is no pre or post scaling of the data, which is normally required for training a neural network, and it also requires less data for parameter tuning. As shown in Fig. 2, the neuro-fuzzy model adjusts its parameters in view of the error generated by the difference between measurements obtained by a standard sensor and the network compensated reading, aiming at producing a final output that is equal to that of a calibrated sensor.

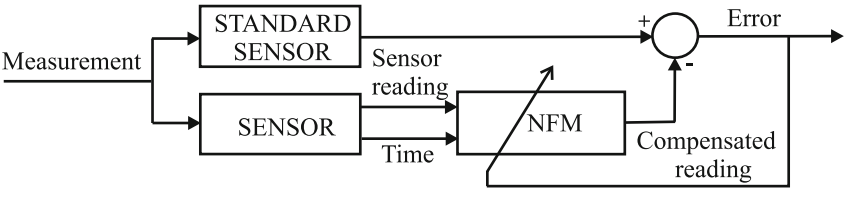


Fig. 2. Self-compensation: training

3 Self-validation Model

The self-validation model is a generic one and can be used to improve monitoring of different sensors in industrial processes. Here, the only information used to establish reliability of data for a particular sensor is its behavior in time. Neural net-based implementations, such as reported in [3], [4], [5], require a considerable amount of data for correct training, which, in this particular case, is not always possible.

The proposed model to establish reliability of data measured by the instrument is based on a Fuzzy Inference System (FIS) of the Mamdani type. Self-validation of data generated by the sensor is achieved through the difference between the compensated signal and the signal measured by the sensor, thus generating a final error that may vary in time. The error and the rate of error change (change-in-error) are the inputs to the system, as shown in Fig. 3.

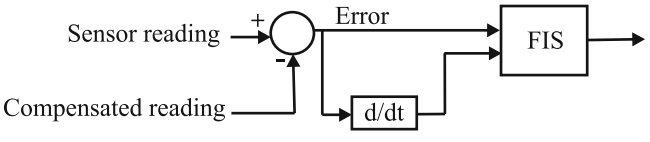


Fig. 3. Self-validation model

Three fuzzy sets, defined by the membership functions shown in Fig. 4, are associated to the chosen variables.

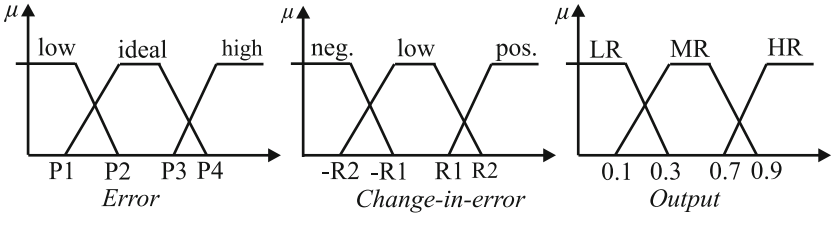


Fig. 4. Membership functions for input and output fuzzy sets

Parameters P1, P2, P3, P4, R1, and R2 must be specified for each sensor. Regarding the *error*, data that fall into the ideal region are highly reliable, whereas data outside this region present low reliability. With respect to the *change-in-error*, high rates (negative or positive) are considered undesirable. The system's final output, computed through Center of Gravity defuzzification, estimates the degree of reliability of measurements. This output is classified in the intervals low (LR), medium (MR), and high reliability (HR). The set of rules is shown in Table 1.

Table 1. Rules for the self-validation FIS

		Change-in-error		
		Negative	**Low**	**Positive**
Error	**Low**	MR	LR	LR
	Ideal	MR	HR	MR
	High	LR	MR	MR

4 Experiments with a Differential Pressure Sensor

This sections considers the application of the proposed Self-compensation and Self-validation models to a differential pressure sensor. Tests were conducted upon a set of curves that model the variation in time of calibration curves [3] of a Rosemount 3051S differential pressure sensor [6]. This sensor has a linear behavior, to ensure damping and a stable signal, when the input approaches zero. It may be scaled to various pressure ranges, according to the adjustment of zero (offset) and gain (span). Its output is a direct current in the range [4, 20] mA. In a real application, data bases with measurements for level and pressure control have been used.

4.1 Self-compensation

The purpose of compensation is to anticipate the output value corresponding to the measurement by a calibrated sensor. In order to represent the values of a decalibrated sensor, 36 curves (three years of operation) based on the equations for a standard curve were generated. Each curve represent a signal that is decompensated. Several simulations were made to obtain an adequate data base for training an ANFIS model with five Gaussian membership functions. Eventually, it consisted of 50 samples for each curve and of the time corresponding to the measurements presented to the ANFIS model, totaling 1,800 data. The data base was divided into two sets: 80% for training (assuring coverage of the whole sensor's operation range) and 20% for testing. Both the training and the evaluation vectors are of the form $[t\ vd\ vp]^T$ where t represents the time in operation, vd represents the measurements by the decalibrated sensor, and vp is the desired output value of the measurements given by the standard sensor (Fig. 5).

The inputs to the Self-compensation model that presented decalibration were compensated and became equal to the values that would have been returned by a calibrated sensor. The mean square errors (MSE) for training and testing were 7.53×10^{-5} and 5.3×10^{-5}, respectively.

The behavior of the model is shown in Fig. 6. The upper, smoother curve corresponds to the calibrated sensor, whereas the lower one corresponds to the response of the decalibrated sensor in six months. The circles on top of the upper curve represent the compensated response.

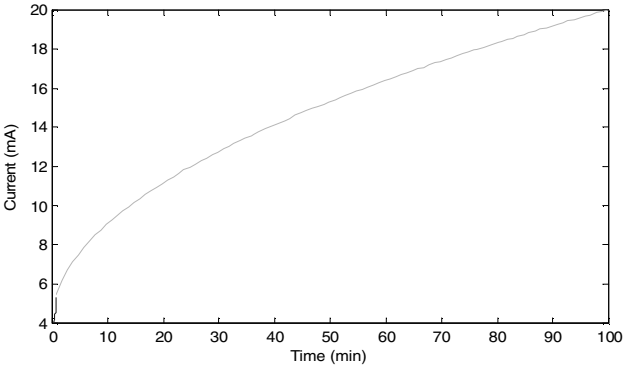


Fig. 5. Standard curve for Rosemount Sensor

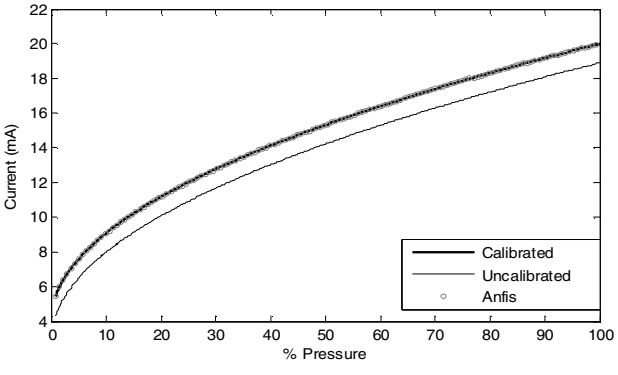


Fig. 6. Self-compensation curves for the Rosemount Sensor

To evaluate the model's performance in a real application, two data bases obtained by the SCADA (Supervisory Control and Data Acquisition) system for the Barrick mine in Peru have been used. The first one refers to level control for the Barren Tank, which stores gold, silver and a cyanide solution. The second one refers to pressure control for the Prensa filter, which stores a gold and silver solution. Both contain information on the number of samples, the operation time in minutes, and the measurements during one week of operation. The number of samples stored for both cases is 1009. Table 2 shows the parameters.

Table 2. Database parameters

Control	Average	Std.Dev. (%)	Zero	Span
Level	75.35 mBar	8.28	0	100mBar
Pressure	368.47 kPa	112.9	0	100 kPa

The data were transformed into current values in the range of 4 to 20 mA. Based on the curves simulated before, a data base for each month was created. In the results for the Barren Tank (Figs. 7a and 7b) and for the Prensa filter (Figs. 8a and 8b), the continuous upper curve represents the calibrated sensor response for the original data base, the lower curve represents decalibrated sensor data and the black circles represent the compensated values. In both cases, the NFM responds very well, with an MSE in the magnitude of 10^{-6}.

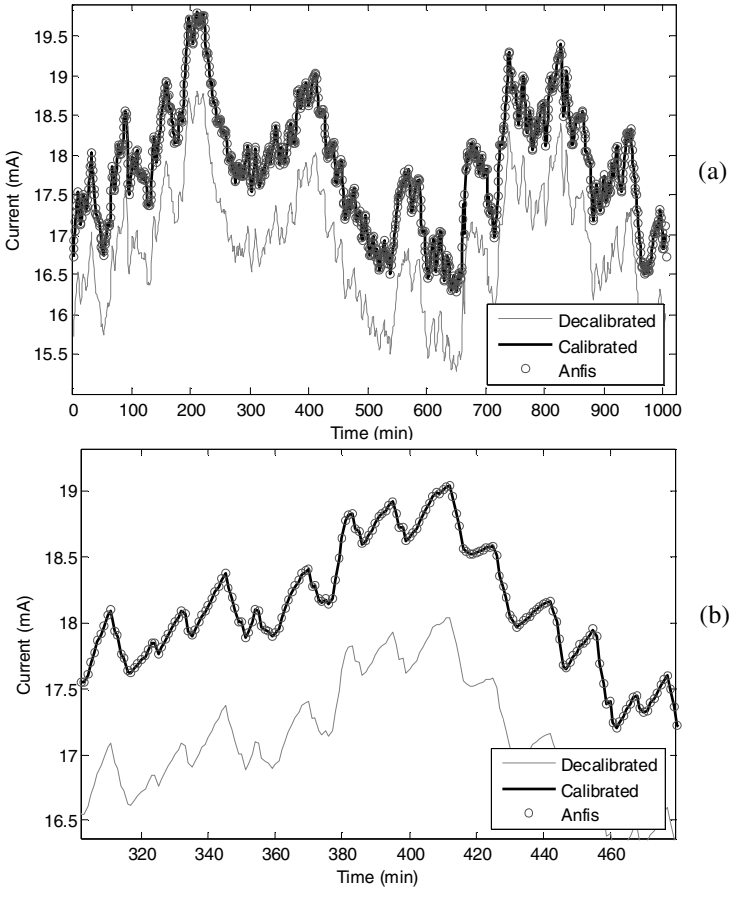


Fig. 7. Self-compensation: Barren Tank – (b) enhanced version of (a)

4.2 Self-validation

In the generic self-validation system, based on [7], the fuzzy inference system is expected to provide an appropriate output given the error and change-in-error input values.

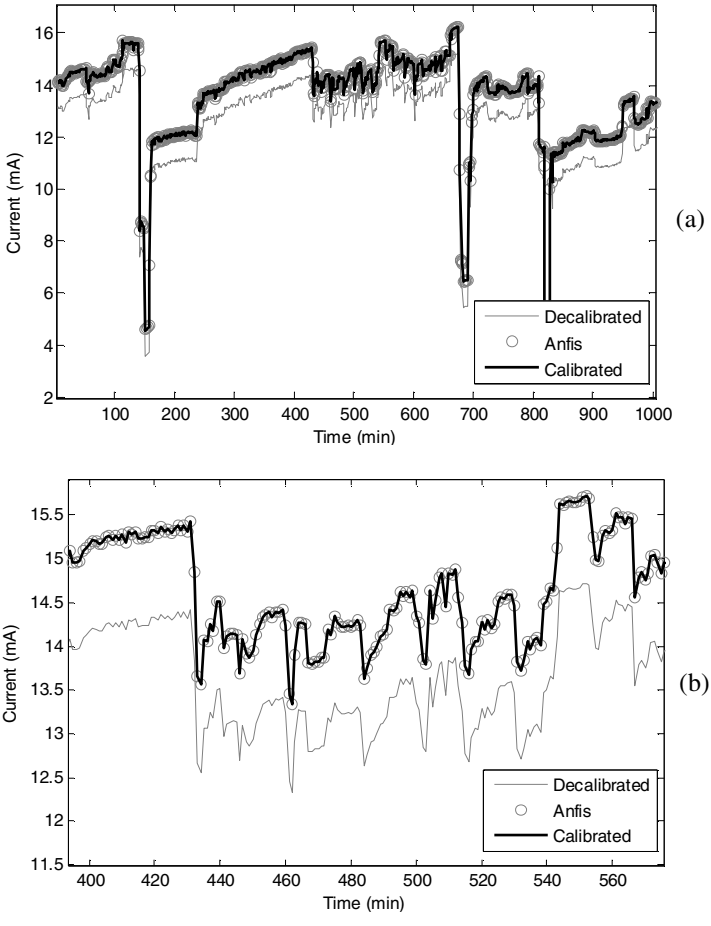


Fig. 8. Self-compensation: Prensa Filter – (b) enhanced version of (a)

Fuzzy sets parameters for both cases were chosen as:

[P1 P2 P3 P4 R1 R2] = [-0.4 -0.1 0.1 0.4 0.3 0.6]

Values for P2 and P3, for example, were defined by assuming that an error of 0.1 (mBar or kPa) would be acceptable. Values for R1 and R2 were chosen to be greater than the standard deviation of the error. If they are smaller, the self-validation model generates an oscillatory response (for reliability), which is of difficult interpretation. For level control (Barren tank), the response after defuzzification was a constant value of 0.8, indicating high reliability. From the response for the Prensa filter (Fig. 9), it can be observed that sensor reliability for pressure control is low at some instants, as a result of a large derivative. However, the return of the signal to its right value after some time indicates that the sensor maintains its initial degree of reliability.

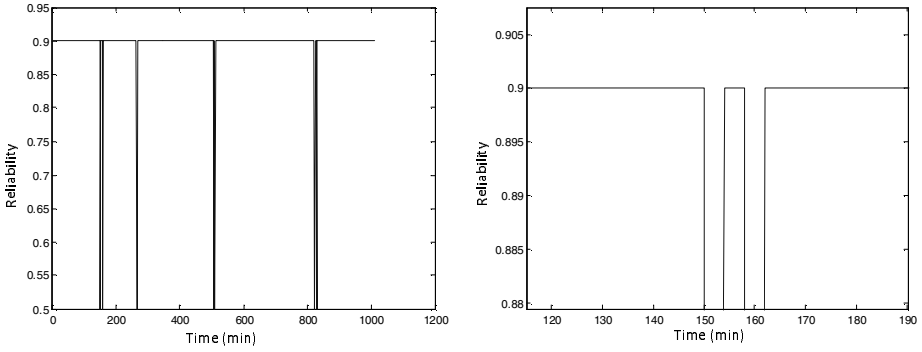


Fig. 9. Self-validation: Prensa filter

5 Conclusions

In this paper, a methodology was developed for self-compensation and self-validation of sensor measurements. Self-compensation is performed by a Neuro-Fuzzy model, which corrects measurements errors in a decalibrated sensor. Self-validation is performed by a fuzzy inference system that gives as its output the reliability of the measured signal. Although the tests were performed for a specific case study, the methodology is generic and can be used for any application. The compensation and validation systems are applicable to data generated by independent sensors. If a sensor presents a fault, it is possible to identify when a new calibration must be performed. Future work involves developing a model that is capable of self-compensating and self-validating measurements with errors resulting from deviations, off-sets and noise.

References

1. Tian, G.Y., Zhao, Z.X., Baines, R.W.: A Fieldbus-based Intelligent Sensor. Mechatronics 10, 835–849 (1999)
2. Jang, J.S.R.: ANFIS: Adaptive-Network-based Fuzzy Inference System. IEEE Transactions on Systems, Man & Cybernetics 23(3), 665–685 (1993)
3. Pereira, D., Bezerra, J., Duarte Neto, A., Melo, J.: Instrumentacao Inteligente aplicada ao Processo de Medicao de Vazao e BSW (in portuguese). In: XV Brazilian Congress on Automation, Brazil (2004)
4. Mahamoud, W.: Hardware Implementation of Automated Sensor Self-validation System for Cupola Furnaces. In: 9th IEEE Symposium on Field-Programmable Custom Computing Machines, Napa, USA (2001)
5. Abdelrahman, M., Subramanian, S.: An Intelligent Signal Validation System for a Cupola furnace - P. 1: Methodology. In: American Control Conf., San Diego (1999)
6. Emerson Reference Manual-Rosemount 3051S Series. Scalable Pressure, Flow, and Level Solutions (2009)
7. PhaniShankar, C.V., Orth, S., Frolik, J., Abdelrahman, M.: Fuzzy Rules for Automated Sensor Self-Validation and Confidence Measure. In: American Control Conference, Chicago, USA (2000)

Author Index