STUDIES IN *FUZZINESS* AND *SOFT COMPUTING*

Humberto Bustince
Francisco Herrera
Javier Montero
Editors

# Fuzzy Sets and Their Extensions: Representation, Aggregation and Models

Springer

Fuzzy Sets and Their Extensions: Representation, Aggregation and Models

Studies in Fuzziness and Soft Computing, Volume 220

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

---

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

Vol. 206. A. Sengupta (Ed.)
*Chaos, Nonlinearity, Complexity,* 2006
ISBN 978-3-540-31756-2

Vol. 207. Isabelle Guyon, Steve Gunn, Masoud Nikravesh, Lotfi A. Zadeh (Eds.)
*Feature Extraction,* 2006
ISBN 978-3-540-35487-1

Vol. 208. Oscar Castillo, Patricia Melin, Janusz Kacprzyk, Witold Pedrycz (Eds.)
*Hybrid Intelligent Systems,* 2007
ISBN 978-3-540-37419-0

Vol. 209. Alexander Mehler, Reinhard Köhler
*Aspects of Automatic Text Analysis,* 2007
ISBN 978-3-540-37520-3

Vol. 210. Mike Nachtegael, Dietrich Van der Weken, Etienne E. Kerre, Wilfried Philips (Eds.)
*Soft Computing in Image Processing,* 2007
ISBN 978-3-540-38232-4

Vol. 211. Alexander Gegov
*Complexity Management in Fuzzy Systems,* 2007
ISBN 978-3-540-38883-8

Vol. 212. Elisabeth Rakus-Andersson
*Fuzzy and Rough Techniques in Medical Diagnosis and Medication,* 2007
ISBN 978-3-540-49707-3

Vol. 213. Peter Lucas, José A. Gámez, Antonio Salmerón (Eds.)
*Advances in Probabilistic Graphical Models,* 2007
ISBN 978-3-540-68994-2

Vol. 214. Irina Georgescu
*Fuzzy Choice Functions,*2007
ISBN 978-3-540-68997-3

Vol. 215. Paul P. Wang, Da Ruan, Etienne E. Kerre (Eds.)
*Fuzzy Logic*, 2007
ISBN 978-3-540-71257-2

Vol. 216. Rudolf Seising
*The Fuzzification of Systems*, 2007
ISBN 978-3-540-71794-2

Vol. 217. Masoud Nikravesh, Janusz Kacprzyk, Lofti A. Zadeh (Eds.)
*Forging the New Frontiers: Fuzzy Pioneers I*, 2007
ISBN 978-3-540-73181-8

Vol. 218. Masoud Nikravesh, Janusz Kacprzyk, Lofti A. Zadeh (Eds.)
*Forging the New Frontiers: Fuzzy Pioneers II*, 2007
ISBN 978-3-540-73184-9

Vol. 219. Roland R. Yager, Liping Liu (Eds.)
*Classic Works of the Dempster-Shafer Theory of Belief Functions*, 2007
ISBN 978-3-540-25381-5

Vol. 220. Humberto Bustince, Francisco Herrera, Javier Montero (Eds.)
*Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision, 2008
ISBN 978-3-540-73722-3

Humberto Bustince · Francisco Herrera ·
Javier Montero
Editors

# Fuzzy Sets and Their Extensions: Representation, Aggregation and Models

## Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision

With 126 Figures and 44 Tables

Foreword by Didier Dubois

Springer

Humberto Bustince
Department of Automatics and Computation
Universidad Pública de Navarra
Campus Arrosadia s/n
31006 Pamplona
SPAIN
Email: bustince@unavarra.es

Francisco Herrera
Department of Computer Science
and Artificial Intelligence (DECSAI)
University of Granada
Periodista Daniel Saucedo Aranda s/n
18071 Granada
SPAIN
Email: herrera@decsai.ugr.es

Javier Montero
Faculty of Mathematics
Universidad Complutense
Plaza de las Ciencias 3
28040 Madrid
SPAIN
Email: monty@mat.ucm.es

Library of Congress Control Number: 2007932776

ISSN print edition: 1434-9922
ISSN electronic edition: 1860-0808
ISBN 978-3-540-73722-3 Springer Berlin Heidelberg New York



Springer is a part of Springer Science+Business Media
springer.com


Typesetting: Integra Software Services Pvt. Ltd., India
Cover design: WMX Design, Heidelberg

Printed on acid-free paper SPIN: 11679158 89/3180/Integra 5 4 3 2 1 0

# Foreword

## Fuzzy Sets and Their Extensions: Representation, Aggregation and Models

### *Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision*

Fuzzy sets are now more than 40 years old, and have come of age. However, the development of fuzzy set theory at the theoretical level, and its successful applications to science and technology have often run in isolation. Only a little part of the theoretical apparatus was effectively used in past applications. The most prominent ones, namely fuzzy rule-based modeling and control engineering, were directly inspired from a seminal paper by Lotfi Zadeh in 1973, suggesting how to use expert knowledge for synthetizing control laws, and from the first experiments published by Abe Mamdani. Later in the eighties, when spectacular applications were blossoming in Japan, fuzzy rule-based systems were systematized and simplified by Michio Sugeno and colleagues, and became a basic approach to non-linear system modeling and control, soon hybridized with neural networks in the nineties. Thus fuzzy systems significantly contributed to the raise of computational intelligence, and a lot of learning techniques for the construction of (supposedly interpretable) fuzzy models from data were developed under the flag of soft computing.

Even if this area was quite successful, it is patent that the role, in the success of fuzzy logic, of new fuzzy set-related concepts developed quite at the same time in the mathematical niche of the fuzzy set community was limited. To wit, the notion of fuzzy rule is now much better understood, but fuzzy extensions of material implications have not been popular in systems engineering so far. Other notions, like aggregation operations, fuzzy intervals, fuzzy preference relations, possibility measures have also acquired strong theoretical underpinnings in the meantime. The end of the nineties have also witnessed tremendous progress in the foundations of formal fuzzy logic, due to the impulse given by reputed logicians like Petr Hajek and Daniele Mundici. Yet they only had a limited impact on domains like databases and information retrieval, risk analysis, artificial intelligence, mathematical programming, and multiple criteria decision-making, where their use looks natural. One reason is the lack of communication between mathematics and engineering in many cases.

From scanning its contents, this book looks like an outline of what could be the next generation of fuzzy logic applications. Strikingly enough it does not deal with systems engineering nor soft computing. Instead, new exciting topics are proposed like data mining and web intelligence, as well as more traditional ones like decision-making and computer vision, for which fuzzy set methods are available for a long time, but the impact of fuzzy sets was clearly marginal so far. With the tremendous gain in maturity of fuzzy set mathematics, observed in the last twenty years, no doubts that new generations of researchers, trained with the new results, will propose new techniques for such application areas as appearing in this book. The extensive series of chapters in this book, devoted to aggregation operators, their axiomatization, their construction and their identification from data is very symptomatic of the renewal of research topics in fuzzy set theory.

The other part of the theoretical section of this book raises some questions. Several chapters are typical of the new trends towards hybrid constructions, namely fuzzy sets with rough sets, and fuzzy sets with random sets. The other two topics, that is fuzzy sets of higher order and Atanassov's intuitionistic fuzzy sets are more problematic at this stage of their development. Type-two fuzzy sets as well as interval-valued fuzzy sets are an old construct suggested by Zadeh already in the seventies. They are natural concepts addressing the apparent paradox, faced by standard fuzzy sets, of modeling imprecise concepts using precise membership grades. Of course, they fall under the usual objection of regression to infinity (why then should the imprecise membership grades be modeled by type 1 fuzzy sets ?). But this is not at all a valid practical objection. A first more actual difficulty is that type-2 fuzzy systems may involve many more tuning parameters than type-1 fuzzy systems. So they may not be a very concise tool. Another more basic difficulty lies in the extension of fuzzy connectives via the extension principle, studied in the seventies by Mizumoto and Tanaka (and Dubois and Prade !), that hides dependency phenomena. If an interval-valued or type-2 fuzzy set represents uncertainty about membership grades, they are uncertain type-1 fuzzy sets, not fuzzy sets with membership grades in a more complex scale than the unit interval. So for instance even if the membership grade $\mu_A(x)$ is ill-known, we nevertheless know that $\mu_A(x) + \mu^c_A(x) = 1$, where $A^c$ is the complement of $A$, hence $\min(\mu_A(x), \mu^c_A(x)) \leq 0.5$. But, if $\mu_A(x) \in$ [a,b], computing the range of the membership grade of $x$ in $A \cup A^c$ as

$$min([a, b], 1 - [a, b]) = [min(a, 1 - b), min(b, 1 - a)]$$

using the extension principle on fuzzy connectives, cannot retrieve this upper bound, as we may have min(b,1-a) $>$ 0.5. This is one more example of the lack of truth-functionality in the presence of uncertainty. This is clearly a limitation of the truth-functional calculus of interval-valued fuzzy sets (and a fortiori, type-2 fuzzy sets) one should be aware of. Interestingly, the interpretation of a fuzzy set as an ill-known (random) crisp set leads to the same loss of truth-functionality, as explained in the chapter by Lawry.

The case of intuitionistic fuzzy sets, advocated by Atanassov, in this book, has been discussed elsewhere. They suffer from a lack of connection with current intuitionistic logic (despite the fact that some ideas were borrowed by their founder from early intuitionism). Moreover, the algebraic operations proposed for them make them formally equivalent to interval-valued fuzzy sets (hence they also suffer from the difficulty pointed out in the previous paragraph). Yet, the idea of Atanassov, namely evaluating degrees of membership and degrees of non-membership in a separate way, is very important, as it refers to the phenomenon of bipolarity. Indeed, it has been observed by cognitive scientists that the human brain tends to process positive and negative aspects of information separately. More work is needed to clarify the specific role of Atanassov membership/non-membership pairs, as distinct from the older interval-valued fuzzy sets, and from what is nowadays referred to as intuitionism.

The second part of the book will be very useful to charter the most promising future applications areas for fuzzy set theory. In the case of decision-making, preference modeling will certainly benefit from the progress made in the theory of fuzzy relations, because the natural concept of gradual preference is not really encountered in classical decision theory but for mere utility functions. The success of fuzzy reference relations will depend on the development of measurement-theoretic foundations for them, and the existence of tools that demonstrate their simplicity of use and efficiency in solving issues crisp relations cannot address (like in voting theories). Another challenge in fuzzy decision theory is its capability to found linguistic approaches. Here, the difficulty is the same as in qualitative reasoning at large: finding a compromise between mathematically well-behaved but poorly expressive frameworks (as like the sign calculus), and more expressive mathematically-ill behaved settings (like absolute orders of magnitude). In particular, it is not clear that linguistic variables grounded on non-measurable numerical scales (like using [0, 1] to evaluate abstract notions like beauty) are the way to go. Several chapters offer the state of the art in linguistic decision-making.

Data mining and web intelligence are clearly new topics where fuzzy sets have a role to play. The first data-mining tools presupposed a Boolean world. Adapting them to numerical attributes is problematic if crisp partitions of attribute ranges are used. Fuzzy sets may address the issue of sensitivity to thresholds, even if this advantage needs to be properly assessed. Nevertheless the concept of fuzzy association rule and more generally fuzzy rule appears to be more flexible than its crisp counterpart. The web is also an excellent opportunity to combine several fuzzy set methods developed in the recent years: formal fuzzy logic for the description of fuzzy ontologies, fuzzy preference modeling and aggregation techniques for search engines, linguistic information processing in recommender systems and other kinds of e-services. These topics are well documented in the book.

Lastly, the selection of papers in computer vision also witnesses how fuzzy techniques have slowly but strongly entered the various tool-boxes of the field. That this area could benefit from fuzzy sets was pointed out quite early by the late A. Rosenfeld in the seventies, with continued efforts until recently. Filtering, thresholding and segmentation techniques, colour processing and pixel classification methods, among other subproblems include fuzzy ingredients as witnessed by various survey chapters in this volume.

Systems and control engineers used to be the main advertisers of fuzzy sets till the end of the twentieth century. This book suggests that the new century will see fuzzy set theory becoming an important methodology for information science and engineering at large.

March 1, 2007 Didier Dubois

# Preface

## Fuzzy Sets and Their Extensions: Representation, Aggregation and Models

### *Intelligent Systems from Decision Making to Data Mining, Web Intelligence and Computer Vision*

This book has its origins in the invited talks presented at the "Second International Workshop of Artificial Intelligence, Fuzzy Logic and Computer Vision", held in Pamplona, Spain (November 30 to December 2, 2005). These invited talks covered the different tasks that we must take into consideration for "fuzzy logic based real applications", from fuzzy logic foundations (representation and aggregation operators) to information fusion and specific model constructions. Although model construction depends on the kind of problem and presents a great variety of situations, the Workshop paid special attention to computer vision applications, and it also included some invited talks on decision-making, web intelligence and data mining.

During the conference we realized that there was scientific demand for a book offering a good state-of-the-art collection of papers together with a wide-ranging view of applications, in such a way that readers could find, in a single volume, the three most important tasks to take into consideration for fuzzy logic real applications: representation, aggregation procedures and a variety of models in different application areas, considering the different semantics for fuzzy membership functions that exist in the literature (similarity, preference and uncertainty).

The book has been conceived according to a fixed scheme, covering a wide view of past, present and future research related to this field, together with a strict selection of prestigious authors as a guarantee for quality papers, but still maintaining a standard anonymous peer-review for every paper (other researchers, also prestigious in each topic, kindly accepted to collaborate in this project). We paid attention to non-standard representations that extend fuzzy sets, aggregation procedures, and the whole process of intelligent information management using fuzzy logic, focusing on four important application areas: decision-making, data mining, web intelligence and computer vision.

The present book brings together many of the mentioned invited talks at the "Second International Workshop of Artificial Intelligence, Fuzzy Logic and Computer Vision", plus a collection of well recognized researcher contributions, with the aim of presenting an extensive background on each topic. It collects a set of papers and gives a triple perspective: papers for revision, papers prospecting these areas and papers presenting interesting novel approaches.

Once all the papers had been revised and corrected, Prof. D. Dubois kindly accepted to write the Preface. The book contains 34 chapters divided into two parts:

Part I devoted to foundation issues (Sect. 1 and Sect. 2) and Part II to the four application areas previously mentioned (Sects. 3, 4, 5 and 6).

More specifically, the first part is divided into two sections. Section 1 contains four review papers introducing some non-standard representations that extend fuzzy sets (type-2 fuzzy sets, Atanassov's IFS, fuzzy rough sets and computing with words under the fuzzy sets perspective). Section 2 contains six review and prospect papers that revise different aggregation issues from a theoretical and practical point of view. The second part is divided into four sections. Section 3 is devoted to decision-making, containing seven papers that show how fuzzy sets and their extensions are an important tool for modeling choice problems (e.g., sensory evaluation, preference representation, group decision making, consensus and voting systems). Section 4 collects eight papers that cover different aspects on the use of fuzzy sets and their extensions in data mining, classification, association rules, non-supervised classification, subgroup discovery, etc., giving an illustrative revision of the state of the art in the subject. Section 5 is devoted to the emergent topic of web intelligence and contains four papers that show the use of fuzzy set theory in certain problems that we can tackle under this heading (information retrieval, web meta-search engines, e-services and recommender systems). Section 6 is devoted to the use of fuzzy sets and their extensions in the field of computer vision, presenting how these can be a useful tool in this area (image thresholding, segmentation, fuzzy measures and color processing).

We believe that this volume presents an up-to-date state of current research in the use of fuzzy sets and their extensions in the whole process of intelligent information management. It will be useful to non-expert readers, whatever their background, who are keen to learn more about this area of research. It will also support those specialists who wish to discover the latest results as well as the latest trends in the mentioned areas.

Finally, we would like to express our most sincere gratitude to Springer-Verlag and in particular to Prof. J. Kacprzyk (editor-in-chief of the series "Studies in Fuzzyness and Soft Computing"), for having given us the opportunity to prepare the text and for having supported and encouraged us throughout its preparation. We would also like to acknowledge our gratitude to all those who have contributed to the books by producing articles that we consider to be of the highest quality. We also like to mention the somewhat obscure and altruistic, though absolutely essential, task carried out by a group of referees (all the contributions have been reviewed by two referees), who, through their comments, suggestions, and criticisms, have contributed to raising the quality of this volume.

March 1, 2007

H. Bustince
F. Herrera
J. Montero

# Contents

# Contributors

**Sergio Alonso**
Department of Computer Science and Artificial Intelligence
University of Granada, 18071 - Granada, Spain
salonso@decsai.ugr.es

**Ana del Amo**
Mission Management Dept., Smiths Aerospace, Grand Rapids, MI.,U.S.A.
Ana.Del.Amo@smiths-aerospace.com

**Krassimir T. Atanassov**
CLBME - Bulgarian Academy of Sciences P.O.Box 12, Sofia-1113, Bulgaria krat@bas.bg,
krat@clbme.bas.bg

**Bernard De Baets**
Department of Applied Mathematics, Biometrics and Process Control Ghent University, Coupure links 653, B-9000 Gent, Belgium
bernard.debaets@ugent.be

**Edurne Barrenechea**
Departamento de Automática y Computación Universidad Pública de Navarra Campus de Arrosadía, s/n, 31006 Pamplona, Spain
edurne.barrenechea@unavarra.es

**Gleb Beliakov**
School of Engineering and Information Technology, Deakin University 221 Burwood Hwy, Burwood 3125, Australia
gleb@deakin.edu.au

**Greg S. Biging**
Environmental Sciences, University of California, Berkeley, CA., U.S.A.
biging@nature.berkeley.edu

**Bernadette Bouchon-Meunier**
Université Pierre et Marie Curie-Paris 6, CNRS, UMR 7606,
LIP6 8 rue du Capitaine Scott, Paris, F-75015, France
{marie-jeanne.lesot,maria.rifqi,
bernadette.bouchon-meunier}@lip6.fr

**Humberto Bustince**
Departamento de Automática y Computación Universidad Pública de Navarra Campus de Arrosadía, s/n, 31006 Pamplona, Spain
bustince@unavarra.es

**Tomasa Calvo**
Departamento de Ciencias de la Computación,
Universidad de Alcalá 28871-Alcalá de Henares (Madrid), Spain
tomasa.calvo@uah.es

**Tamalika Chaira**
A-7, Aurobinda Nagar
Judge court Road
Midnapore(W)-721101
West Bengal, India
tchaira@yahoo.com

**QingCai Chen**
Department of Computer Science and Technology, Harbin
Institute of Technology, Harbin, China
djchen@iastate.edu

**Francisco Chiclana**
Centre for Computational Intelligence
School of Computing, De Montfort University Leicester LE1 9BH, UK
chiclana@dmu.ac.uk

**Simon Coupland**
Centre for Computational Intelligence
De Montfort University The Gateway, Leicester, LE1 9BH, United Kingdom
simonc@dmu.ac.uk

**Pedro Couto**
CETAV-University of Tras-os-Montes e Alto Douro Ap. 1014, 5001-911 Vila Real, Portugal
pamcouto@gmail.com

**Glad Deschrijver**
Fuzziness and Uncertainty Modelling Research Unit
Department of Applied Mathematics and Computer Science, Ghent University
Krijgslaan 281 (S9), B–9000 Gent, Belgium
Glad.Deschrijver@UGent.be

**Mario Fedrizzi**
Dipartimento di Informatica e Studi Aziendali, Universitá degli Studi di Trento via Inama 5, I-38100 Trento, Italy
Mario.Fedrizzi@unitn.it

**János Fodor**
John von Neumann Faculty of Informatics, Budapest Tech Bécsi út 96/b, H-1034 Budapest, Hungary
fodor@bmf.hu

**José Luis García-Lapresta**
Dep. de Economía Aplicada (Matemáticas), PRESAD
Research Group Universidad de Valladolid, Spain
*lapresta*@eco.uva.es

**Salvatore Greco**
Faculty of Economics, University of Catania, Corso Italia, 55 95129 Catania, Italy
salgreco@unict.it

**Daniel Gómez**
Faculty of Statistics, Complutense University, Madrid, Spain
dagomez@estad.ucm.es

**Pedro González**
Department of Computer Science
University of Jaén, 23071 – Jaén, Spain
pglez@ujaen.es

**Francisco Herrera**
Department of Computer Science and Artificial Intelligence University of Granada, 18071 – Granada, Spain
herrera@decsai.ugr.es

**Enrique Herrera-Viedma**
Dept. of Computer Science and Artificial Intelligence,
University of Granada 18071 - Granada, Spain
viedma@decsai.ugr.es

**Tu-Bao Ho**
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
bao@jaist.ac.jp

**Tzung-Pei Hong**
Department of Electrical Engineering, National University of Kaohsiung
tphong@nuk.edu.tw

**Eyke Hüllermeier**
University of Magdeburg, Faculty of Computer Science
Universitätsplatz 2, D-39106 Magdeburg, Germany
eyke.huellermeier@iti.cs.uni-magdeburg.de

**Van-Nam Huynh**
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
{huynh,bao,nakamori}@jaist.ac.jp

**Hisao Ishibuchi**
Osaka Prefecture University,
Sakai, Osaka 599-8531, Japan
hisaoi@cs.osakafu-u.ac.jp

**María José del Jesus**
Department of Computer Science
University of Jaén, 23071 – Jaén, Spain
mjjesus@ujaen.es

**Robert John**
Centre for Computational Intelligence
De Montfort University The Gateway, Leicester, LE1 9BH,
United Kingdom
rij@dmu.ac.uk

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

**Etienne Kerre**
Fuzziness and Uncertainty Modelling Research Unit
Department of Applied Mathematics and Computer Science, Ghent University
Krijgslaan 281 (S9), B–9000 Gent, Belgium
Etienne.Kerre@UGent.be

**Anna Kolesárová**
Institute IAM, Fac. of Chemical and Food Technology Slovak University of Technology Radlinského 9, 812 37 Bratislava, Slovakia
anna.kolesarova@stuba.sk

**Magda Komorníková**
Department of Mathematics, SvF
Slovak University of Technology
Radlinského 11, 813 68 Bratislava, Slovakia and ÚTIA AV ČR, 182 08 Prague, Czech Republik
anna.kolesarova@stuba.sk

**Jonathan Lawry**
Department of Engineering Mathematics, University of Bristol Bristol, BS8 1TR, UK
j.lawry@bris.ac.uk

**Yeong-Chyi Lee**
Institute of Information Engineering,
I-Shou University
d9003007@stmail.isu.edu.tw

**Enrique Lerma**
Pathology Dept Hospital Sant Pau
Autonomous University. Barcelona, Spain
elerma@hsp.santpau.es

**Marie-Jeanne Lesot**
Université Pierre et Marie Curie-Paris6, CNRS, UMR 7606,
LIP6 8 rue du Capitaine Scott, Paris, F-75015, France
marie-jeanne.lesot@lip6.fr

**Jun Liu**
School of Computing and Mathematics
University of Ulster at Jordanstown
Newtownabbey, Co. Antrim BT37 0QB,
Northern Ireland, UK
j.liu@ulster.ac.uk

**Bonifacio Llamazares**
Dep. de Economía Aplicada
(Matemáticas), PRESAD
Research Group Universidad de
Valladolid, Spain
*boni*@eco.uva.es

**Vincenzo Loia**
Dipartimento di Matematica e
Informatica Universitá degli Studi di
Salerno via Ponte don Melillo, 84084 -
Fisciano (SA), Italy
loia@unisa.it

**Antonio Gabriel López-Herrera**
Dept. of Computer Science, University
of Jaén. 23071 - Jaén, Spain
agabriel@ujaen.es

**Jie Lu**
Faculty of Information Technology,
University of Technology, Sydney
(UTS) PO Box 123, Broadway, NSW
2007, Australia
jielu@it.uts.edu.au

**Luis Martínez**
Computer Sciences Department
University of Jaen Campus Las
Lagunillas s/n 23071- Jaén, Spain
martin,
martin@ujaen.es

**Benedetto Matarazzo**
Faculty of Economics, University
of Catania, Corso Italia, 55 95129
Catania, Italy
matarazz@unict.it

**Pedro Melo-Pinto**
CETAV-University of Tras-os-Montes
e Alto Douro Ap. 1014, 5001-911 Vila
Real, Portugal
pmelo@utad.pt

**Radko Mesiar**
Department of Mathematics, SvF
Slovak University of Technology
Radlinského 11, 813 68 Bratislava,
Slovakia and ÚTIA AV ČR, 182 08
Prague, Czech Republik
mesiar@math.sk

**Javier Montero**
Faculty of Mathematics, Complutense
University, Madrid, Spain
monty@mat.ucm.es

**Eduard Montseny**
Computer Engineering and Industrial
Systems Dept. Technical University of
Catalonia. Barcelona, Spain
eduard.montseny@upc.edu

**Mike Nachtegael**
Ghent University Fuzziness &
Uncertainty Modelling Research Unit
Krijgslaan 281 (building S9) 9000
Ghent - Belgium
mike.nachtegael@ugent.be

**Yoshiteru Nakamori**
Japan Advanced Institute of Science
and Technology
1-1 Asahidai, Nomi, Ishikawa
923-1292, Japan
{huynh,bao,nakamori}
@jaist.ac.jp

**Yusuke Nojima**
Osaka Prefecture University, Sakai,
Osaka 599-8531, Japan
nojima@cs.osakafu-u.ac.jp

**Hannu Nurmi**
Department of Political Science,
University of Turku FIN-20014 Turku,
Finland
hannu.nurmi@utu.fi

**José A. Olivas**
Department of Technologies and Information Systems SMILe Research Group (Soft Management of Internet e-Laboratory) Escuela Superior de Informática, Universidad de Castilla - La Mancha Paseo de la Universidad 4, 13071-Ciudad Real, Spain
JoseAngel.Olivas@uclm.es

**Miguel Pagola**
Departamento de Automática y Computación Universidad Pública de Navarra Campus de Arrosadía, s/n, 31006 Pamplona, Spain
miguel.pagola@unavarra.es

**Gabriella Pasi**
Università degli Studi di Milano Bicocca Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
pasi@disco.unimib.it

**Ricardo Alberto Marques Pereira**
Dipartimento di Informatica e Studi Aziendali Università degli Studi di Trento Via Inama 5, 38100 Trento, Italy
mp@economia.unitn.it

**Luis G. Pérez**
Computer Sciences Department University of Jaen Campus Las Lagunillas s/n 23071- Jaén, Spain
lgonzaga@ujaen.es

**Carlos Porcel**
Dept. of Computing and Numerical Analysis, University of Córdoba 14071 - Córdoba, Spain
carlos.porcel@uco.es

**Maria Rifqi**
Université Pierre et Marie Curie-Paris6, CNRS, UMR 7606,
LIP6 8 rue du Capitaine Scott, Paris, F-75015, France
rifqi@lip6.fr

**Da Ruan**
Belgian Nuclear Research Centre (SCKCEN) Boeretang 200, 2400 Mol, Belgium
druan@sckcen.be

**Stefan Schulte**
Ghent University Fuzziness & Uncertainty Modelling Research Unit Krijgslaan 281 (building S9) 9000 Ghent - Belgium
Stefan.Schulte@UGent.be

**Roman Słowiński**
Institute of Computing Science, Poznań University of Technology 60-965 Poznań and Institute for Systems Research, Polish Academy of Sciences 01-447 Warsaw, Poland
roman.slowinski@cs.put.poznan.pl

**Pilar Sobrevilla**
Applied Maths II Dept. Technical University of Catalonia. Barcelona, Spain
pilar.sobrevilla@upc.edu

**Eulalia Szmidt**
Systems Research Institute, Polish Academy of Sciences ul. Newelska 6, 01–447 Warsaw, Poland
szmidt@ibspan.waw.pl

**Hamid R. Tizhoosh**
University of Waterloo Systems Design Engineering Pattern Recognition and Machine Intelligence Group Waterloo, ON, N2L 3G1, Canada
tizhoosh@uwaterloo.ca

**Eric C.C. Tsang**
Department of Computing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
csetsang@inet.polyu.edu.hk

**Vicenç Torra**
Institut d'Investigació en Intel.ligència Artificial - CSIC Campus UAB s/n, E-08193 Bellaterra (Catalunya, Spain)
vtorra@iiia.csic.es

**Mario Veniero**
Dipartimento di Matematica e Informatica Universitá degli Studi di Salerno via Ponte don Melillo, 84084 - Fisciano (SA), Italy
mveniero@unisa.it

**Xizhao Wang**
Department of Mathematics and Computer Science, Hebei University, Baoding, Hebei, 071002, P.R.China
wangxz@mail.hbu.edu.cn

**Dietrich Van der Weken**
VRT - Research & Innovation
August Reyerslaan 42
1043 Brussel
Dietrich.vanderweken@vrt.be

**Valérie De Witte**
Ghent University Fuzziness & Uncertainty Modelling Research Unit Krijgslaan 281 (building S9) 9000 Ghent - Belgium
Valerie.DeWitte@UGent.be

**Zeshui Xu**
Department of Management Science and Engineering School of Economics and Management Tsinghua University, Beijing 100084, China
xu_zeshui@263.net

**Daniel S. Yeung**
Department of Computing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
csdaniel@comp.polyu.edu.hk

**Sławomir Zadrożny**
Systems Research Institute, Polish Academy of Sciences ul. Newelska 6, 01-447 Warsaw, Poland
zadrozny@ibspan.waw.pl

**Guangquan Zhang**
Faculty of Information Technology, University of Technology, Sydney (UTS) PO Box 123, Broadway, NSW 2007, Australia
zhang@it.uts.edu.au

**Suyun Zhao**
Department of Computing, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
csszhao@comp.polyu.edu.hk

# Part I
# Foundations: Representation and Aggregation

# Type-2 Fuzzy Logic and the Modelling of Uncertainty

**Simon Coupland and Robert John**

**Abstract** This chapter provides a review of the development of the field of type-2 fuzzy logic. We explore some underpinning philosophical arguments that support the notion of type-2 fuzzy logic. We give the fundamental definitions of type-2 fuzzy sets and basic logical operations. The key stages in development of the field are reviewed and placed in a historical context. In addition, we report an example application of type-2 fuzzy logic to mobile robot navigation, demonstrating the potential of type-2 fuzzy systems to outperform type-1 fuzzy systems.

## 1 Introduction

Recently there has been significant growth in scientific interest in type-2 fuzzy logic. Type-2 fuzzy logic is an extension of type-1 (regular) fuzzy logic where the membership grade in a fuzzy set is itself measured as a fuzzy number. Much of this growth in interest has only been concerned with type-2 interval fuzzy systems, a subset of type-2 fuzzy systems, where the membership grade of a fuzzy set is given as an interval set. We take the view that type-2 interval fuzzy logic is a compromise and that (generalised) type-2 fuzzy sets have a great deal to offer. To demonstrate this point this Chapter begins by discussing works of philosophy that relate to fuzzy logic. It is clear from these works that type-2 fuzzy logic has a role to play in modelling uncertainty. Section 3 gives the fundamental definitions of general and interval type-2 fuzzy set along with basic logic operations. Section 4 of this Chapter presents the historical development of the field of type-2 fuzzy logic. It becomes clear that the dominance in the literature of interval techniques has only come about recently. Section 5 presents the application of type-1, type-2 interval and generalised type-2 fuzzy logic to a mobile robot control application. This example application demonstrates the potential of generalised type-2 fuzzy logic to give an improved performance over type-2 interval fuzzy logic. Section 6 draws conclusions from this work.

## 2 Philosophy, Uncertainty and Fuzzy Logic

This Section explores some of the work in the field of Philosophy that relates to the field of fuzzy logic.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008 

### 2.1 Vagueness and Imprecision

The problems of uncertainty, imprecision and vagueness have been discussed for many years. These problems have been major topics in philosophical circles with much debate, in particular, about the nature of vagueness and the ability of traditional Boolean logic to cope with concepts and perceptions that are imprecise or vague (Williamson [63] and Keefe [35]).

As early as 1923 the well known philosopher Russell discussed the notion of *Vagueness* [55]. In this early work he discusses, for example, the word 'red'. This word has no precise meaning - it is vague. The average person would clearly associate this word, for example, with the predominant colour of a London bus. However, as he points out, there are certain colours, as one moves through the colour spectrum, which could not perhaps unequivocally be described as red. There would be some uncertainty in associating the word red with that particular colour. There are other colours - the yellow of New York taxi for example - which are clearly not red. 'Red' is a vague term yet is one which is used every day in common language. Another example Russell uses in his discussion of vagueness is that of baldness[1]. If a man starts with a full head of hair and the hairs are removed one by one he will eventually be bald. At what point does he become bald? There is no precise point, no particular hair that defines the move from a position of not being bald to one of baldness. Clearly baldness is an imprecise, vague concept. He also discusses the idea that quantitative words used in science are (more or less) vague. For instance a two kilogram bag of sugar will hardly ever be exactly two kilograms. Even if it were perceived to be exactly two kilograms using the most modern measuring equipment it is unlikely to be **exactly** two kilograms since the measurement obtained is limited to the accuracy of that equipment. A two kilogram bag of sugar is in effect *about* two kilograms. All measurements have this imprecision. As Russell says *'It follows that every proposition that can be framed has a certain degree of vagueness'* [55, p. 99]. In other words all propositions are vague to some degree. His argument is that the notion of Boolean AND and Boolean OR lose their real meaning when the symbols used are vague (Other philosophers (e.g. [63]) dispute this and argue that traditional Boolean logic can deal with these vague terms. The problem is one of ignorance. Concepts such as red and not red have crisp boundaries - we just don't know where these boundaries lay. From the perspective of the arguments presented here whether Russell or Williamson is right does not affect the arguments in favour of type-2 fuzzy logic. If there is some unknown boundary between 'red' and 'not red' for many applications concepts and perceptions like red need modelling.). His definition of vagueness is: *a representation is vague when the relation of the representing system to the represented system is not one-to-one but one-too-many.* What

[1] There are a number of puzzles considered by philosophers that are known as $s\bar{o}rit\bar{e}s$. The baldness problem is one such example. Another is the question of a heap. At what point does a pile of grains of wheat make a heap? Is one grain a heap? two? a thousand? There is no particular grain that makes a heap so this raises the question of whether a heap of wheat can exist? Clearly, human beings understand the notion of a heap of wheat yet traditional logic is unable to cope with this type of puzzle.

is meant by this? Well, if the system that is being represented can, to some degree, be related to more than one representation in another system then the representation is vague. He also points out that of course the law of excluded middle[2] does not hold for imprecise or vague symbols but only for precise symbols. Black [1] also discusses the problems of modelling vagueness. He differs from Russell in that he proposes that traditional logic can be used by representing vagueness at an appropriate level of detail and suggests that Russell's definition of vagueness confuses vagueness with generality. He discusses vagueness of terms or symbols by using borderline cases where its unclear whether the term can be used to describe the case. When discussing scientific measurement he points out *"... the indeterminacy which is characteristic in vagueness is present also in all scientific measurement"* [1, p. 429] and *"Vagueness is a feature of scientific as other discourse."* [1, p. 429]. An idea put forward by Black is the idea of a consistency profile or curve to enable some analysis of the ambiguity of a word or symbol. He uses three notions - a language, a situation when a user is trying to apply a symbol $L$ to an object $x$ and the consistency of the application of $L$ to $x$. These notions are used to determine a curve that describes the consistency of application of $L$ to $x$ which is the number of observers who would apply the symbol $L$ to $x$ divided by the number of observers who would apply an alternative symbol ($\sim L$). The graph has on the vertical axis this consistency measure and on the horizontal axis the $x$'s ranked according to the consistency. He notes that the curve will be different for different symbols. Figure 1 is similar to the figure in Black [1, p. 443] and shows the consistency profile for a vague symbol $L$ and $\sim L$. Figure 2 shows the profile for a more precise symbol. To the fuzzy logic researcher of today these curves bear a strong resemblance to the membership functions of type-1 fuzzy sets Zadeh [67] but, as will be seen later, consistency profiles are different from membership functions.

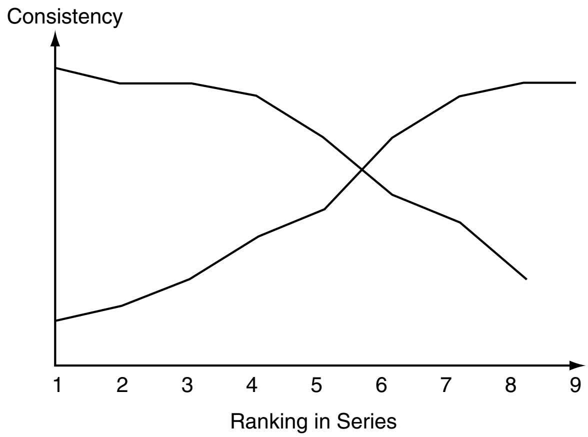
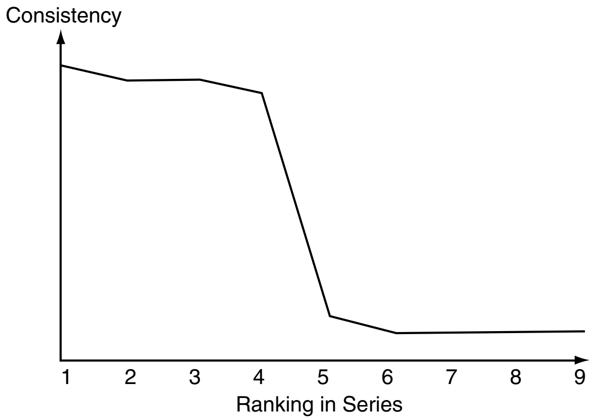



**Fig. 1** A consistency profile for a vague symbol

[2] The law of excluded middle states that if we have two sets $A$ and its complement $\overline{A}$ then the union of $A$ and $\overline{A}$ is the universal set $X$. In other words $A \cup \overline{A} = X$.

**Fig. 2** A consistency profile for a precise symbol




The notion of 'loose concepts' is presented in a seminal article by [2] where the concept of 'tallness' is discussed at length. These loose concepts relate very closely to the notion of vagueness. A loose concept according to Black is where there is no sharp boundary where a concept $C$ becomes $not\ C$. He gives the example of a sharply bounded concept "short" by using the following formal definition for short: *"There is a certain height, h, such that a man of that height is short, while a man of height* $h + \delta$ *is not short, no matter how small* $\delta$ *may be"* Black [2, p. 4] (clearly everyday use of short does not match this definition - short is another word that is vague). Any concept not sharply bounded he calls a loose concept (or vague concept) - short is therefore a loose concept. Goguen [15] considered inexact concepts and how (type-1) fuzzy sets can be used to represent such concepts. He introduces the notion of of a $J-set$ (a type of fuzzy set). $J$ denotes the closed unit interval, the set$\{a \in Y|0 \leq a \leq 1\}$. The expression $\{a \in Y|P(a)\}$ denotes the set of all elements $a$ of $Y$ such that the proposition $P$ is true of a. A $J-set$ then is a function $S : X \rightarrow J$ where J is the truth-set of S. It is interesting that Goguen briefly mentions higher type $J-set$s ($J-set$s which contain other $J-set$s) which he points out seem to measure the abstractness of a concept. This is a similar argument to the one made about type-2 fuzzy sets.

More recently (Zadeh [72]), the modelling of perceptions has become an important topic. Consider this quote:

> "... the remarkable human capability to perform a wide variety of physical and mental tasks without any measurements and any computations. Familiar examples of such tasks are parking a car; driving in heavy traffic; playing golf; understanding speech, and summarizing a story. Underlying this remarkable ability is the brain's crucial ability to manipulate perceptions - perceptions of size, distance, weight, speed, time, direction, smell, color, shape, force, likelihood, truth and intent, amongst others"
>
> Zadeh [72, pp. 106–107]

In this quote Zadeh has highlighted that the real world is imprecise in many ways. The human being is capable of handling perceptions to carry out complex tasks that cannot successfully be modelled by traditional mathematical techniques. The central tenet of his work is that fuzzy logic can be used to model perceptions. Furthermore

Zadeh, Mendel and Turksen appear to agree that model perceptions requires the use of type-2 (possibly type-2 interval.) fuzzy sets.

The real world is not precise and the notions of vagueness, looseness, uncertainty, imprecision, concepts and perceptions are central to the way human beings solve problems. The discussions, briefly described here, have informed the debate about the problems of modelling notions, concepts or perceptions that are somehow vague, imprecise or uncertain.

## 2.2 Fuzzy Logic, Imprecision and Vagueness

Fuzzy sets (Zadeh [67]) have, over the past forty years, laid the basis for a successful method of modelling uncertainty, vagueness and imprecision in a way that no other technique has been able. The use of fuzzy sets in real computer systems is extensive, particularly in consumer products and control applications.

Fuzzy logic (a logic based on fuzzy sets) is more mature than artificial neural networks with which it is often bracketed and indeed the reality is that the application of fuzzy logic is more pervasive. It is without doubt that fuzzy logic is now a mainstream technique in everyday use across the world. The number of applications is many, and growing, in a variety of areas, for example, heat exchange, warm water pressure, aircraft flight control, robot control, car speed control, power systems, nuclear reactor control, fuzzy memory devices and the fuzzy computer, control of a cement kiln, focusing of a camcorder, climate control for buildings, shower control and mobile robots (see for example Lee [37], Schwartz [58], Saffiotti [56], etc.). The use of fuzzy logic is not limited to control. Successful applications, for example, have been reported in train scheduling, system modelling, computing (OMRON [53]), stock tracking on the Nikkei stock exchange (Schwatz [58]), information retrieval (Nakamura and Iwai [52]) and the scheduling of community transport (John and Bennett [27]). The fuzzy set approach to modelling is both intuitive and exciting. That this relatively simple idea can be used to model quite complex situations is extraordinary.

Zadeh [72] presents a powerful argument for the use of fuzzy logic for manipulating perceptions. As has been discussed, his argument is that perceptions (for example, perceptions of size, safety, health and comfort) cannot be modelled by traditional mathematical techniques and that fuzzy logic is more suitable. The discussion about perception modelling is both new and exciting. We argue that type-2 fuzzy sets, since they have non-crisp fuzzy membership functions, can model these perceptions more effectively than type-1 fuzzy sets where the membership grades are crisp in nature.

So, we take the position that although fuzzy logic has many successful applications there are a number of problems with the 'traditional' fuzzy logic approach that require a different set of fuzzy tools and techniques for modelling high levels of uncertainty. In particular the argument presented here is that fuzzy logic, as it is commonly used, is essentially **precise** in nature and that for many applications it is unable to model knowledge from an expert adequately. We argue that the modelling

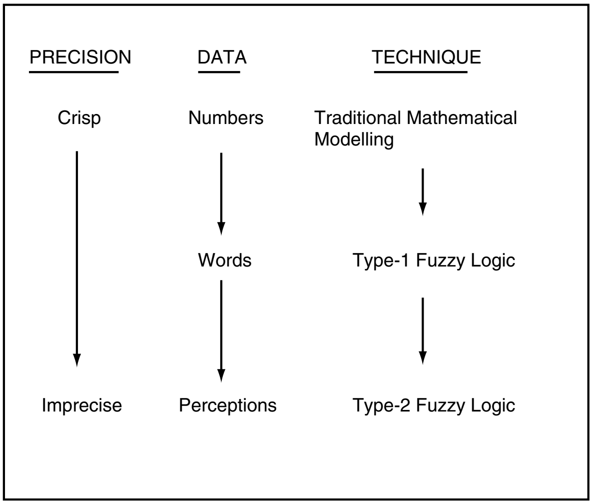



**Fig. 3** Relationships between imprecision, data and fuzzy technique

of imprecision can be enhanced by the use of type-2 fuzzy sets - providing a higher level of imprecision. Indeed, the tenet of this work is that the success of fuzzy logic can be built on by type-2 fuzzy sets and taken into the next generation of (type-2) fuzzy systems. The use of type-2 fuzzy sets allows for a better representation of uncertainty and imprecision in particular applications and domains. This argument is presented with the use of a mobile robot control application.

The more imprecise or vague the data is, then type-2 fuzzy sets offer a significant improvement on type-1 fuzzy sets. Figure 22 shows the view taken in this work of the relationships between levels of imprecision, data and technique. As the level of imprecision increases then type-2 fuzzy logic provides a powerful paradigm for potentially tackling the problem. Problems that contain crisp, precise data do not, in reality, exist. However some problems can be tackled effectively using mathematical techniques where the assumption is that the data is precise. Other problems (for example, in control) use imprecise terminology that can often be effectively modelled using type-1 fuzzy sets. Perceptions, it is argued here, are at a higher level of imprecision and type-2 fuzzy sets can effectively model this imprecision.

## 3 Key Definitions

This Section provides a core set of definitions that give the fundamental technical detail of type-2 systems.

### *3.1 Definitions of Type-2 Fuzzy Sets*

The concepts of generalized and interval type-2 fuzzy sets are now formally defined.

**Definition 1.** ***A Generalised Type-2 Fuzzy Set***
*At each value of x, such that $x \in X$, in the generalised type-2 fuzzy set $\widetilde{A}$, i.e., $\mu_{\widetilde{A}}(x)$ maps to a secondary membership function $f(x)$, which map values in* $[0, 1]$ *to values in* $[0, 1]$. *Let the domain of the secondary membership function denoted by $J_x$ then;*

$$\widetilde{A} = \int_{x \in X} \left[ \int_{u \in J_x} f_x(u)/u \right] \bigg/ x \tag{1}$$

*Where $J_x \subseteq [0, 1]$, $x \in X$, $u \in [0, 1]$ and $f_x(u) \in [0, 1]$.*

*Adapted from Mendel And John [47]*

**Definition 2.** ***A Type-2 Interval Fuzzy Set***
*At each value of x, such that $x \in X$, in the type-2 type-2 fuzzy set $\widetilde{A}$, i.e., $\mu_{\widetilde{A}}(x)$ maps to a secondary membership function $f(x)$, which map values in* $[0, 1]$ *to values in* $\{0, 1\}$. *Let the domain of the secondary membership function denoted by $J_x$ then;*

$$\widetilde{A} = \int_{x \in X} \left[ \int_{u \in J_x} 1/u \right] \bigg/ x \tag{2}$$

*Where $J_x \subseteq [0, 1]$, $x \in X$ and $u \in [0, 1]$.*

*Adapted from Mendel [45]*

**Definition 3.** *For discrete universes of discourse X and U, an embedded type-2 set $\widetilde{A}_e$ has N elements, where $\widetilde{A}_e$ contains exactly one element from $J_{x_1}, J_{x_2}, \ldots, J_{x_N}$, namely $u_1, u_2, \ldots, u_N$, each with its associated secondary grade, namely $f_{x_1}(u_1), f_{x_2}(u_2), \ldots, f_{x_N}(u_N)$, i.e.,*

$$\widetilde{A}_e = \sum_{i=1}^{N} [f_{x_i}(u_i)/u_i]/x_i \quad u_i \in J_{x_i} \subseteq U = [0, 1] \tag{3}$$

*Where $\widetilde{A}_e^j$ is the $j^{th}$ embedded set in $\widetilde{A}$ and $M_i$ is the number of points in the domain of the $i^{th}$ secondary membership function of $\widetilde{A}$. Set $\widetilde{A}_e$ is embedded in $\widetilde{A}$, and, there are a total[3] of $\prod_{i=i}^{N} M_i$.*

*Adapted from Mendel and John (2002)*

## 3.2 Definitions of Type-2 Fuzzy Logic Operators

The logical operations required for reasoning with type-2 fuzzy sets are now formally defined

[3] For continuous type-2 fuzzy sets, there are an uncountable number of embedded type-2 fuzzy sets, and this concept is not very useful.

**Definition 4.** ***The Join of Two Generalised Type-2 Fuzzy Sets***
*The join ($\sqcup$) operation finds the conjunction of two secondary membership functions $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$. Let $\mu_{\tilde{A}}(x) = \sum_{i=1}^{M} \alpha_i / v_i$ and let $\mu_{\tilde{B}}(x) = \sum_{j=1}^{N} \beta_j / w_j$. The conjunction of $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ is given by*

$$\mu_{\tilde{A} \sqcup \tilde{B}}(x) = \sum_{i=1}^{M} \sum_{j=1}^{N} (\alpha_i^* \beta_j)/(v_i \vee w_j) \tag{4}$$

*Adapted from Mizumoto and Tanaka [50]*

*where $\vee$ is a t-conorm, generally taken to be maximum and $^*$ is a t-norm such as minimum or product.*

**Definition 5.** ***The Join of Two Type-2 Interval Fuzzy Sets***
*Let $\mu_{\tilde{A}}(x) = \sum_{i=1}^{M} 1/v_i$ and let $\mu_{\tilde{B}}(x) = \sum_{j=1}^{N} 1/w_j$. The conjunction of $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ is given by*

$$\mu_{\tilde{A} \sqcup \tilde{B}}(x) = \sum_{i=1}^{M} \sum_{j=1}^{N} 1/(v_i \vee w_j) \tag{5}$$

*Adapted from Mendel [45]*

*where again $\vee$ is a t-conorm, generally taken to be maximum and $^*$ is a t-norm such as minimum or product.*

**Definition 6.** ***The Meet of Two Generalised Type-2 Fuzzy Sets***
*The meet ($\sqcap$) operation finds the disjunction of two secondary membership functions $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$. The disjunction of $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ is given by*

$$\mu_{\tilde{A} \sqcap \tilde{B}}(x) = \sum_{i=1}^{M} \sum_{j=1}^{N} (\alpha_i^* \beta_j)/(v_i^* w_j) \tag{6}$$

*Adapted from Mizumoto and Tanaka [50]*

*where again $^*$ is a t-norm.*

**Definition 7.** ***The Meet of Two Type-2 Interval Fuzzy Sets***
*Let $\mu_{\tilde{A}}(x) = \sum_{i=1}^{M} 1/v_i$ and let $\mu_{\tilde{B}}(x) = \sum_{j=1}^{N} 1/w_j$. The disjunction of $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ is given by*

$$\mu_{\tilde{A} \sqcup \tilde{B}}(x) = \sum_{i=1}^{M} \sum_{j=1}^{N} 1/(v_i^* w_j) \tag{7}$$

*Adapted from Mendel [45]*

*where again $^*$ is a t-norm.*

**Definition 8.** ***The Centroid of a Generalised Type-2 Fuzzy Set***
*The generalized centroid $(GC)$ gives a possiblistic distribution of the centroids of a Generalised type-2 fuzzy set. Let $\widetilde{A}$ be a discrete type-2 fuzzy set with $L$ discrete points in its domain. Let $n$ be the number of embedded type-2 sets required to represent $\widetilde{A}$ using the representation theorem [47]. The generalised centroid of $\widetilde{A}$ may be given as*

$$GC_{\widetilde{A}} = \sum_{i=1}^{n}[\bigstar_{j=1}^{L}\mu_{\widetilde{A}_e^i}(x_j, u_j)] \Big/ \frac{\sum_{j=1}^{L} x_j u_j}{\sum_{j=1}^{L} x_j} \tag{8}$$

*Adapted from Karnik and Mendel [34]*

*where $\mu_{\widetilde{A}_e^i}(x_j, u_j)$, $x_j$ and $u_j$ follow from the definition of a type-2 embedded set given in (3) and $\bigstar_{j=1}^{L}\mu_{\widetilde{A}_e^i}(x_j, u_j)$ is the t-norm of all values of $\mu_{\widetilde{A}_e^i}(x_j, u_j)$ from 1 to L.*

**Definition 9.** ***The Centroid of a Type-2 Interval Fuzzy Set***
*The generalised centroid of a type-2 interval fuzzy set $\widetilde{A}$ over the domain $X$ is defined as:*

$$GC_{\widetilde{A}} = \int_{\theta_1 \in J_{x_1}} \cdots \int_{\theta_1 \in J_{x_N}} 1 \Big/ \frac{\sum_{i=1}^{N} x_i \theta_i}{\sum_{i=1}^{N} \theta_i} = [C_l, C_r] \tag{9}$$

*Adapted from Mendel [45]*

*where $J_{x_N}$ is the secondary membership grade at $N$ in the secondary membership function $J_x$ and $x \in X$. The type-reduced set $C$ only needs two endpoints to define it, $C_l$ and $C_r$. Each of these points come from the centroid values of a set that is embedded in $\widetilde{A}$. The iterative method [34] exploits the properties of the centroid operation to find these two sets with a relatively low amount of computational effort.*

The following Section of this Chapter explores the development of the field of type-2 fuzzy logic.

## 4 The Historical Development of Type-2 Fuzzy Logic

Type-2 fuzzy logic is a growing research topic. Figure 4 illustrates the growth in academic activity in the field since its inception. In this Section the main themes reported in literature are discussed in a historical context. Figure 5 depicts a time line showing the historical development of type-2 methods. Each block on the time line relates to a paragraph presented in this Section.

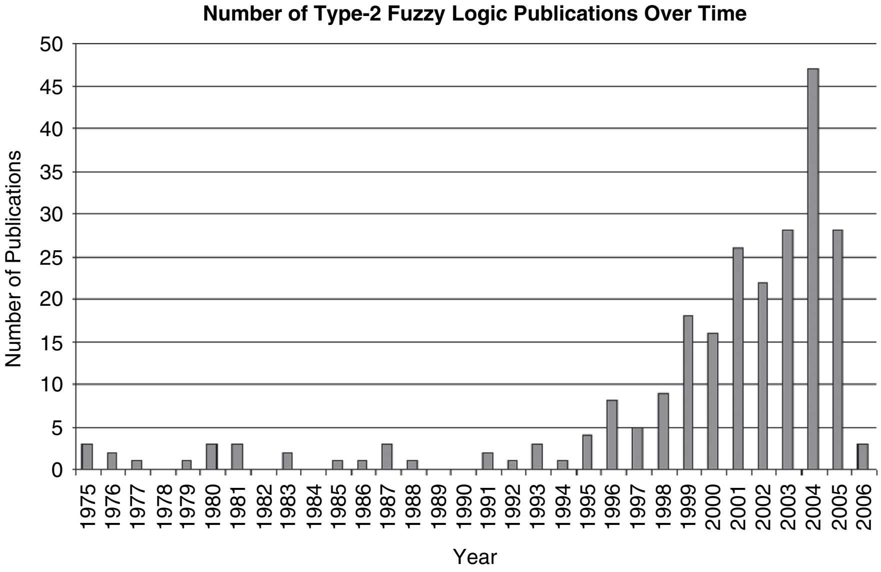



**Fig. 4** The number of type-2 related publications over time
Source: http://www.type2fuzzylogic.org/publications/ accessed 01/03/2006

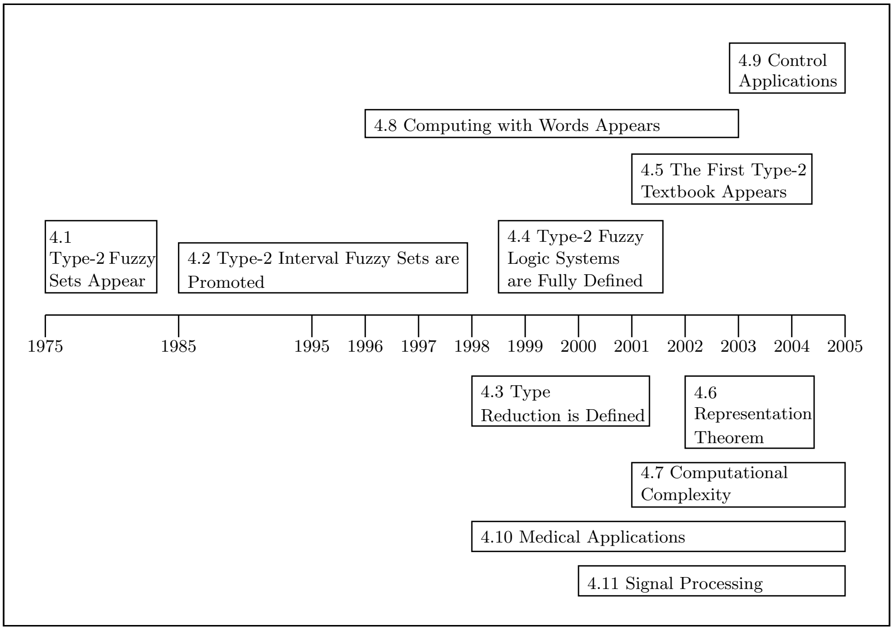



**Fig. 5** A time line depicting the historical development of type-2 fuzzy logic

### 4.1 Type-2 Fuzzy Sets Appear

Type-2 fuzzy sets were first defined and discussed in a trilogy of papers by Zadeh [68, 69, 70]. These papers concentrated on the notion of a fuzzy set where the memberships grades of a fuzzy set are measured with linguistic terms such as *low*, *medium* and *high*. Logical operators for such sets were also given, although the terms join and meet were not used. Zadeh only explored the use of the minimum and maximum operators t-norm and t-conorm when investigating the logical operations. Mizumoto and Tanaka [49, 50] and Dubois and Prade [12] both studied the logical connectives of what became known as secondary membership functions. Mizumoto and Tanaka were the first to use the terms join and meet for these logical connectives. Both Dubois and Prade and Mizumoto and Tanaka studied the join and meet under a variety of t-norm and t-conorm operators.

### 4.2 Type-2 Interval Fuzzy Sets are Promoted

Turksen [59, 60, 61], Schwartz [57], Bustince [3], Gorzalçany [16] and Klir and Folger [36] promoted the use of type-2 fuzzy sets, at that time called interval valued or IV fuzzy sets. Schwartz believes that type-2 interval fuzzy sets should be employed when the linguistic uncertainty of a term cannot be sufficiently modelled by the type-1 methods. Klir and Folger advocate the use of IV fuzzy sets when the membership functions of type-1 fuzzy sets could not be agreed upon. These arguments were explored in greater detail by Mendel [43]. Turksen put forward a collection of logical operators for type-2 interval fuzzy sets noting that the expressive power of type-2 fuzzy reasoning lies in the ability to retain the uncertainty throughout the inferencing process.

### 4.3 Type-reduction is Defined

Karnik and Mendel [31, 32, 34] defined type-reduction, the technique used for defuzzifing type-2 fuzzy sets, by applying the extension principle to a variety of type-1 defuzzifiers. The notion of an output processing stage of a type-2 fuzzy system was developed in these papers.

### 4.4 Type-2 Fuzzy Logic Systems are Fully Defined

Karnik and Mendel [31, 34] gave a complete description of the fuzzy inferencing process. This allowed work on the application of type-2 fuzzy logic to proceed. Around this time John [23, 24, 25, 26] published a series of review papers on type-2 fuzzy systems. Early applications of the technology also began to appear (see for example John [22, 28] and Karnik and Mendel [33]). As Fig. 4 demonstrates, the recent growth in type-2 fuzzy publications began around this time.

## 4.5 The First Textbook on the Subject of Type-2 Fuzzy Logic Appears

Following the consolidation of the definitions and existing literature by John and Karnik and Mendel, the field was opened up to a wider potential audience with the publication of the first type-2 textbook. Uncertain Rule-Based Fuzzy Logic System: Introduction and New Directions was written by Mendel [45] and published in 2001. This textbook references a great deal of the work on type-2 fuzzy logic that had been published to date, bringing together many of Mendel's earlier publications.

## 4.6 The Representation Theorem is Defined

Mendel and John [47] gave the representation theorem of type-2 fuzzy sets. By representing a type-2 fuzzy set as a collection of simpler type-2 embedded sets it is possible to define operations of type-2 fuzzy sets without the use of the extension principle. The motivation behind this work was that by eliminating the need to learn about the extension principle, the field would be more accessible to type-1 fuzzy practitioners. However, the representation theorem has its own learning curve, and is not significantly simpler to understand than the extension principle.

## 4.7 Issues of Computational Complexity Begin to be Explored

The complexity of join and meet operations and type-reduction of a type-2 fuzzy set limit the applicability of type-2 methods. Although type-2 interval sets are simpler, type-reduction is still a problem, due to inherent complexity and redundancies. The iterative method (Karnik and Mendel [34]) and the Wu-Mendel [65, 66] approximation were developed to make the type-reduction of type-2 interval fuzzy sets more efficient. This has led to the majority of the publications in the field of type-2 only discussing type-2 interval methods. Indeed, many authors refer to type-2 interval fuzzy set as type-2 fuzzy sets and add the qualifying term 'generalised' when discussing actual type-2 fuzzy sets. The computational problems of join and meet were effectively resolved by Karnik and Mendel [34]. This work is also discussed by the author, along with some aspects of the geometric approach in Coupland et al. [6, 8, 9]. Greenfield et al. [17] give an efficient method for approximating the the type-reduced set of a type-2 fuzzy set using a stochastic approach.

### 4.8 Computing with Words Appears

Zadeh [72, 71] made the claim that fuzzy logic, approximately at least, equates to computing with words (CWW). In CWW numbers are replaced with words not only when reasoning, but also when solving calculations. Zadeh's examples use fuzzy granules to model words. A fuzzy granule is actually the FOU of a type-2 interval fuzzy set. Both Mendel [44, 46] and Turksen [62] point out that CWW requires type-2 fuzzy sets, both opting to use the simpler type-2 interval representations. Mendel [43] re-emphasised this point by demonstrating that human models of words as obtained through a survey require at least interval representations. The authors' opinion is that type-2 fuzzy logic does not constitute computing with words. Type-2 fuzzy sets can model perceptions. Linguistic terms are not only models of perception, they also may contain other characteristics, such as contextual dependence, that fuzzy logic bears no relation to.

### 4.9 Control Applications

With the iterative method and the Wu-Mendel approximation allowing fast execution of type-2 fuzzy systems, control applications began to emerge. Melin and Castillo [41, 42] used type-2 interval systems in the context of plant control. Hagras [18] demonstrated that a type-2 interval fuzzy logic controller could outperform a type-1 fuzzy controller under large uncertainties. Wu and Tan [64] applied type-2 interval systems to the control of a complex multi-variable liquid level process. Figueroa et al. [13] used a type-2 interval control for non-autonomous robots in the context of a robot football game. The authors' have performed a comprehensive study of both general and type-2 interval fuzzy controllers for an autonomous mobile robot. Some aspects of these studies are presented in Sect. 5 of this work and in Coupland [9]. Doctor et al. [11] used a type-2 interval system to model and adapt to the behaviour of people in an intelligent dormitory room. Lynch et al. [39] are continuing to build a type-2 interval control system for large marine diesel engines. Melgarejo et al. [40] have developed a limited hardware implementation of a type-2 interval controller.

### 4.10 Medical Applications

Medical applications are one of the few areas where a generalised type-2 fuzzy logic has been used in preference to type-2 interval fuzzy logic. This is largely because such systems do not require fast execution times but do contain large uncertainties. John et al. [28, 21] used a type-2 fuzzy system for the pre-processing of tibia radiographic images. Garibaldi et al. [14, 54] have done extensive work on assessing the health of a new born baby using knowledge of acid-base balance in the blood from the umbilical cord. Innocent and John [19] proposed the use of fuzzy cognitive maps to aid the differential diagnosis of confusable diseases and suggest that type-2 cognitive maps may yield improved results. Di Lascio et al. [10] also used type-2 fuzzy sets to model differential diagnosis of diseases, modelling the compatibility of the symptom to a disease as a linguistic term. John et al. [20, 29] used type-2 fuzzy sets to model the perception of clinical opinions of nursing staff as linguistic terms.

### *4.11 Signal Processing*

Signal processing, like control, has to date only used type-2 interval methods. Liang and Mendel [38] implemented a fuzzy adaptive filter for the equalization of non-linear time-varying channels. Mitchell [48] defined a similarity measure for use with type-2 fuzzy sets which was used in a radiographic image classifier. Karnik and Mendel [33] used a type-2 interval system to predict the next value in a chaotic time series. Musikasuwan et al. [51] investigated how the learning capabilities of type-1 and type-2 interval systems differ according to the number of learning parameters used. Both systems were designed to to predict a Mackey-Glass time series.

### *4.12 Summary*

This Section has given the major developments that have taken place in the field of type-2 fuzzy logic and places them in a historical context. Type-2 literature has become predominately concerned with type-2 interval methods. The likely reason for this is the elimination of the computational problems for type-2 interval methods. The authors' view is that generalised type-2 fuzzy logic has a great deal to offer as will be demonstrated in the following section.

## 5 A Comparison of a Generalised and an Interval Type-2 Fuzzy Logic Controller

There are currently no reported systems (except Coupland et al. [4] which reports some aspects of this experiment) where generalised type-2 fuzzy logic has been applied to a control application. Recent theoretical advances by Coupland et al. [5, 6, 7, 8, 9] and Greenfield et al. [17] make such applications possible. This Section describes the first such application which has been made possible with the introduction of geometric type-2 fuzzy logic. We present a comparison of three fuzzy logic controllers which are given the task of navigating around the curved obstacle depicted in Fig. 6. The task of the FLC is essentially to minimise the deviation from

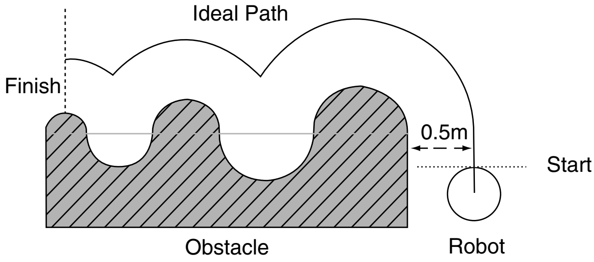



**Fig. 6** Mobile robot and obstacle

the ideal path between the start and finish lines. Each of the three controllers is based on a different fuzzy technology:

- Controller 1 uses type-1 fuzzy logic;
- Controller 2 uses type-2 interval fuzzy logic, and
- Controller 3 uses hybrid type-2 fuzzy logic.

The type-1 controller was designed first and provides a basis for controllers 2 and 3. The hybrid type-2 controller makes use of geometric type-2 fuzzy logic in order to achieve the execution speeds requires by the robot control system.

The task of mobile robot navigation represents a significant challenge for a type-2 FLC. The control system has to operate in real time on limited hardware resources. The environment which the robot has to operate in is challenging. The sensors on the robot are operating in the real world and are prone to noise and error. Typically a type-2 system will be applied in applications with high levels of uncertainty, as has we have just described.

## *5.1 Results*

The path each robot FLC took around the obstacle was tracked fifty times. These tracked paths are depicted in Figs. 7, 8 and 9. The error for each point in this tracked path relative to an ideal path was calculated. The RMSE for each tracked run around the obstacle was then calculated. The mean, median, standard deviation and coefficient of variance over the fifty runs was then calculated for each robot FLC. These results are given in Table 1.

An initial visual comparison would suggest that the controller 3 performed most consistently. Controller 2 had a wide but consistent spread. Controller 1 had spread

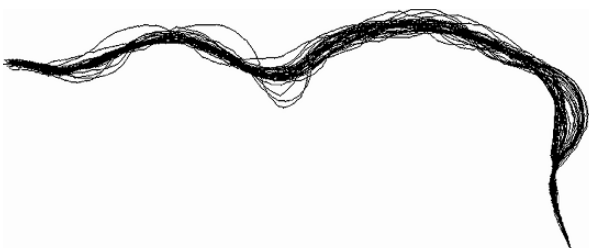


**Fig. 7** Paths taken by controller 1

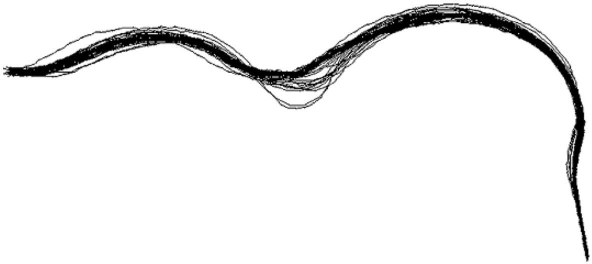


**Fig. 8** Paths taken by controller 2

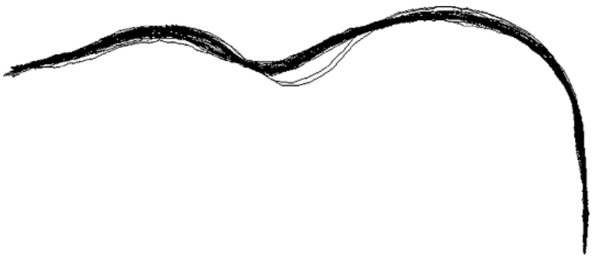


**Fig. 9** Paths taken by controller 3

of paths somewhere between the two with a few paths quite far outside the main spread. It is difficult to judge the error of the controllers visually, although the Controller 3 path appear more tightly packed than the other two.

The results from the experiment did not display either normality or equality of variance. Therefore the non-parametric Kruskal-Wallis test was used to assess whether or not there are any differences between the controllers' performance. The test gave a $H$ statistic value of 97.01 and a $p$ value of $< 0.0005$, suggesting strong evidence of differences between the controllers. The Kruskal-Wallis test works by ranking the data by median value. Table 2 gives the median values and the average ranking of the three controllers. No statistically significant conclusions can be drawn from these rankings. However the median positions and mean rankings do point to controller 3 having the best performance, followed by controller 2 and by controller 1. This performance ranking is identical to the ordering of the $R\bar{M}SE$ of the FLC. Looking at consistency of performance both the test for equal variances and the values of $\sigma RMSE$ suggest that controller 1 and controller 3 FLC were equally consistent. Controller 2 had a less consistent performance. To summarise these points:

- The type-2 FLC performed consistently well.
- The interval type-2 FLC performed quite well, but was a little inconsistent.
- The type-1 FLC performed relatively badly, but was consistent in this level of error.

**Table 1** The mean, median, standard deviation and coefficient of variance of error for the six Robot FLC over fifty runs. All numbers quoted to 4 decimal places

| Controller | Mean Error | Median Error | St Dev of Error | Co Var of Error |
|---|---|---|---|---|
| 1 | 13.5852 | 13.4185 | 1.0995 | 0.0809 |
| 2 | 12.5394 | 11.9779 | 2.0543 | 0.1638 |
| 3 | 9.8171 | 9.7783 | 1.0185 | 0.1038 |

**Table 2** The Median and Average Rank of the Three Controllers from the Kruskal-Wallis Test Procedure

| Controller | 1 | 2 | 3 |
|---|---|---|---|
| Median | 13.392 | 11.961 | 9.802 |
| Average Rank | 113.3 | 84.2 | 29.0 |

These findings are supported by a visual inspection of taken and by a statistical analysis of those paths.

## 6 Conclusion

This Chapter has reviewed the ideas at the core of type-2 fuzzy logic. In Sect. 2 we examined the philosophical arguments that show everyday reasoning uses on vague and uncertain concepts and argued that type-2 fuzzy sets are better placed than type-1 fuzzy sets to model such concepts. Definitions of generalised and interval type-2 fuzzy sets and basic logical operations were given in Sect. 3. This give the reader the technical detail required to begin working with type-2 fuzzy systems. Section 4 presented the historical development of type-2 fuzzy logic, addressing the main developments in the field and noting the emergence of different aspect of the technology. This historical review shows that type-2 interval publications only began to become predominant after the iterative procedure was introduced. The iterative procedure made the time-critical type-2 interval applications possible. The authors' opinion is that interval techniques have not been promoted solely on the basis of merit but, in part, due to their ease of application. Section 5 presented a comparison of type-1, type-2 interval and generalised type-2 fuzzy logic controllers for a mobile robot. This investigation showed that type-2 fuzzy logic controllers can give a improved performance over both type-1 and type-2 interval fuzzy controllers. We feel that the field of generalised type-2 fuzzy logic will be of great interest in the coming years.

## References


1. M. Black. Vagueness. *Philosophy of Science*, pp. 427–455, 1937.
2. M. Black. Vagueness. *Reasoning with Loose Concepts*, pp. 1–12, 1963.
3. H. Bustince. Indicator of inclusion grade for interval-valued fuzzy sets. application to approximate reasoning based on interval-valued fuzzy sets. 23(3):137–209, 2000.
4. S. Coupland, M. Gongora, R. John, and K. Wills. A Comparative Study of Fuzzy Logic Controllers for Autonomous Robots. In *Proc. IPMU 2006*, Paris, France, July 2006. Accepted for publication.
5. S. Coupland and R. John. A New and Efficient Method for the Type-2 Meet Operation. In *Proc. FUZZ-IEEE 2004*, pp. 959–964, Budapest, Hungary, July 2004.
6. S. Coupland and R. John. Fuzzy Logic and Computational Geometry. In *Proc. RASC 2004*, pp. 3–8, Nottingham, England, December 2004.
7. S. Coupland and R. John. Geometric Interval Type-2 Fuzzy Systems. In *Proc. EUSFLAT 2005*, pp. 449–454, Barcelona, Spain, September 2005.

8. S. Coupland and R. John. Towards More Efficient Type-2 Fuzzy Logic Systems. In *Proc. FUZZ-IEEE 2005*, pp. 236–241, Reno, NV, USA, May 2005.
9. S. Coupland and R. John. Geometric Type-1 and Type-2 Fuzzy Logic Systems. *IEEE Transactions on Fuzzy Systems*, 2006. Accepted for publication.
10. L. Di Lascio, A. Gisolfi, and A. Nappi. Medical differential diagnosis through Type-2 Fuzzy Sets. In *Proc. FUZZ-IEEE 2005*, pp. 371–376, Reno, NV, USA, May 2005.
11. F. Doctor, H. Hagras, and V. Callaghan. A Type-2 Fuzzy Embedded Agent For Ubiquitous Computing Environments. In *Proc. FUZZ-IEEE 2004*, pp. 1105–1110, Budapest, Hungary, July 2004.
12. D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York, 1980.
13. J. Figueroa, J. Posada, J. Soriano, M. Melgarejo, and S. Rojas. A Type-2 Fuzzy Controller for Tracking Mobile Objects in the Context of Robotic Soccer Games. In *Proc. FUZZ-IEEE 2005*, pp. 359–364, Reno, AZ, USA, May 2005.
14. J. M. Garibaldi, J. A. Westgate, E. C. Ifeachor, and K. R. Greene. The Development and Implementation of an Expert System for the Analysis of Umbilical Cord Blood. *Artificial Intelligence in Medicine*, 10(2):129–144, 1997.
15. J. Goguen. The logic of inexact concepts. *Synthese*, pp. 325–373, 1979.
16. M. B. Gorzalçany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, pp. 1–17, 1987.
17. S. Greenfield, R. John, and S. Coupland. A Novel Sampling Method for Type-2 Defuzzification. In *Proc. UKCI 2005*, pp. 120–127, 2005.
18. H. Hagras. A Hierarchical Type-2 Fuzzy Logic Control Architecture for Autonomous Mobile Robots. *IEEE Transactions on Fuzzy Systems*, 12:524–539, 2004.
19. P. Innocent and R. I. John. Computer Aided Fuzzy Medical Diagnosis. *Information Sciences*, 162:81–104, 2004.
20. R. John and S. Lake. Modelling nursing perceptions using type-2 fuzzy sets. In *EUROFUSE 2001 Workshop on Preference Modelling and Applications*, pp. 241–246, 2001.
21. R. I. John, P. R. Innocent, and M. R. Barnes. Neuro-fuzzy clustering of radiographic tibia image data using type-2 fuzzy sets. *Information Sciences*, 125:203–220, 2000.
22. R. I. John. Type-2 inferencing and community transport scheduling. In *Proc. Fourth European Congress on Intelligent Techniques and Soft Computing, EUFIT 1996*, pp. 1369–1372, Aachen, Germany, September 1996.
23. R. I. John. Type–2 Fuzzy Sets for Knowledge Representation and Inferencing. In *Proc. 7th Intl. Conf. on Fuzzy Systems FUZZ-IEEE 1998*, pp. 1003–1008, 1998.
24. R. I. John. Type 2 Fuzzy Sets: An Appraisal of Theory and Applications. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, 6(6):563–576, 1998.
25. R. I. John. Fuzzy sets of type-2. *Journal of Advanced Computational Intelligence*, 3(6): 499–508, 1999.
26. R. I. John. Type-2 fuzzy sets. *Expert Update*, 2(2), 1999. ISSN 1465-4091.
27. R. I. John and S. C. Bennett. The use of fuzzy sets for resource allocation in an advance request vehicle brokerage system - a case study. *Journal of the Operational Research Society*, 1997.
28. R. I. John, P. R. Innocent, and M. R. Barnes. Type–2 Fuzzy Sets and Neuro-Fuzzy Clustering of Radiographic Tibia Images. In *Proc. FUZZ-IEEE 1998*, pp. 1373–1376, 1998.
29. R. I. John and S. Lake. Type-2 fuzzy sets for modelling nursing intuition. In *Proc. Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pp. 1920–1925, July 2001.
30. N. N. Karnik and J. M. Mendel. An Introduction to Type-2 Fuzzy Logic Systems. Technical report, University of Southern California, 1998.
31. N. N. Karnik and J. M. Mendel. Introduction to Type-2 Fuzzy Logic Systems. In *Proc. IEEE World Congress on Computational Intelligence*, pp. 915–920, Anchorage, Alaska, USA, 1998.
32. N. N. Karnik and J. M. Mendel. Type-2 Fuzzy Logic Systems: Type-Reduction. In *Proc. IEEE Systems, Man and Cybernetics*, pp. 2046–2051, 1998.
33. N. N. Karnik and J. M. Mendel. Application of Type-2 Fuzzy Logic System to Forecasting of Time-Series. *Information Sciences*, 120:89–111, 1999.

34. N. N. Karnik and J. M. Mendel. Centroid of a type-2 fuzzy Set. *Information Sciences*, 132: 195–220, 2001.
35. R. Keefe and P. Smith. *Vagueness: A Reader*. MIT press, 1997.
36. G. J. Klir and T. A. Folger. *Fuzzy Sets, Uncertainty, and Information*. Prentice-Hall, 1988.
37. C. Lee. Fuzzy logic in control systems: Fuzzy logic controller, part ii. *IEEE Transactions on Systems, Man and Cybernetics*, 20(2):419–435, 1990.
38. Q. Liang and J. M. Mendel. Equalization of Nonlinear Time-Varying Channels Using Type-2 Fuzzy Adaptive Filters. *IEEE Transactions on Fuzzy Systems*, 8:551–563, 2000.
39. C. Lynch, H. Hagras, and V. Callaghan. Embedded Type-2 FLC for Real-Time Speed Control of Marine and Traction Diesel Engines. In *Proc. FUZZ-IEEE 2005*, pp. 347–352, Reno, AZ, USA, May 2005.
40. M. Melgarejo and C Pena-Reyes. Hardware Architecture and FPGA Implementation of a Type-2 Fuzzy System. In *Proc. GLSVSLI 2004*, pp. 458–461, Boston, Massachusetts, USA, April 2004.
41. P. Melin and O. Castillo. Fuzzy Logic for Plant Monitoring and Diagnostics. In *Proc. NAFIPS 2003*, pp. 20–25, July 2003.
42. P. Melin and O. Castillo. Intelligent Control of Non-Linear Dynamic Plants Using Type-2 Fuzzy Logic and Neural Networks. In *Proc. FUZZ-IEEE 2004*, Budapest, Hungary, July 2004.
43. J. M. Mendel. Computing With Words, When Words Mean Different Things to Different People. In *Proc. of Third International ICSC Symposium on Fuzzy Logic and Applications*, Rochester Univ., Rochester, NY., 1999.
44. J. M. Mendel. The Perceptual Computer: an Architecture for Computing With Words. In *Proc. FUZZ-IEEE 2001*, Melbourne, Australia, 2001.
45. J. M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Upper Saddle River, NJ, 2001.
46. J. M. Mendel. Fuzzy sets for words: a new beginning. In *Proc. FUZZ-IEEE 2003*, pp. 37–42, St. Louis, MO, USA, 2003.
47. J. M. Mendel and R. I. John. Type-2 Fuzzy Sets Made Simple. *IEEE Transaction on Fuzzy Systems*, 10(2):117–127, 2002.
48. H. B. Mitchell. Pattern Recognition Using Type-II Fuzzy Sets. *Information Sciences*, 170:409–418, 2005.
49. M. Mizumoto and K. Tanaka. Some properties of fuzzy set of type-2. *Information and control*, 31:312–340, 1976.
50. M. Mizumoto and K. Tanaka. Fuzzy Sets of Type 2 Under Algebraic Product and Algebraic Sum. *Fuzzy Sets and Systems*, 5:277–290, 1981.
51. S. Musikasuwan, T. Ozen, and J. M. Garibaldi. An investigation into the effect of number of model parameters on performance in type-1 and type-2 fuzzy logic systems. In *Proc. 10th Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU 2004)*, pp. 1593–1600, Perugia, Italy, 2004.
52. K. Nakamura and S. Iwai. Topological fuzzy sets as a quantative description of analogical inference and its application to question-answering systems for information retrieval. *IEEE Transactions on Systems, Man and Cybernetics*, pp. 193–204, 1982.
53. OMRON. Clearly fuzzy. Technical report, OMRON, 1992.
54. T. Ozen and J. M. Garibaldi. Investigating Adaptation in Type-2 Fuzzy Logic Systems Applied to Umbilical Acid-Base Assessment. In *Proc. of the 2003 European Symposium on Intelligent Technologies*, pp. 289–294, Oulu, Finland, July 2003.
55. B. Russell. Vagueness. *Austrian Journal of Philosophy*, 1:84–92, 1923.
56. A. Saffiotti. Fuzzy Logic in Autonomous Robotics: Behavior Co-ordination. In *Proc. IEEE Int. Conf. Fuzzy Systems*, pp. 573–578, 1997.
57. Daniel G. Schwartz. The case for an interval-based representation of linguistic truth. *Fuzzy Sets and Systems*, 17:153–165, 1985.
58. T. J. Schwarz. Fuzzy systems in the real world. *AI Expert*, 1990.
59. I. B. Türkşen. Interval-valued fuzzy sets and fuzzy connectives. *Interval Computations*, 4: 35–38, 1993.

60. I. B. Türkşen. Interval-valued fuzzy uncertainty. In *Proc. Fifth IFSA World Congress*, pp. 35–38, Seoul, Korea, July 1993.
61. I. B. Türkşen. Knowledge representation and approximate reasoning with type ii fuzzy sets. In *Proc. FUZZ-IEEE 1995*, Vol. 2, pp. 1911–1917, Yokohama, Japan, March 1995.
62. I. B. Türkşen. Type 2 Representation and Reasoning for CWW. *Fuzzy Sets and Systems*, 127:17–36, 2002.
63. T. Williamson. *Vagueness*. Routledge, 1998.
64. D. Wu and W. W. Tan. A Type-2 Fuzzy Logic Controller for the Liquid-level Process. In *Proc. FUZZ-IEEE 2004*, pp. 953–958, Budapest, Hungary, July 2004.
65. H. Wu and J. M. Mendel. Introduction to Uncertainty Bounds and Their Use in the Design of Interval Type-2 Fuzzy Logic Systems. In *Proc. FUZZ-IEEE 2001*, Melbourne, Australia, 2001.
66. H. Wu and J. M. Mendel. Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems. *IEEE Transactions on Fuzzy Systems*, pp. 622–639, October 2002.
67. L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–353, 1965.
68. L. A. Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning. *Information Sciences*, 8:199–249, 1975.
69. L. A. Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning – II. *Information Sciences*, 8:301–357, 1975.
70. L. A. Zadeh. The Concept of a Linguistic Variable and its Application to Approximate Reasoning – III. *Information Sciences*, 9:43–80, 1975.
71. L. A. Zadeh. Fuzzy Logic = Computing with Words. *IEEE Transactions on Fuzzy Systems*, 4:103–111, 1996.
72. L. A. Zadeh. From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions. *IEEE Transactions on Circuits and Systems – I:Fundamental Theory and Applications*, 45:105–119, 1999.

# My Personal View on Intuitionistic Fuzzy Sets Theory

**Krassimir T. Atanassov**

**Abstract** In this chapter, some remarks are given on the history, theory, applications and research on the extension of fuzzy sets model proposed by the author in 1983.

## 1 Introduction

This chapter[1] discusses the origin, motivation and current state of research of one of the extensions of Lotfi Zadeh's fuzzy sets [41].

The author would like to ask the reader to let him use, whenever personal attitude or opinion is involved, the first person singular, reserving the usual 'we' for statements whose truth is not subjective.

The origin of my idea of intuitionistic fuzziness was a happenstance: as a mathematical game. I read the Russian translation of A. Kaufmann's book [50][2] and decided to add to the definition a second degree (degree of non-membership) and studied the properties of a set with both degrees. Of course, I observed that the new set is an extension of the ordinary fuzzy set, but I did not notice immediately that it has essentially different properties. So the first research works on IFS followed step by step the existing results in fuzzy sets theory. Of course, it is not very difficult to extend formally some concepts. It is interesting to show that the respective extension has specific properties, absent in the basic concept.

Only when I convinced myself that the so-constructed sets really have worthy properties, I discussed them with my former lecturer from the time when I was a student at the Mathematical Faculty of Sofia University - George Gargov (7 April 1947–9 Nov. 1996) - one of the most colourful Bulgarian mathematicians. He proposed the name *"Intuitionistic Fuzzy Set" (IFS)* (see, e.g. [32]).

Let us have a fixed universe $E$ and its subset $A$. The set

$$A^* = \{\langle x, \mu_A(x), \nu_A(x)\rangle \mid x \in E\},$$

[1] The present chapter uses elements of author's research [9, 10, 11].

[2] In early 80's, only Russian translations of the books [14, 41, 50] were available in Bulgaria and for this reason it was these books that influenced the development of the first steps of IFS theory.

where

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

is called IFS and functions $\mu_A : E \to [0, 1]$ and $\nu_A : E \to [0, 1]$ represent *degree of membership (validity, etc.)* and *non-membership (non-validity, etc.)*. Now, we can define also function $\pi_A : E \to [0, 1]$ through

$$\pi(x) = 1 - \mu(x) - \nu(x)$$

and it corresponds to *degree of indeterminacy (uncertainty, etc.)*.

For brevity, we shall write below $A$ instead of $A^*$, whenever this is possible.

Obviously, for every ordinary fuzzy set $A$: $\pi_A(x) = 0$ for each $x \in E$ and these sets have the form $\{\langle x, \mu_A(x), 1 - \mu_A(x)\rangle | x \in E\}$.

It is clear that IFS *can be* different from ordinary fuzzy sets.

Throughout the past twenty-five years other extensions of fuzzy set theory have been given. The most representative ones are the following: Type 2 fuzzy sets, interval-valued fuzzy sets, vague sets, neutrosophic sets and fuzzy rough sets.

The concept of a *type 2 fuzzy set* was introduced in 1975 by Zadeh (see [41]) as a generalization of an ordinary fuzzy set. Type 2 fuzzy sets are characterized by a fuzzy membership function, that is, the membership value for each element of the set, is itself a fuzzy set in [0, 1]. A particular case of type 2 fuzzy sets are *interval type 2 fuzzy sets*, (see [39, 40]).

Interval-valued fuzzy sets, that is, sets such that the membership degree of each element to the set is given by a closed subinterval of the interval [0, 1]. In [39, 40] it turns out that an interval type 2 fuzzy set is the same as an IVFS in some special conditions.

In [23] H. Bustince and P. Burillo proved the coincidence of vague sets with IFSs, while in [30] K. Georgiev showed that on the one hand the neutrosophic sets have been incorrectly defined and on the other, if they obtain a correct definition, they will again coincide with the IFSs and they will not be IFS-extension, as their author F. Smarandache asserts, e.g., in [46].

In [46] and [1], together with G. Gargov, we discussed the equipolence (in the sense of [34]) of this concept with IFS. From our construction it is seen that each IFS can be represented by an IVFS and each IVFS can be represented by an IFS. I write these years to emphasize that then I believed IFS were defined prior to IVFS. Now, I know (merely as a fact, without having seen the original texts) that IVFS are essentially older. In 2003 this equivalence was mathematically proven by Deschrijver and Kerre (see [26]). Therefore, from a mathematical point of view, the results that we obtain for IVFSs are easily adaptable to IFSs and vice versa. Nevertheless we need to point out that conceptually and semantically both types of sets are totally different. This is made clear when applications of these sets are constructed (see [31, 47, 48]).

Finally, I must mention that Cornelis et al. (see [27]), Samanta et al. (see [45]), Rizvi et al. (see [43]) studied a possible combination of fuzzy rough sets and IFSs.

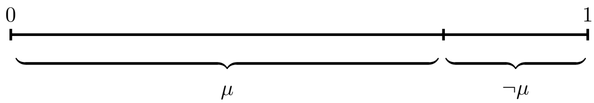
## 2 IFS and Some other Fuzzy Set and IFS-extensions

As we noted above, IFSs are an extension of the standard fuzzy sets. All results which hold for fuzzy sets have their interpretation here. Also, any research based on fuzzy sets, can be described in terms of IFS.

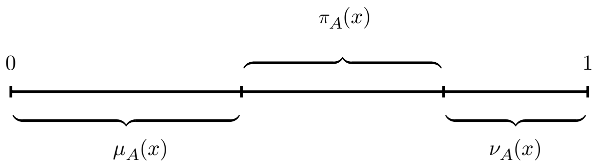
First, we will discuss the relations between ordinary fuzzy sets and IFSs from two aspects: geometrical and probabilistic. Initially, I would like to note that some authors discuss the fact that in the case of ordinary fuzzy sets

$$\mu \vee \neg\mu \leq 1$$

as a manifest of the idea of intuitionism. Really, this inequality, in its algebraic interpretation of "$\vee$" by $max$, does not satisfy the Law for Excluded Middle (LEM). But this is not the situation in a geometrical interpretation. Having in mind that in fuzzy set theory $\neg\mu = 1 - \mu$, we obtain that the geometrical interpretation is:




The situation in the IFS case is essentially different:




Now, the geometrical sums of both degrees can *really* be smaller than 1, i.e., LEM is not valid here. From probabilistic point of view, for case of the ordinary fuzzy sets, if $\mu \& \neg\mu = 0$, then the probability

$$p(\mu \vee \neg\mu) = p(\mu) + p(\neg\mu) = 1,$$

like in the geometrical case, while in IFS case we will have the inequality

$$p(\mu \vee \neg\mu) \leq 1.$$

It is important to note, that all of these constructions are only on the level of the definition of the set (fuzzy set or IFS).

IFS have different geometrical interpretations. The first of them (see Fig. 1) is a trivial modification of the fuzzy set one.

Its analogue is given in Fig. 2.

Another interpretation is shown on Fig. 3. The author's opinion is that it played very important role in IFS theory development.

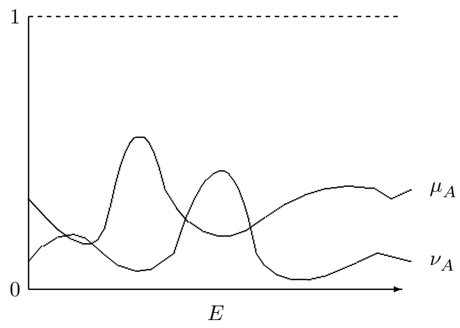



**Fig. 1** First geometric interpretation

Similarly to the fuzzy set theory, a large number of relations and operations over IFSs are defined (see, e.g., [36]), but more interesting are the modal operators that can be defined over the IFSs. They do not have analogues in fuzzy set theory.

For example, for every two IFSs $A$ and $B$ we can define:

$$A \subset B \ iff \ (\forall x \in E)(\mu_A(x) \leq \mu_B(x) \& \nu_A(x) \geq \nu_B(x));$$
$$A = B \ iff \ (\forall x \in E)(\mu_A(x) = \mu_B(x) \& \nu_A(x) = \nu_B(x));$$

$$\overline{A} = \{\langle x, \nu_A(x), \mu_A(x)\rangle | x \in E\};$$
$$A \cap B = \{\langle x, min(\mu_A(x), \mu_B(x)), max(\nu_A(x), \nu_B(x))\rangle | x \in E\};$$
$$A \cup B = \{\langle x, max(\mu_A(x), \mu_B(x)), min(\nu_A(x), \nu_B(x))\rangle | x \in E\};$$

$$A + B = \{\langle x, \mu_A(x) + \mu_B(x) - \mu_A(x).\mu_B(x), \nu_A(x).\nu_B(x)\rangle \mid x \in E\};$$
$$A.B = \{\langle x, \mu_A(x).\mu_B(x), \nu_A(x) + \nu_B(x) - \nu_A(x).\nu_B(x)\rangle \mid x \in E\};$$
$$A@B = \{\langle x, (\frac{\mu_A(x) + \mu_B(x))}{2}, \frac{(\nu_A(x) + \nu_B(x))}{2}\rangle | x \in E\}.$$

Let $A$ be a fixed IFS. The first form of operation "negation", introduced in 1983, was

$$\neg_1 A = \{\langle x, \nu_A(x), \mu_A(x)\rangle \mid x \in E\}.$$

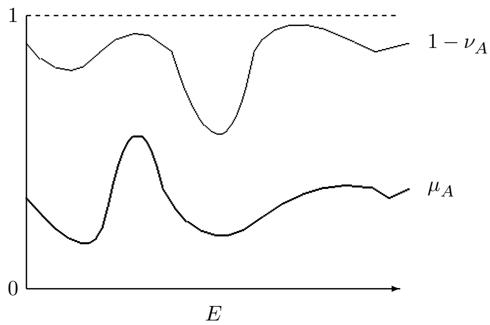



**Fig. 2** Second geometric interpretation

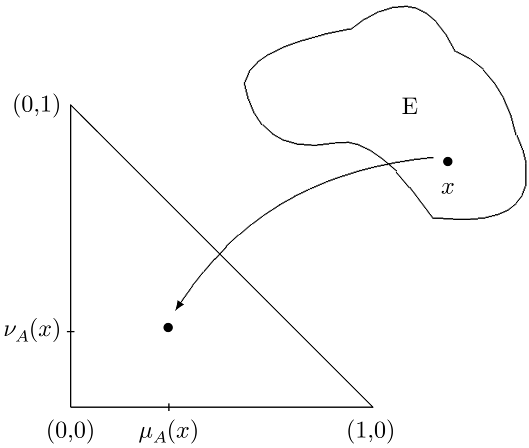



**Fig. 3** Third geometric interpretation

One of my mistakes is that for a long time I made use of just the simplest form of negation. In its case the equality

$$\neg\neg A = A \tag{2}$$

holds, which may resemble classical logic. Now, a series of *new* negations were constructed. The first four of them are defined in [14]:

$$\neg^2 A = \{\langle x, 1 - \text{sg}(\mu_A(x)), \text{sg}(\mu_A(x))\rangle | x \in E\},$$
$$\neg^3 A = \{\langle x, \nu_A(x), \mu_A(x)\nu_A(x) + \mu_A(x)^2\rangle | x \in E\},$$
$$\neg^4 A = \{\langle x, \nu_A(x), 1 - \nu_A(x)\rangle | x \in E\},$$
$$\neg^5 A = \{\langle x, \overline{sg}(1 - \nu_A(x)), \text{sg}(1 - \nu_A(x))\rangle | x \in E\},$$

where

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

$$\overline{sg}(x) = \begin{cases} 0 & \text{if } x > 0 \\ 1 & \text{if } x \leq 0 \end{cases}$$

The last four negations satisfy strongly intuitionistic properties.

In [12, 13] it is checked that $\neg_1$ satisfies all three properties below, while the rest negations satisfy only properties P1 and P3, where

**Property P1**: $A \to \neg\neg A$,
**Property P2**: $\neg\neg A \to A$,
**Property P3**: $\neg\neg\neg A = \neg A$.

By analogy with [13] we can show that negations $\neg_2$, ..., $\neg_5$ do not satisfy LEM ($P \vee \neg P$, where $P$ is a propositional form) and they satisfy some of its modifications (e.g., $\neg\neg P \vee \neg P$). In [15] it is shown that the same negations do not satisfy De Morgan's Laws, but they satisfy some of their modifications.

Independently on or in relation with some negation, a lot of implications can be defined over IFSs. Initially, in a series of papers they were introduced in the frames of the intuitionistic fuzzy logic, but in [14] they obtained the following IFS-analogues.

Let $A$ and $B$ be two fixed IFSs and let

$$X_i = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in E\} \rightarrow_i \{\langle x, \mu_B(x), \nu_B(x)\rangle | x \in E\}.$$

We shall introduce the IFS-implications.

$X_1 = \{\langle x, \max(\nu_A(x), \min(\mu_A(x), \mu_B(x))), \min(\mu_A(x), \nu_B(x)))\rangle | x \in E\}$,
$X_2 = \{\langle x, 1 - \text{sg}(\mu_A(x) - \mu_B(x)), \nu_B(x).\text{sg}(\mu_A(x) - \mu_B(x))\rangle | x \in E\}$,
$X_3 = \{\langle x, 1-(1-\mu_B(x)).\text{sg}(\mu_A(x)-\mu_B(x)), \nu_B(x).\text{sg}(\mu_A(x)-\mu_B(x))\rangle | x \in E\}$,
$X_4 = \{\langle x, \max(\nu_A(x), \mu_B(x)), \min(\mu_A(x), \nu_B(x))\rangle | x \in E\}$,
$X_5 = \{\langle x, \min(1, \nu_A(x) + \mu_B(x)), \max(0, \mu_A(x) + \nu_B(x) - 1)\rangle | x \in E\}$,
$X_6 = \{\langle x, \nu_A(x) + \mu_A(x).\mu_B(x), \mu_A(x).\nu_B(x)\rangle | x \in E\}$,
$X_7 = \{\langle x, \min(\max(\nu_A(x), \mu_B(x)), \max(\mu_A(x), \nu_A(x)), \max(\mu_B(x), \nu_B(x))),$
$\max(\min(\mu_A(x), \nu_B(x)), \min(\mu_A(x), \nu_A(x)), \min(\mu_B(x), \nu_B(x)))\rangle | x \in E\}$,
$X_8 = \{\langle x, 1 - (1 - \min(\nu_A(x), \mu_B(x))).\text{sg}(\mu_A(x) - \mu_B(x)), \max(\mu_A(x), \nu_B(x)).$
$\text{sg}(\mu_A(x) - \mu_B(x)).\text{sg}(\nu_B(x) - \nu_A(x))\rangle | x \in E\}$,
$X_9 = \{\langle x, \nu_A(x) + \mu_A(x)^2\mu_B(x), \mu_A(x)\nu_A(x) + \mu_A(x)^2\nu_B(x)\rangle | x \in E\}$,
$X_{10} = \{\langle x, \mu_B(x).\overline{sg}(1 - \mu_A(x)) + \text{sg}(1 - \mu_A(x)).(\overline{sg}(1 - \mu_B(x)) + \nu_A(x).$
$\text{sg}(1 - \mu_B(x))), \nu_B(x).\overline{sg}(1 - \mu_A(x)) + \mu_A(x).\text{sg}(1 - \mu_A(x)).\text{sg}(1 - \mu_B(x))\rangle$
$| x \in E\}$,
$X_{11} = \{\langle x, 1-(1-\mu_B(x)).\text{sg}(\mu_A(x)-\mu_B(x)), \nu_B(x).\text{sg}(\mu_A(x)-\mu_B(x)).\text{sg}(\nu_B(x)$
$-\nu_A(x))\rangle | x \in E\}$,
$X_{12} = \{\langle x, \max(\nu_A(x), \mu_B(x)), 1 - \max(\nu_A(x), \mu_B(x))\rangle | x \in E\}$,
$X_{13} = \{\langle x, \nu_A(x) + \mu_B(x) - \nu_A(x).\mu_B(x), \mu_A(x).\nu_B(x)\rangle | x \in E\}$,
$X_{14} = \{\langle x, 1 - (1 - \mu_B(x)).\text{sg}(\mu_A(x) - \mu_B(x)) - \nu_B(x).\overline{sg}(\mu_A(x) - \mu_B(x)).$
$\text{sg}(\nu_B(x) - \nu_A(x)), \nu_B(x).\text{sg}(\nu_B(x) - \nu_A(x))\rangle | x \in E\}$,
$X_{15} = \{\langle x, 1 - (1 - \min(\nu_A(x), \mu_B(x))).\text{sg}(\text{sg}(\mu_A(x) - \mu_B(x)) + \text{sg}(\nu_B(x)$
$-\nu_A(x))) - \min(\nu_A(x), \mu_B(x)).\text{sg}(\mu_A(x) - \mu_B(x)).\text{sg}(d - \nu_A(x)), 1-$
$(1 - \max(\mu_A(x), \nu_B(x))).\text{sg}(\overline{sg}(\mu_A(x) - \mu_B(x)) + \overline{sg}(\nu_B(x) - \nu_A(x)))$
$\max(\mu_A(x), \nu_B(x)).\overline{sg}(\mu_A(x) - \mu_B(x)).\overline{sg}(\nu_B(x) - \nu_A(x))\rangle | x \in E\}$,
$X_{16} = \{\langle x, \max(1 - \text{sg}(\mu_A(x)), \mu_B(x)), \min(\text{sg}(\mu_A(x)), \nu_B(x))\rangle | x \in E\}$,
$X_{17} = \{\langle x, \max(\nu_A(x), \mu_B(x)), \min(\mu_A(x).\nu_A(x) + \mu_A(x)^2, \nu_B(x))\rangle | x \in E\}$,
$X_{18} = \{\langle x, \max(\nu_A(x), \mu_B(x)), \min(1 - \nu_A(x), \nu_B(x))\rangle | x \in E\}$,
$X_{19} = \{\langle x, \max(1 - \text{sg}(\text{sg}(\mu_A(x)) + \text{sg}(1 - \nu_A(x))), \mu_B(x)), \min(\text{sg}(1 - \nu_A(x)),$
$\nu_B(x))\rangle | x \in E\}$,
$X_{20} = \{\langle x, \max(1 - \text{sg}(\mu_A(x)), 1 - \text{sg}(1 - \text{sg}(\mu_B(x)))), \min(\text{sg}(\mu_A(x)), \text{sg}(1-$
$sg(\mu_B(x))))\rangle | x \in E\}$,

$$X_{21} = \{\langle x, \max(\nu_A(x), \mu_B(x)(\mu_B(x) + \nu_B(x))), \min(a(\mu_A(x) + \nu_A(x)), \nu_B(x)(\mu_B(x)^2 + d + \mu_B(x)\nu_B(x)))\rangle | x \in E\},$$
$$X_{22} = \{\langle x, \max(\nu_A(x), 1 - \nu_B(x)), \min(1 - \nu_A(x), \nu_B(x))\rangle | x \in E\},$$
$$X_{23} = \{\langle x, 1 - \min(\text{sg}(1 - \nu_A(x)), \text{sg}(1 - \text{sg}(1 - \nu_B(x)))), \min(\text{sg}(1 - \nu_A(x)), \text{sg}(1 - \text{sg}(1 - \nu_B(x))))\rangle | x \in E\},$$
$$X_{24} = \{\langle x, \overline{sg}(\mu_A(x) - \mu_B(x)).\overline{sg}(\nu_B(x) - \nu_A(x)), \text{sg}(\mu_A(x) - \mu_B(x)).\text{sg}(\nu_B(x) - \nu_A(x))\rangle | x \in E\},$$
$$X_{25} = \{\langle x, \max(\nu_A(x).\overline{sg}(\mu_A(x)).\overline{sg}(1 - \nu_A(x)), \mu_B(x).\overline{sg}(\nu_B(x)).\overline{sg}(1 - \mu_B(x))), \min(\mu_A(x).\text{sg}(1 - \nu_A(x)), \nu_B(x).\text{sg}(1 - \mu_B(x)))\rangle | x \in E\},$$

In [14] other 149 implications are introduced, but some of them coincide and one of the actual problems is to determine the sugnificantly different implications. The above ones are examples of them, but there at least 50 other that are different than the rest.

For all 174 implications in [14] it is proved that in some sense they are extensions of the classical first order logic implication, because, if

$$\overline{O} = \{\langle x, 0, 1\rangle | x \in E\},$$
$$\overline{E} = \{\langle x, 1, 0\rangle | x \in E\},$$

then for the $i$-th implication ($1 \le i \le 25$) equalities

$$\overline{O} \rightarrow_i A = \overline{E},$$
$$A \rightarrow_i \overline{E} = \overline{E},$$
$$\overline{E} \rightarrow_i \overline{O} = \overline{O}$$

hold for every IFS $A$.

Now, I hope that the argument *"the IFSs do not have intuitionistic behaviour"* will fail.

The simplest operators are

$$\Box A = \{\langle x, \mu_A(x), 1 - \mu_A(x)\rangle | x \in E\};$$
$$\Diamond A = \{\langle x, 1 - \nu_A(x), \nu_A(x)\rangle | x \in E\}.$$

They are analogous to the modal logic operators *"necessity"* and *"possibility" (see Fig. 4).*

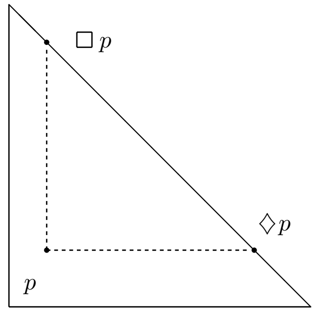
For a fuzzy set $A$ (whose special case is, e.g., Takeuti and Titani's sets), the modal operators $\Box$ and $\Diamond$ would satisfy

$$\Box A = A = \Diamond A,$$

while for a proper IFS $A$:

$$\Box A \subset A \subset \Diamond A$$

**Fig. 4** Necessity and Possibility




and

$$\Box A \neq A \neq \Diamond A.$$

This example shows that IFSs are essential extensions of fuzzy sets.

In the framework of the IFSs theory we can extend these operators in a step by step manner. The first group of extended modal operators are the following:

$$D_\alpha(A) = \{\langle x, \mu_A(x) + \alpha.\pi_A(x), \nu_A(x) + (1-\alpha).\pi_A(x)\rangle | x \in E\},$$
$$F_{\alpha,\beta}(A) = \{\langle x, \mu_A(x) + \alpha.\pi_A(x), \nu_A(x) + \beta.\pi_A(x)\rangle | x \in E\}, \text{ where } \alpha + \beta \leq 1,$$
$$G_{\alpha,\beta}(A) = \{\langle x, \alpha.\mu_A(x), \beta.\nu_A(x)\rangle | x \in E\},$$
$$H_{\alpha,\beta}(A) = \{\langle x, \alpha.\mu_A(x), \nu_A(x) + \beta.\pi_A(x)\rangle | x \in E\},$$
$$H^*_{\alpha,\beta}(A) = \{\langle x, \alpha.\mu_A(x), \nu_A(x) + \beta.(1 - \alpha.\mu_A(x) - \nu_A(x))\rangle | x \in E\},$$
$$J_{\alpha,\beta}(A) = \{\langle x, \mu_A(x) + \alpha.\pi_A(x), \beta.\nu_A(x)\rangle | x \in E\},$$
$$J^*_{\alpha,\beta}(A) = \{\langle x, \mu_A(x) + \alpha.(1 - \mu_A(x) - \beta.\nu_A(x)), \beta.\nu_A(x)\rangle | x \in E\},$$

where $\alpha, \beta \in [0, 1]$ are fixed numbers. The geometrical interpretations of the seven operators are given on Figs. 5–11.

These operators are extended to the operators

$$F_B(A) = \{\langle x, \mu_A(x) + \mu_B(x).\pi_A(x), \nu_A(x) + \nu_B(x).\pi_A(x)\rangle \mid x \in E\},$$

$$G_B(A) = \{\langle x, \mu_B(x).\mu_A(x), \nu_B(x).\nu_A(x)\rangle \mid x \in E\},$$

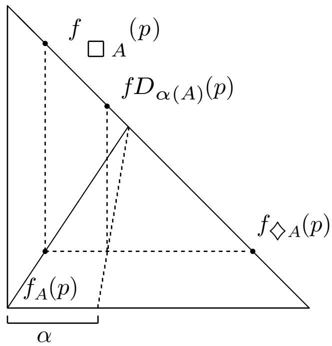



**Fig. 5** Geometric interpretation of $D_\alpha$

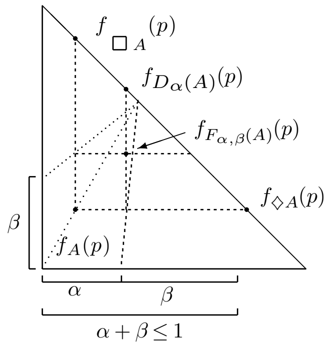



**Fig. 6** Geometric interpretation of $F_{\alpha,\beta}$

$$H_B(A) = \{\langle x, \mu_B(x).\mu_A(x), \nu_A(x) + \nu_B(x).\pi_A(x)\rangle \mid x \in E\},$$
$$H_B^*(A) = \{\langle x, \mu_B(x).\mu_A(x), \nu_A(x) + \nu_B(x).(1 - \mu_B(x).\mu_A(x) - \nu_A(x))\rangle \mid x \in E\},$$
$$J_B(A) = \{\langle x, \mu_A(x) + \mu_B(x).\pi_A(x), \nu_B(x).\nu_A(x)\rangle \mid x \in E\},$$
$$J_B^*(A) = \{\langle x, \mu_A(x) + \mu_B(x).(1 - \mu_A(x) - \nu_B(x).\nu_A(x)), \nu_B(x).\nu_A(x)\rangle \mid x \in E\},$$

where $B$ is a given IFS; and modified to operators

$$d_\alpha(A) = \{\langle x, \nu_A(x) + \alpha.\pi_A(x), \mu_A(x) + (1-\alpha).\pi_A(x)\rangle \mid x \in E\},$$
$$f_{\alpha,\beta}(A) = \{\langle x, \nu_A(x) + \alpha.\pi_A(x), \mu_A(x) + \beta.\pi_A(x)\rangle \mid x \in E\}, \text{ where } \alpha + \beta \leq 1,$$
$$g_{\alpha,\beta}(A) = \{\langle x, \alpha.\nu_A(x), \beta.\mu_A(x)\rangle \mid x \in E\},$$
$$h_{\alpha,\beta}(A) = \{\langle x, \alpha.\nu_A(x), \mu_A(x) + \beta.\pi_A(x)\rangle \mid x \in E\},$$
$$h^*_{\alpha,\beta}(A) = \{\langle x, \alpha.\nu_A(x), \mu_A(x) + \beta.(1 - \alpha.\nu_A(x) - \mu_A(x))\rangle \mid x \in E\},$$
$$j_{\alpha,\beta}(A) = \{\langle x, \nu_A(x) + \alpha.\pi_A(x), \beta.\mu_A(x)\rangle \mid x \in E\},$$
$$j^*_{\alpha,\beta}(A) = \{\langle x, \nu_A(x) + \alpha.(1 - \nu_A(x) - \beta.\mu_A(x)), \beta.\mu_A(x)\rangle \mid x \in E\}$$

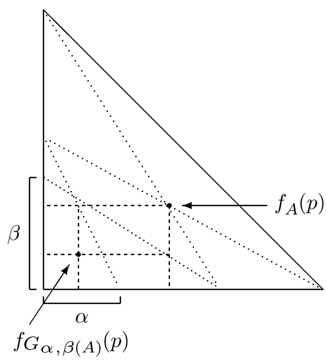
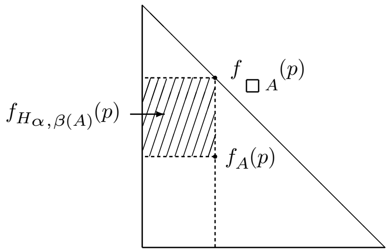



**Fig. 7** Geometric interpretation of $G_{\alpha,\beta}$

**Fig. 8** Geometric interpretation of $H_{\alpha,\beta}$




that can also be extended to the form of second group of operators:

$$f_B(A) = \{\langle x, \nu_A(x) + \mu_B(x).\pi_A(x), \mu_A(x) + \nu_B(x).\pi_A(x)\rangle \mid x \in E\},$$
$$g_B(A) = \{\langle x, \mu_B(x).\nu_A(x), \nu_B(x).\mu_A(x)\rangle \mid x \in E\},$$
$$h_B(A) = \{\langle x, \mu_B(x).\nu_A(x), \mu_A(x) + \nu_B(x).\pi_A(x)\rangle \mid x \in E\},$$
$$h^*_B(A) = \{\langle x, \mu_B(x).\nu_A(x), \mu_A(x) + \nu_B(x).(1 - \mu_B(x).$$
$$\nu_A(x) - \nu_A(x))\rangle \mid x \in E\},$$
$$j_B(A) = \{\langle x, \nu_A(x) + \mu_B(x).\pi_A(x), \nu_B(x).\mu_A(x)\rangle \mid x \in E\},$$
$$j^*_B(A) = \{\langle x, \nu_A(x) + \mu_B(x).(1 - \nu_A(x) - \nu_B(x).\mu_A(x)),$$
$$\nu_B(x).\mu_A(x)\rangle \mid x \in E\},$$

where $B$ is a given IFS.

A series of new extensions of the modal operators were introduced in the last two years.

In modal logic both operators $\square$ and $\diamond$ are related to the following two equalities that are valid for each IFS $A$, too:

$$\square \neg A = \neg \diamond A,$$
$$\diamond \neg A = \neg \square A,$$

but no other connection between them has been observed. In the IFS-case, we can see that operators $D_\alpha$ and $F_{\alpha,\beta}$ ($\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$) are their direct extensions, because:

$$\square A = D_0(A) = F_{0,1}(A),$$
$$\diamond A = D_1(A) = F_{1,0}(A).$$

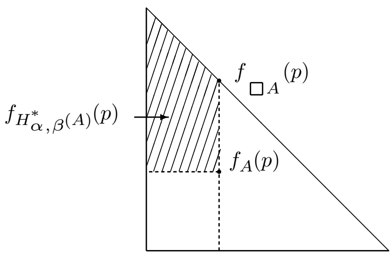
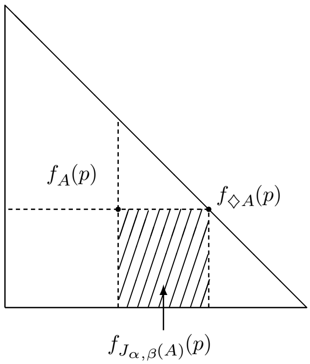



**Fig. 9** Geometric interpretation of $H^*_{\alpha,\beta}$

**Fig. 10** Geometric interpretation of $J_{\alpha,\beta}$




These equalities show a deeper interconnection between the two ordinary modal logic operators.

Some examples using the modal operators are described in [20, 21].

Let $s \geq 2$ be a fixed natural number. Following [36], the $s$-tuple $(X_1, \ldots, X_s)$, where $X_1, \ldots, X_s \in S = \{D_\alpha, F_{\alpha,\beta}, G_{\alpha,\beta}, H_{\alpha,\beta}, H^*_{\alpha,\beta}, J_{\alpha,\beta}, J^*_{\alpha,\beta}\}$ will be called a *basic s-tuple of operators* from $S$ if every one of the operators of $S$ can be represented by the operators of the $s$-tuple, using the above operations and the "composition" operation over operators.

**Theorem 1 [36]**: $(D, G), (F, G), (H, J), (H, J^*), (H^*, J)$ and $(H^*, J^*)$ are the only basic 2-tuples of operators.

**Corollary 1**: The only basic 3-tuples of operators are:

$$\begin{array}{lllll} (D, F, G), & (D, G, H), & (D, G, H^*), & (D, G, J), & (D, G, J^*), \\ (D, H, J), & (D, H, J^*), & (D, H^*, J), & (D, H^*, J^*), & (F, G, H), \\ (F, G, H^*), & (F, G, J), & (F, G, J^*), & (F, H, J), & (F, H, J^*), \\ (F, H^*, J), & (F, H^*, J^*), & (G, H, J), & (G, H, J^*), & (G, H^*, J), \\ (G, H^*, J^*), & (H, H^*, J), & (H, H^*, J^*), & (H, J, J^*), & (H^*, J, J^*). \end{array}$$

**Corollary 2**: $(D, F, H, H^*), (D, F, J, J^*)$ are the only 4-tuples of operators that are not basic 4-tuples of operators.

**Corollary 3**: All 5-tuples of operators are basic 5-tuples of operators.

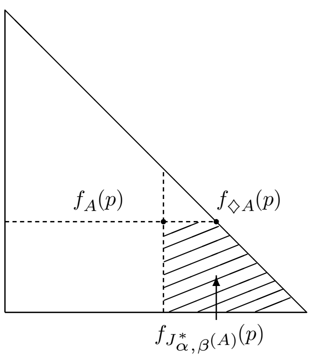



**Fig. 11** Geometric interpretation of $J^*_{\alpha,\beta}$

Two analogues of the topological operators can be defined over the IFSs, too: operator *"closure"* $C$ and operator *"intersection"* $I$:

$$C(A) = \{\langle x, \sup_{y \in E} \mu_A(y), \inf_{y \in E} \nu_A(y)\rangle | x \in E\},$$
$$I(A) = \{\langle x, \inf_{y \in E} \mu_A(y), \sup_{y \in E} \nu_A(y)\rangle | x \in E\}.$$

It is very interesting to note that the IFS-interpretations of both operators coincide, respectively, with the IFS-interpretations of the logic quantifiers $\exists$ and $\forall$ (see, e.g. [36]).

Let the IFS $A$ over the universe $E$ be called *proper*, if there exists at least one $x \in E$ for which $\pi_A(x) > 0$.

**Theorem 2 [36]**: Let $A, B$ be two proper IFSs for which there exist $y, z \in E$ such that $\mu_A(y) > 0$ and $\nu_B(z) > 0$. If $C(A) \subset I(B)$, then there are real numbers $\alpha, \beta, \gamma, \delta \in [0, 1]$, such that $J_{\alpha,\beta}(A) \subset H_{\gamma,\delta}(B)$.

Let

$$P_{\alpha,\beta}(A) = \{\langle x, \max(\alpha, \mu_A(x)), \min(\beta, \nu_A(x))\rangle | x \in E\},$$
$$Q_{\alpha,\beta}(A) = \{\langle x, \min(\alpha, \mu_A(x)), \max(\beta, \nu_A(x))\rangle | x \in E\},$$

for $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$.

The degrees of membership and non-membership of the elements of a given universe to its subset can be directly changed by these operators.

Obviously, for every IFS $A$ and for $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 1$:

$$P_{\alpha,\beta}(A) = A \cup \{\langle x, \alpha, \beta\rangle | x \in E\},$$
$$Q_{\alpha,\beta}(A) = A \cap \{\langle x, \alpha, \beta\rangle | x \in E\},$$
$$Q_{\alpha,\beta}(A) \subset A \subset P_{\alpha,\beta}(A).$$

**Theorem 3 [36]**: For every two IFSs $A$ and $B$, $C(A) \subset I(B)$, iff there exist two real numbers $\alpha, \beta \in [0, 1]$ such that $\alpha + \beta \leq 1$ and $P_{\alpha,\beta}(A) \subset Q_{\alpha,\beta}(B)$.

Here, for a first time we shall introduce extensions of the last two theorems for the extended modal operators from type $O_B(A)$, where $A$ and $B$ are IFSs and $O \in \{F, G, H, H^*, J, J^*, f, g, h, h^*, j, j^*\}$.

**Theorem 4**: Let $A, B$ be two proper IFSs for which there exist $y, z \in E$ such that $\mu_A(y) > 0$ and $\nu_B(z) > 0$. If $C(A) \subset I(B)$, then there are IFSs $X$ and $Y$, such that $J_X(A) \subset H_Y(B)$.

**Theorem 5**: For every two IFSs $A$ and $B$, $C(A) \subset I(B)$, iff there exists an IFS $X$ such that $P_X(A) \subset Q_X(B)$.

The two topological operators are generalized to the following four new ones (see [7]):

$$C_\mu(A) = \{\langle x, K, inf(1 - K, \nu_A(x))\rangle | x \in E\};$$
$$C_\nu(A) = \{\langle x, \mu_A(x), L\rangle | x \in E\};$$
$$I_\mu(A) = \{\langle x, k, \nu_A(x))\rangle | x \in E\};$$
$$I_\nu(A) = \{\langle x, inf(1 - l, \mu_A(x)), l\rangle | x \in E\},$$

where $K, L, k, l$ are defined above.

The weight operator $W$ is also defined over IFSs (see [16]) by

$$W(A) = \{\langle x, \frac{\sum\limits_{y \in E} \mu_A(y)}{card(E)}, \frac{\sum\limits_{y \in E} \nu_A(y)}{card(E)} \rangle | x \in E\},$$

where $card(E)$ is the number of the elements of the (finite) set $E$.

These operators have no analogues in fuzzy set theory and in the theories of the other fuzzy set extensions.

It will be interesting to mention that IFSs can be represented in the form $\langle A, B \rangle$, where $A$ and $B$ are ordinary fuzzy sets (see, e.g. [34, 23]). For this reason it may be wrongfully considered that IFSs are trivial extensions of ordinary fuzzy sets. Against such a claim we use the following argument. The statement above is analogous to the assertion that the set of the complex numbers is a trivial extension of the set of the real numbers.

All operators discussed above can be transformed for the IVFS case, too. On the other hand, using IFS-form, we can work easier with interval data, than with IVFS-form. Also, we can easily interpret interval data as points of the IFS-interpretation triangle. For example, let us have the set of intervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$. Let $A \leq \min a_i < \max b_i \leq B$. Of course, $A < B$, because otherwise for all $i$: $a_i = b_i$. Now, for interval $[a_i, b_i]$ we can construct numbers

$$\mu_i = \frac{a_i - A}{B - A},$$
$$\nu_i = \frac{B - b_i}{B - A}$$

that satisfy the condition $0 \leq \mu_i + \nu_i \leq 1$ and have the geometrical interpretation from Fig. 3. This idea is introduced for a first time in [2] and it is used in a series of joint research of V. Kreinovich, M. Mukaidono, H. Nguyen, B. Wu, M. Koshelev, B. Rachamreddy, H. Yasemis and the author (see, [36, 37, 35][3]).

Obviously, it is more convenient to work with points than with intervals. If the above points have the geometrical interpretation from Fig. 12, then by topological operators $C$ and $I$ we can determine points $U$ and $V$ and the region in which all above constructed points lie (see Fig. 13). If it is necessary, by operator $W$ we can find the point of the triangle that is the mass centre of this region.

Some colleagues have objections against the above assertion of mine, but in a future paper I plan to comment in more details the differences and usefulness of the IVFS and IFS.

Algorithms for separation of the different areas in the triangle are discussed in [18].

More interesting is the case (see [10]) when interval data are elements of two (or more) sets. Then we can obtain, e.g., IFS-geometrical interpretation from Fig. 14.

---

[3] One of the referees shown me also the paper [29].

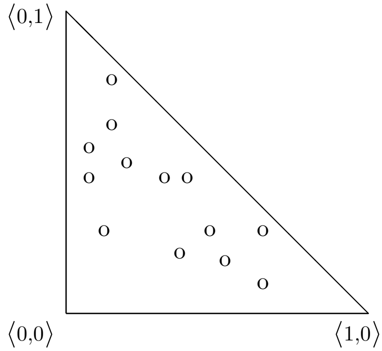



**Fig. 12** Puntual representation of intervals

and using topological operators, defined over IFS, we can separate these sets to the form from Fig. 15, using the algorithm from [18].

The concept of IVFS was extended in the sense of the IFS to *"interval-valued IFS"*. An IVIFS $A^*$ (for brevity $A$) over $E$ is an object of the form:

$$A = \{\langle x, M_A(x), N_A(x)\rangle \mid x \in E\},$$

where $M_A(x) \subset [0, 1]$ and $N_A(x) \subset [0, 1]$ are intervals and for all $x \in E$:

$$supM_A(x) + supN_A(x) \leq 1.$$

For the new sets it was shown that they have some essentially new properties different than the ordinary IFS.

Similar is the situation with the $L$-fuzzy sets. Really, each IFS can be interpreted as an $L$-fuzzy set, because the IFS-interpretation triangle can be interpreted as a complete lattice. On the other hand, many kinds of $L$-fuzzy sets have IFS-interpretation, while still many more *cannot* represent a given IFS. Following the idea of the $L$-fuzzy sets, in 1984 Stefka Stoeva and the author introduced the concept of *"Intuitionistic L-Fuzzy Set.* Later, in [24] Dogan Čoker proved that Pawlak's fuzzy rough sets are intuitionistic $L$-fuzzy sets, while Guo-jun Wang and Ying-Yu

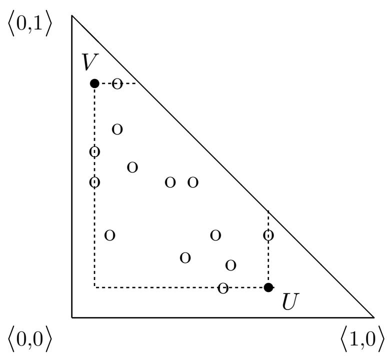



**Fig. 13** Points $U$ and $V$ and constructed region

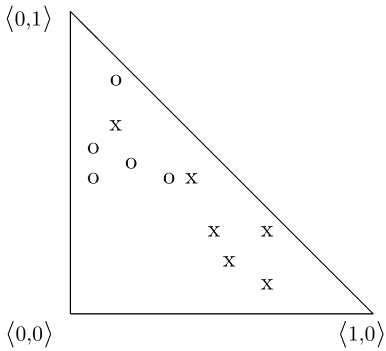



**Fig. 14** IFS-geometrical representation

He in [4], as well as Ofer Arieli, Chris Cornelis, Etienne Kerre and Glad Deschrijver in [1, 25, 26] discussed the nature of relations between $L$-fuzzy sets and IFSs.

In [36] there are IFSs over different universes and IFSs of type 2, for which (1) is changed to

$$\mu_A(x)^2 + \nu_A(x)^2 \leq 1.$$

It is clear that the latter inequality is a modification of the ordinary fuzzy set condition $\mu_A(x) \in [0, 1]$, as well as (1). Of course, we can continue in the direction of increasing the powers. Therefore, for a natural number $n \geq 2$ we can define IFSs of type $n$, for which (1) is changed to

$$\mu_A(x)^n + \nu_A(x)^n \leq 1.$$

This kind of extension of IFS is currently investigated by Peter Vassilev and the author. Other results in this direction were obtained by Arieli O., C. Cornelis, G. Deschrijver, E. Kerre, (see [1]), A. Pankowska and M. Wygdalak (see [42]) and others.

We can easily see that for every natural number $n \geq 2$, if a given set is an IFS of type $n$, then it is an IFS of type $n + 1$, but the opposite is not always valid.

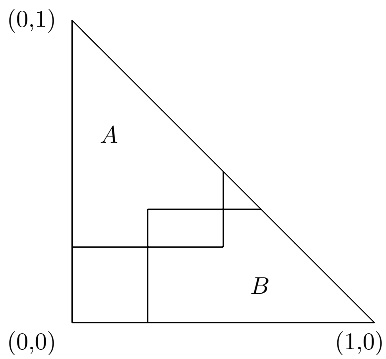



**Fig. 15** Set representation of Fig. 14

The IFSs of type $n$ have more theoretical sense, but it is not rather applicable for $n \geq 2$. To put it in other words, the ordinary IFS are more useful than the IFSs of type $n$, for $n \geq 2$. However, the situation with $L$-fuzzy sets is the same, when the lattice has a more complex form. Of course, each IFS can be interpreted as an $L$-fuzzy set, whose lattice $L$ has the form of the above triangle $F$. We can write

$$L^* = \{\langle a, b\rangle \mid a, b, a + b \in [0, 1]\}.$$

The opposite interpretation is more interesting. If we have a fixed $L$-fuzzy set and if $L$ is a discrete complete lattice, such that $L$ has exactly one minimal and exactly one maximal element, we can put it in a triangle $L^*$, so that its minimal vertex coincides with $L^*$-vertex $\langle 0, 1\rangle$, and its maximal vertex coincides with $L^*$-vertex $\langle 1, 0\rangle$. Now, having in mind its discrete nature, we can find such a region of the lattice, and respectively, of its $L^*$-interpretation, in which there are no lattice points, but there are IFS-triangle points. Therefore, the IFS will be richer than the fixed $L$-fuzzy set. If $L$ is a compact set that can be topologically included in $L^*$, then both (IFS- and $L$-fuzzy set) interpretations will have equivalent properties. The most complex case is when lattice $L$ has transfinite elements. In this case it is not clear whether we can put them in $L^*$, but I doubt the chance for success is big. Therefore, in this case, it is possible that the $L$-fuzzy set (with this special property of its lattice) will have more properties that the IFSs.

Some other extensions of the IFSs are introduced by S. Rizvi, H.J. Naqvi and D. Nadeem (see [43]), called *"rough IFSs",* by P.K. Maji, R. Biswas and A.R. Roy (see [38]), named *"intuitionistic fuzzy soft sets"* and by S. Samanta and T.K. Mondal (see [45]), called *"intuitionistic fuzzy rough sets"* and *"rough IFSs"* and others.

The author thinks that one of the most useful extensions of the IFS are so called *"temporal IFS"* (see Fig. 16) introduced in 1990 (see [5, 36]). All operations, relations and operators over IFS can be transferred to them, too. They have the form

$$A(T) = \{\langle x, \mu_A(x, t), \nu_A(x, t)\rangle | \langle x, t\rangle \in E \times T\},$$

where $E$ is a universe, $T$ is a non-empty set and
**(a)** $A \subset E$ is a fixed set,
**(b)** $\mu_A(x, t) + \nu_A(x, t) \leq 1$ for every $\langle x, t\rangle \in E \times T$,
**(c)** $\mu_A(x, t)$ and $\nu_A(x, t)$ are the degrees of membership and non-membership, respectively, of the element $x \in E$ at the time-moment $t \in T$.

The voting example could be much improved if we can test the social attitude to the respective parties and to the government at some time. Using temporal IFSs we can trace it in a longer period of time.

Similarly to fuzzy sets', IFS theory has different aspects, too.

The algebraic research within IFS theory is aimed at defining intuitionistic fuzzy subgroups, constructing the category $IFuz$ of IFS and other related categories. Intuitionistic fuzzy filters and ideals of lattices have also been introduced.

There are different aproaches to define IF numbers. P. Burillo, H. Bustince, V. Mohedano and M. Nikolova have worked in this area. T. Buhaescu has investigated interval valued real numbers.

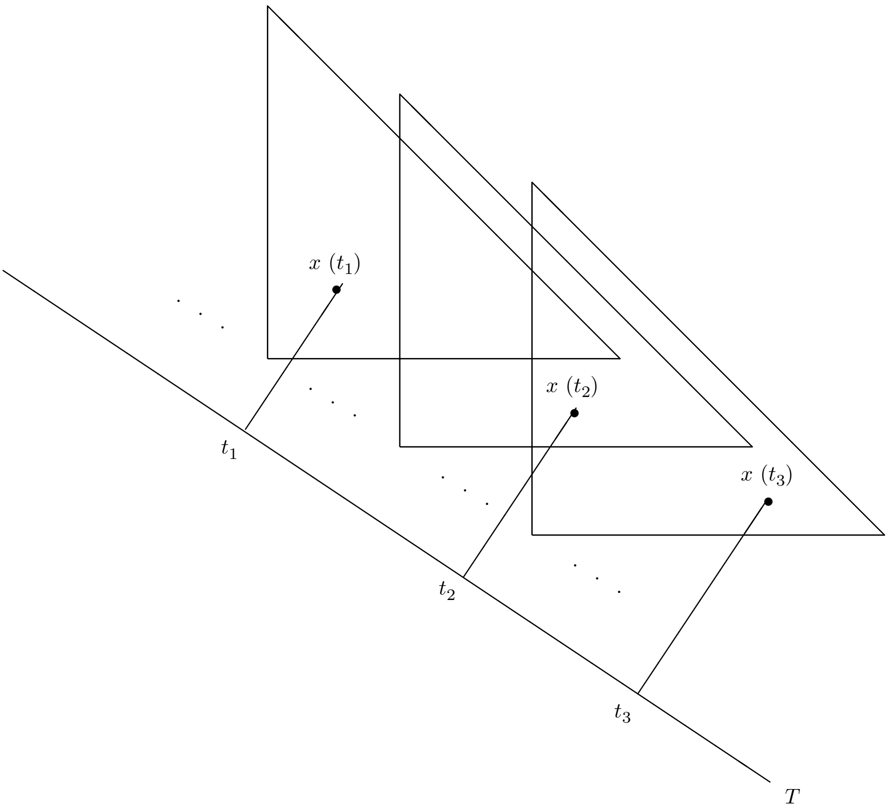



**Fig. 16** Representation of temporal IFS

Concepts of convexity and concavoconvexity for IFSs and temporal IFSs are introduced. The concept of intuitionistic (fuzzy) measure is defined. So far, A. Ban has introduced the limit of a sequence of IFSs. With the help of an abstract integral he has described a family of intuitionistic fuzzy entropies, introduced earlier by P. Burillo and H. Bustince. It is proved that certain intuitionistic fuzzy entropies are intuitionistic fuzzy measures. On this theme there are papers of T. Gerstenkorn, J. Manko, P. Burillo, H. Bustince, and others.

The notion of intuitionistic fuzzy metric space is presented by O. Ozbakir and D. Čoker. The solution concept for a semi linear equation with the fuzzy parameters studied by K. Peeva and S. Melliani.

A lot of research is devoted to *Intuitionistic Fuzzy Logic (IFL).* There are a lot of papers by G. Gargov, A. Ban, H. Bustince, E. Kerre, C. Cornelis, N. Nikolov, and the author in which intuitionistic fuzzy propositional and predicate calculus, intuitionistic fuzzy modal and temporal logic have been discussed. Norms and metrics over intuitionistic fuzzy logics and relations between the quantifiers and the modal type of operators in intuitionistic fuzzy logics have been studied. Rules of inference and the notion of intuitionistic fuzzy deductive closure are investigated.

Intuitionistic fuzzy model of the axioms of the paraconsistent set theory $NF_1$, intuitionistic logic and others are presented. Intuitionistic fuzzy interpretation of the conditional logic VW and Kun's axiom are made. It is proved that the Hauber's law is an intuitionistic fuzzy tautology.

In the last ten years IFS were applied in different areas. The IF-approach to artificial intelligence includes treatment of decision making and machine learning, neural networks and pattern recognition, expert systems, database, machine reasoning, logic programming and IF Prolog, Petri nets and generalized nets[4].

In the last ten years IFS were used in the process of decision making. Eulalia Szmidt and Janusz Kacprzyk, Humberto Bustince and Pedro Burillo, Adrian Ban and Cecilia Temponi, Gabriella Pasi, Ronald Yager and the author obtained interesting results in this direction. E. Szmidt and J. Kacprzyk extend the classical Bellman and Zadeh's general approach to decision making under fuzziness, originally termed decision making in a fuzzy environment, to the case of IFS.

Intuitionistic fuzzy versions of one of the basic statistical nonparametrical methods and the k-NN method, are proposed by Ludmila Kouncheva, Stefan Hadjitodorov, and others.

Currently, IFSs have applications in various areas. There are applications of IFSs in medical diagnosis and in decision making in medicine, developed by Anthony Shannon, Soon Ki-Kim, Eulalia Szmidt, Janusz Kacprzyk, Humberto Bustince, Joseph Sorsich and others.

Plamen Angelov has solved some optimization problems by means of intuitionistic fuzzy sets and has also worked on optimization in an intuitionistic fuzzy environment.

There are many applications of IFS in chemistry. Some more interesting of them are following: a method for simulation of complex technological system by use of IF generalized nets, an IF generalized net approach for optimal scheduling of iron ore delivering, discharge and blending yards creation and others.

IFS approach in credit risk assessment is proposed in a series of works by Dinko Dimitrov.

Olympia Georgieva and Tania Pencheva have described the key process variable and corrective actions of the waste water treatment plant with biosorption using the theory of IFSs. There are also IF generalized nets models of the gravitational field, in astronomy, sociology, biology, musicology, controllers, and others.

Intuitionistic fuzzy systems and IF abstract systems are defined and studied by Valentina Radeva, Hristo Aladjov and the author.

A first step to describe a theory of the IF-graphs and temporal IF-graphs was made by Anthony Shannon and the author. Application of IF-graphs and IF-relation methods have also been developed.

Of course, the list of the authors and their research is essentially longer and it will be an object of a new research, continuation of [41].

---

[4] Perhaps, the first paper that discusses the idea of fusing of fuzziness and Petri nets, is [3], published in 1985 in France.

## 3 Conclusion, or About the IFS-future

Like all young theories, the theory of IFSs contains a lot of open problems. While well-established theories contain famous problems with solutions that seem a matter of distant future, a number of "technical" problems persist in new theories, perhaps not so hard but requiring plenty of time and research effort. IFS theory is now at its beginning, and most of its problems are of the second type - nevertheless, there are some mathematical challenges of great interest to every researcher.

At the moment some students of mine are defining and studying the properties of operations over IFSs that are analogous to the new implications and negations.

In near future a new extension of the IFSs will be described: granulated IFSs, that are extension of L. Zadeh's idea for granulation. The research on the IFSs from $n$-th type ($n \geq 2$) will be published soon.

In [36, 8, 10] the author provided a list of open problems.

Finally, I will repeat with additional assurance the last words of my paper [10]:

*The author is an optimist for the future of the IFSs.*

**Acknowledgments** I would like to express gratitude to the Editors of the present book and the anonymous reviewes for their precions comments and remarks on the first draft of this paper. I would also thank to all colleagues, who have contributed to the development of the theory of intuitionistic fuzzy sets and their applications.

## References

1. Arieli, O., C. Cornelis, G. Deschrijver, E. Kerre, Relating intuitionistic fuzzy sets and interval-valued sets through bilattices. Proceedings of the 6th International FLINS Conference Applied Computational Intelligence, Blankenberge, 1–3 Sept. 2004, 57–64.
2. Asparoukhov, O., K. Atanassov, Intuitionistic fuzzy interpretation of confidencial intervals of criteria for decision making, Proc. of the First Workshop on Fuzzy Based Expert Systems (D. Lakov, Ed.), Sofia, Sept. 28–30, 1994, 56–58.
3. Atanassov, K. Generalized nets and their fuzzings, *AMSE Review,* Vol. 2 (1985), No. 3, 39–49.
4. Atanassov, K. Review and new results on intuitionistic fuzzy sets. *Preprint IM-MFAIS-1-88,* Sofia, 1988.
5. Atanassov, K. Temporal intuitionistic fuzzy sets. *Comptes Rendus de l'Academie bulgare des Sciences,* Tome 44, 1991, No. 7, 5–7.
6. Atanassov, K. *Intuitionistic Fuzzy Sets,* Springer Physica-Verlag, Berlin, 1999.
7. Atanassov, K. On four intuitionistic fuzzy topological operators. Mathware & Soft Computing, Vol. 8, 2001, 65–70.
8. Atanassov, K. Open problems in intuitionistic fuzzy sets theory. *Proceedings of 6-th Joint Conf. on Information Sciences,* Research Triangle Park (North Carolina, USA), March 8-13, 2002, 113–116.
9. Atanassov, K. "Twenty years later" or some words about the intuitionistic fuzzy sets. *Notes on Intuitionistic Fuzzy Sets,* Vol. 9 (2003), No. 3, 1–3.
10. Atanassov, K. Intuitionistic fuzzy sets: past, present and future. *Proc. of the Third Conf. of the European Society for Fuzzy Logic and Technology EUSFLAT' 2003,* Zittau, 10-12 Sept. 2003, 12–19.

11. Atanassov, K. Remarks on Intuitionistic fuzzy set theory. *Soft Computing Foundations and Theoretical Aspects* (K. Atanassov, O. Hryniewicz, J. Kacprzyk, Eds.), Academicka Oficyna Wydawnicza EXIT, Warszawa, 2004, 1–17.
12. Atanassov, K. On some types of intuitionistic fuzzy negations. *Notes on Intuitionistic Fuzzy Sets,* Vol. 11, 2005, No. 4, 170–172.
13. Atanassov, K. On some intuitionistic fuzzy negations. *Notes on Intuitionistic Fuzzy Sets,* Vol. 11, 2005, No. 6, 13–20.
14. Atanassov, K. On the implications and negations over intuitionistic fuzzy sets. *Proceedings of the Free University of Burgas Conference,* 9–11 June 2006 (in press).
15. Atanassov, K. On intuitionistic fuzzy negations and De Morgan's laws. accepted for IPMU'2006, Paris, 2006.
16. Atanassov, K., A. Ban. On an operator over intuitionistic fuzzy sets. Comptes Rendus de l'Academie bulgare des Sciences, Tome 53, 2000, No. 5, 39–42.
17. Atanassov, K., G. Gargov. Interval valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems,* Vol. 31, 1989, No. 3, 343–349.
18. Atanassov, K., J. Kacprzyk, E. Szmidt, L. Todorova. On separability of intuitionistic fuzzy sets. *Proc. of the 10-th Int. Fuzzy Systems Assoc. World Congress,* Istanbul, June 30 - July 2, 2003, In:- Lecture Notes in Artificial Intelligence, Vol. 2715: Fuzzy Sets and Systems - IFSA 2003, Springer, Berlin, 2003, 285–292.
19. Atanassov, K., V. Kreinovich. Intuitionistic fuzzy interpretation of intetrval data, *Notes on Intuitionistic Fuzzy Sets,* Vol. 5 (1999), No. 1, 1–8.
20. Atanassov, K., G. Pasi and R. Yager. Intuitionistic fuzzy interpretations of multi-criteria multi-person and multi-measurement tool decision making. International Journal of Systems Research, Vol. 36, 2005, No. 14, 859–868.
21. Atanassova, V. Strategies for decision making in the conditions of intuitionistic fuzziness. In: - Computational Intelligence, Theory and Applications (B. Reusch, Ed.), Springer, Berlin, 2005, 263–269.
22. Buhaescu, T. On the convexity of intuitionistic fuzzy sets, *Itinerant Seminar of functional equations, approximation and convexity,* Cluj-Napoca, 1988, 137–143.
23. Bustince, H., P. Burillo. Vague sets are intuitionistic fuzzy sets, *Fuzzy Sets and Systems,* Vol. 79, 1996, No. 3, 403–405.
24. Čoker, D. Fuzzy rough sets are intuitionistic L-fuzzy sets, *Fuzzy Sets and Systems,* Vol. 96, 1998, No. 3, 381–383.
25. Cornelis, C. Tweezijdigheid in de Representatie en Verwerking van Imprecieze Informatie. Universiteit Gent, Faculteit Wetenschappen, 2004.
26. Deschrijver, G., E. Kerre. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets and Systems,* Vol. 133 (2003), No. 2, 227–235.
27. Cornelis C., M. De Cock, E. Kerre Intuitionistic fuzzy rough sets: at the crossroads of imperfect knowledge. *Expert Systems,* Vol. 20 (5) (2003), No. 2, 260–270.
28. Dubois, D., H. Prade. Fuzzy Sets and Systems: Theory and Applications, Acad. Press, New York, 1980.
29. Gehrke, M. C. Walker, E. Walker. Some comments on interval-valued fuzzy sets. *International Journal of Intelligent Systems,* Vol. 11 (1996), 751–5759.
30. Georgiev, K. A simplification of the Neutrosophic Sets. Neutrosophic Logic and Intuitionistic Fuzzy Sets. *Notes on Intuitionistic Fuzzy Sets,* Vol. 11 (2005), 2, 28–31.
31. D. Gomez, J. Montero, H. Bustince "Sobre los conjuntos intuicionistas fuzzy de Atanassov", *XIII Congreso Español sobre Tecnologías y Lógica Fuzzy, ESTYLF'06* (In Spanish), Ciudad Real (Spain), September 2006.
32. Heyting, A. *Intuitionism. An Introduction.* North-Holland, Amsterdam, 1956.
33. Kaufmann, A. *Introduction a la théorie des sous-ensembles flous.* Masson, Paris, 1977.
34. Kondakov, N. *Logical Dictionary.*, Moskow, Nauka, 1971 (in Russian).
35. Koshelev, M., V. Kreinovich, B. Rachamreddy, H. Yasemis, K. Atanassov, Fundamental justification of intuitionistic fuzzy logic and interval-valued fuzzy methods. Proceedings of the Second International Conference on Intuitionistic Fuzzy Sets (J. Kacprzyk and K. Atanassov, Eds.), Vol. 1; *Notes on Intuitionistic Fuzzy Sets,* Vol. 4 (1998), No. 2, 42–46.

36. Kreinovich, V., M. Mukaidono, K. Atanassov. From fuzzy values to intuitionistic fuzzy values, to intuitionistic fuzzy intervals etc.: can we get an arbitrary ordering? Proceedings of the Third International Conference on Intuitionistic Fuzzy Sets (J. Kacprzyk and K. Atanassov, Eds.), Vol. 1; Notes on Intuitionistic Fuzzy Sets, Vol. 5 (1999), No. 3, 11–18.
37. Kreinovich, V., H. Nguyen, B. Wu, K. Atanassov, Fuzzy justification of heuristic methods in inverse problems and in numerical computations, with applications to detection of business cycles from fuzzy and intuitionistic fuzzy data. Proceedings of the Second International Conference on Intuitionistic Fuzzy Sets (J. Kacprzyk and K. Atanassov, Eds.), Vol. 2; Notes on Intuitionistic Fuzzy Sets, Vol. 4 (1998), No. 2, 47–56.
38. Maji, P.K., R. Biswas, A.R. Roy Intuitionistic fuzzy soft sets. *The Journal of Fuzzy Mathematics,* Vol. 9, 2001, No. 3, 677–692.
39. J.M. Mendel, Robert I. Bob John, "Type-2 fuzzy sets made simple", *IEEE Transactions on Fuzzy Systems*, 10(2) (2002), 117–127.
40. J.M. Mendel, "Advances in type-2 fuzzy sets and systems". *Information Sciences*, In Press, Available online 21 June 2006.
41. Nikolova, M., N. Nikolov, C. Cornelis, G. Deschrijver. Survey of the research on intuitionistic fuzzy sets. *Advanced Studies in Contemporary Mathematics,* Vol. 4, 2002, No. 2, 127–157.
42. Pankowska, A., M. Wygdalak. Intuitionistic fuzzy sets - an alternative look. Proc. of the Third Conf. of the European Society for Fuzzy Logic and Technology EUSFLAT' 2003, Zittau, 10–12 Sept. 2003, 135–140.
43. Rizvi, S., H.J. Naqvi, D. Nadeem. Rough intuitionistic fuzzy sets. *Proceedings of 6-th Joint Conf. on Information Sciences,* Research Triangle Park (North Carolina, USA), March 8-13, 2002, 101–104.
44. Rocha, L. M., V. Kreinovich, R. Baker Kearfott. Computing uncertainty in interval based sets, In: *Applications of Interval Computations* (R. Baker Kearfott and V. Kreinovich, Eds.), Kluwer Academic Publishers, Dordrecht, 1996, 337–380.
45. Samanta, S.K., T.K. Mondal. Intuitionistic fuzzy rough sets and rough intuitionistic fuzzy sets. *The Journal of Fuzzy Mathematics,* Vol. 9, 2001, No. 3, 561–582.
46. Smarandache, F. Neutrosophic set - a generalization of the intuitionistic fuzzy set. *International Journal of Pure and Applied Mathematics,* Vol. 24, No. 3, 2005, 287–297.
47. E. Szmidt. "Applications of Intuitionistic Fuzzy Sets in Decision Making", *D. Sc. dissertation* Techn. Univ., Sofia, (2000).
48. I.K. Vlachos, G.D. Sergiadis "Inner product based entropy in the intuitionistic fuzzy setting", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14 (3), 351–367 (2006).
49. Wang, G., He Y. Intuitionistic fuzzy sets and L-fuzzy sets, *Fuzzy Sets and Systems,* Vol. 110 (2000), No. 2, 271–274.
50. Zadeh, L. *The Concept of a Linguistic Variable and its Application to Approximate Reasoning.* American Elsevier Publ. Co., New York, 1973.

# Hybridization of Fuzzy and Rough Sets: Present and Future

**Eric C. C. Tsang, QingCai Chen, Suyun Zhao, Daniel S. Yeung and Xizhao Wang**

**Abstract** Though fuzzy set theory has been a very popular technique to represent vagueness between sets and their elements, the approximation of a subset in a universe that contains finite objects was still not resolved until the Pawlak's rough set theory was introduced. The concept of rough sets was introduced by Pawlak in 1982 as a formal tool for modeling and processing incomplete information in information systems. Rough sets describe the approximation of an arbitrary subset of a universe by two definable or observable subsets called lower and upper approximations. Even though Pawlak's rough set theory has been widely applied to solve many real world problems, the problem of being not able to deal with real attribute values had been spotted and found. This problem is originated in the crispness of upper and lower approximation sets in traditional rough set theory (TRS). Under the TRS philosophy, two nearly identical real attribute values are unreasonably treated as two different values. TRS theory deals with this problem by discretizing the original dataset, which may result in unacceptable information loss for a large amount of applications. To solve the above problem, a natural way of combining fuzzy sets and rough sets has been proposed. Since 1990's, researchers had put a lot of efforts on this area and two fuzzy rough set techniques that hybridize fuzzy and rough sets had been proposed to extend the capabilities of both fuzzy sets and rough sets. This chapter does not intend to cover all fuzzy rough set theories. Rather, it firstly gives a brief introduction of the state of the art in this research area and then goes into details to discuss two kinds of well developed hybridization approaches, i.e., constructive and axiomatic approaches. The generalization for equivalence relationships, the definitions of lower and upper approximation sets and the attribute reduction techniques based on these two hybridization frameworks are introduced in different sections. After that, to help readers apply the fuzzy rough set techniques, this chapter also introduces some applications that have successfully applied fuzzy rough set techniques. The final section of this chapter gives some remarks on the merits and problems of each fuzzy rough hybridization technique and the possible research directions in the future.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

# 1 Introduction

Modelling imprecise and uncertain information is one of the main research topics in the area of knowledge representation. Many approaches, such as fuzzy set theory and rough set theory etc., have been proposed to address the problems of uncertainty. Theories of fuzzy sets and rough sets are generalizations of classical set theory for modeling uncertainty. Fuzzy set theory deals with the ill-definition of the boundary of a class through a continuous generalization of set characteristic functions. Rough set theory takes into consideration the indiscernibility between objects. They are related but distinct and complementary with each other. Thus a natural question concerning possible connection between rough sets and fuzzy sets is raised. From a theoretical viewpoint, a new mathematical tool to deal with fuzziness and roughness is expected.

It is well known that Pawlak's rough set theory (traditional rough set theory) has been widely applied to solve many real world problems. However, the problem of being not able to deal with real attribute values had been spotted and found. This problem is originated in the crispness of upper and lower approximation sets in traditional rough set theory (TRS). Under the TRS philosophy, two nearly identical real attribute values are unreasonably treated as two different values. For example, for 3 attribute values $a(x) = 0.1$, $a(y) = 0.8$ and $a(z) = 0.101$, according to the definition of indiscernibility relation, the objects $x, y$ and $z$ are partitioned into different equivalence classes no matter how close $a(x)$ and $a(z)$ is. It is obviously unreasonable. Fuzzy set theory models the ambiguous membership between elements and classes. It treats the continuous values as the member of fuzzy sets (linguistic terms) with a certain membership. To a certain extent, the fuzzy sets and the corresponding membership describe the indiscernibility capability among values. Thus a natural way of combining fuzzy sets and rough sets to deal with real problems with nominal values is proposed. In other words, from a practical viewpoint a new methodology of combining fuzzy sets and rough sets for knowledge discovery and data mining is expected.

Since 1990's, researchers had put a lot of efforts on this area and fuzzy rough set techniques that hybridize fuzzy and rough sets had been proposed to extend the capabilities of both fuzzy sets and rough sets. This chapter does not intend to cover all fuzzy rough set theories. Rather, it firstly gives a brief introduction of the state of the art in this research area and then goes into details to discuss two kinds of well developed hybridization approaches, i.e., constructive and axiomatic approaches. The generalization for equivalence relations, the definitions of lower and upper approximations and the attribute reduction techniques based on these two hybridization frameworks are introduced in different sections. In order to help readers apply the fuzzy rough set techniques, this chapter also introduces some successful applications of fuzzy rough set techniques. The final section concludes this chapter and gives some possible research directions in the future.

## 2 Rough Sets

The traditional rough sets theory (TRS) was introduced by Pawlak in 1982 [19]. It has been proved to be a useful mathematical tool in describing and modeling incomplete and insufficient information. It has been widely used in the areas of Artificial Intelligence, Data Mining, Pattern Recognition, Fault Diagnostics, etc.

Firstly, we review some basic definitions[20] of TRS as follows.

**Definition 1.** *Let $U$, called the universe, be a nonempty set of objects, $A$ be a family of attributes. The pair of $(U, R)$ is called an approximation space (also called information system).*

**Definition 2.** *For the universe $U$ and any subset $B$ of $A$, if $IND(B) = \{(x, y) \in U \times U \,|\forall b \in B, b(x) = b(y)\}$. then $IND(B)$ is called the indiscernibility relation. Where $b(x)$ is the value of attribute $b$ for object $x$.*

Actually, $IND(B)$ is a binary relation $R \subseteq U \times U$ and it satisfies:

- Reflexivity, $\forall x \in U$, $R(x, x)$
- Symmetry, for $x, y \in U$, if $R(x, y)$then $R(y, x)$
- Transitivity, for $x, y, z \in U$, if $R(x, y)$ and $R(y, z)$, then $R(x, z)$

That is to say, the indiscernibility relation is an equivalence relation defined on $U$. This equivalence relation partitions the universe $U$ into disjoint subsets, each of which is called an equivalence class and is represented as $[x]_R = \{y \in U \,|R(x, y)\}$.

**Definition 3.** *A concept $X \subseteq U$is called definable if and only if it is a union of equivalence classes of $R$ [19].*

**Definition 4.** *Any concept $X \subseteq U$can be described by two definable sets $\underline{apr}_R X$ and $\overline{apr}_R X$ of $U$. The $\underline{apr}_R X$ and $\overline{apr}_R X$ called the lower and upper approximations of $X$ are defined as*

$$\underline{apr}_R X = \{x | x \varepsilon U, [x]_R \subseteq X\}$$

$$\overline{apr}_R X = \{x | x \in U, [x]_R \cap X \neq \phi\}$$

It is obvious that if $X \subseteq U$ is definable on $R$, then $\underline{apr}_R X = \overline{apr}_R X$. Otherwise, the concept $X$ is called "rough" and the nonempty set $\overline{apr}_R X - \underline{apr}_R X$ is called the boundary of $X$. There are several important properties of lower and upper approximations of $X$ that are treated as axiomatic properties of rough sets [16]:

- P1. $\underline{apr}_R U = U$ , $\overline{apr}_R \phi = \phi$
- P2. $\underline{apr}_R (X \cap Y) = \underline{apr}_R X \cap \underline{apr}_R Y$, $\overline{apr}_R (X \cup Y) = \overline{apr}_R X \cup \overline{apr}_R Y$
- P3. $\underline{apr}_R X^C = (\overline{apr}_R X)^C$, $\overline{apr}_R X^C = (\underline{apr}_R X)^C$
- P4. $\underline{apr}_R X \subseteq X \subseteq \overline{apr}_R X$
- P5. $X = \underline{apr}_R (\overline{apr}_R X)$ , $\overline{apr}_R (\underline{apr}_R X) = X$
- P6. $\underline{apr}_R X = \underline{apr}_R (\underline{apr}_R X)$ , $\overline{apr}_R (\overline{apr}_R X) = \overline{apr}_R X$

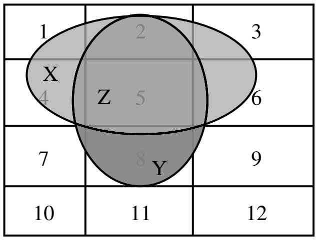



**Fig. 1** Illustration of set approximation

Here $X^C$is the complement set of $X$. Rather than formally prove the above properties, we use Fig. 1 to interpret these properties. In Fig. 1, the universe $U$ is represented by the whole zone and is partitioned into 12 equivalence classes by straight lines, $Z = X \cap Y$ each subset is marked by its index and is denoted by $U_i$, $i$ is the corresponding index number. The lower approximation of a subset is the union of rectangles that are fully contained by the subset and its upper approximation is the union of the rectangles that contain a part of the subset. For example, according to Fig. 1, since $\underline{apr}_R X = \underline{apr}_R Y = \underline{apr}_R Z = U_5$, the first equation of P2 is quite obvious in Fig. 1. For the second equation of P2, $\overline{apr}_R X = (\cup_{i=1,6} U_i) \cup U_8$, $\overline{apr}_R Y = \cup_{i=2,9} U_i$, and $\overline{apr}_R (X \cup Y) = \cup_{i=1,9} U_i = (\overline{apr}_R X) \cup (\overline{apr}_R Y)$. The reader can easily interpret other properties using Fig. 1.

An important application of rough set techniques is attribute reduction. A reduction of the attribute set $A$ is defined as a minimal subset $P \subseteq A$ that preserves the indescernibility relation of $A$. Here the "minimal" means that removing any attribute in $P$ will destroy the original indescernibility relation of $A$. In most cases, there is more than one reduction for $A$. The intersection of all reductions for $A$ is called the *core* of the information system $(U, A)$. By considering the combinational scale of possible reductions for attribute set $A$ and the complexity of verifying the equivalence of two indiscernibility relations on a large scale universe $U$, one may notice that the computing of reductions is not trivial. In fact, it is the bottleneck of rough set applications and many publications had focused on finding efficient algorithms for attribute reductions [1].

The TRS provides a useful tool for data mining and is especially suitable for the data analysis of discrete attribute value space. On the other hand, because of the property of indiscernibility relation mentioned above, it is difficult to directly apply TRS techniques on continuous value space. Since fuzzy set theory is good at describing the nominal values through fuzzy sets and membership, a natural way of combining fuzzy sets and rough sets has been proposed and is described in the following sections.

## 3 Fuzzy Rough Sets: Constructive and Axiomatic Approaches

Rough sets and fuzzy sets are complementary with each other in the sense that fuzzy sets model the ambiguous memberships between elements and classes while rough sets provide a way of approximating indefinable concept with a pair of definable sets within the universe. This observation motivated a lot of researchers to combine fuzzy sets and rough sets together and various kinds of fuzzy rough set models had been proposed in publications. From these publications, at least two kinds of fuzzy rough sets are highlighted. One is the constructive approach that starts with the fuzzy relations on the universe and the lower and upper approximation operators are constructed via these fuzzy relations [7, 8, 28]. The constructive approach of fuzzy rough sets is firstly proposed by Dubois and Prade [7, 8], and Radzikowska and Kerre [21] provide a more general approach of constructing fuzzy rough sets. The other is the axiomatic approach that treats the upper and lower approximations as the primitive notions, and a set of axioms have been studied and applied to characterize upper and lower approximations [15, 17, 18, 23, 24, 25, 27, 28, 30, 35]. When comparing with constructive approach, we know that the axiomatic approach is found to provide a deep view about the mathematical aspects of fuzzy rough sets. However, less has been known about its applications. To provide a clear introduction of fuzzy rough sets, in this section we firstly present the fuzzy logical operators see for example [15, 21], then we introduce each fuzzy rough set approach in detail.

## *3.1 Fuzzy Logical Operators*

**Definition 5.** *A, triangular norm or shortly t-norm, is any function* $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ *that satisfies the following conditions:*

- monotonicity: if $x < \alpha$, $y < \beta$, then $T(x, y) \leq T(\alpha, \beta)$
- commutativity: $T(x, y) = T(y, x)$
- associativity: $T(T(x, y), z) = T(x, T(y, z))$
- boundary condition: $T(x, 1) = x$

The most popular continuous *t*-norms include

- The standard *min* operator (the largest *t*-norm [8])$T_M(x, y) = \min\{x, y\}$;
- The algebraic product $T_P(x, y) = x \cdot y$ ;
- The bounded intersection (also called the Lukasiewicz *t*-norm) $T_L(x, y) = \max\{0, x + y - 1\}$.

**Definition 6.** *A triangular conorm, or shortly t-conorm, is an increasing, commutative and associative function* $S : [0, 1] \times [0, 1] \rightarrow [0, 1]$ *that satisfies the boundary condition* $\forall x \in [0, 1]$, $S(x, 0) = x$.

Three well-known continuous *t*-conorms are listed below:

- The standard *max* operator (the smallest *t*-conorm)

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

- The probabilistic sum $S_P(x, y) = x + y - x \cdot y$ ;
- The bounded sum $S_L(x, y) = \min\{1, x + y\}$.

**Definition 7.** *A negator N is a decreasing function* $N : [0, 1] \to [0, 1]$*that satisfies* $N(0) = 1$ *and* $N(1) = 0$*. A negator N is called involutive iff* $N(N(x)) = x$ *for all* $x \in [0, 1]$*; it is called weakly involutive iff* $N(N(x)) \geq x$ *for all* $x \in [0, 1]$*. The standard negator is defined as* $N_S(x) = 1 - x$*.*

**Definition 8.** *Given a negator N, a t-norm T and a t-conorm S are called dual with respect to N iff De Morgan laws are satisfied, i.e.* $S(N(x), N(y)) = N(T(x, y))$ *and* $T(N(x), N(y)) = N(S(x, y))$*.*

It is also known [15, 21] that given an involutive negator *N* and a *t*-conorm *S*, the function $T_S(x, y) = N(S(N(x), N(y)))$, $x, y \in [0, 1]$, is a *t*-norm such that *T* and *S* are dual with respect to *N*. It is referred to as a *t*-norm dual to *S* with respect to *N*.

Let $X : U \to [0, 1]$ be a fuzzy set and $F(U)$ be the fuzzy power set on *U*, i.e., the collection of all fuzzy sets on $U$ , for every $X \in F(U)$, the symbol $\text{co}_N X$ is used to denote the fuzzy complement of *X* determined by a negator *N*, i.e., for every$x \in U$, $(\text{co}_N X)(x) = N(X(x))$.

**Definition 9.** *An implicator is a function* $I : [0, 1] \times [0, 1] \to [0, 1]$ *satisfying* $I(1, 0) = 0$ *and* $I(1, 1) = I(0, 1)= I(0, 0) = 1$*. An implicator I is called left monotonic (right monotonic) iff for every* $x \in [0, 1]$*,* $I(\cdot, x)$ *is decreasing (*$I(x, \cdot)$ *is increasing). If I is both left and right monotonic, then it is called a hybrid monotonic implicator.*

Three classes of implicators [21] are defined below. Let *T*, *S*, *N* be t-norm, t-conorm and negator respectively. An implicator $I(x, y)$ is called

- an S-implicator based on *S* and *N* iff $I(x, y) = S(N(x), y)$;
- an R-implicator (residual implicator) based on continuous *T* iff for every $x, y \in [0, 1]$, $I(x, y) = \sup\{\sigma \,|\sigma \in [0, 1]\ , T(x, \sigma) < y\}$,
- an QL-implicator based on *T*, *S* and *N* iff for every $x, y \in [0, 1]$, $I(x, y) = S(N(x), T(x, y))$.

Some examples of S-implicator are

- the Lukasiewicz implicator based on $S_L$ and $N_S$:

$$I_L(x, y) = \min\{1, 1 - x + y\};$$

- the Kleene–Dienes implicator based on $S_M$ and $N_S$:

$$I_{KD}(x, y) = \max\{1 - x, y\};$$

- the Kleene–Dienes- Lukasiewicz implicator based on $S_P$ and $N_S$:

$$I_*(x, y) = 1 - x + x \cdot y$$

Examples of R-implicators:

- the Lukasiewicz implicator $I_L$, based on $T_L$;

- the Gödel implicator based on $T_M$ : $I_G(x, y) = \begin{cases} 1, & x \le y \\ y, & elsewhere \end{cases}$
- the Gaines implicator based on $T_P$ : $I_\Delta(x, y) = \begin{cases} 1, & x \le y \\ y/x & elsewhere \end{cases}$

Examples of QL-implicators:

- the Kleene–Dienes implicator $I_{KD}$
- the Early Zadeh implicator $I_Z(x, y) = \min\{1 - x, \ \min\{x + y\}\}$

All implicators given above belong to a more general implicator class, i.e. the border implicators. Formally, an implicator $I$ is called a border implicator iff for every $x \in [0, 1]$, $I(1, x) = x$.

## 3.2 Constructive Approach of Fuzzy Rough Sets

The fuzzy rough sets were firstly proposed by Dubois and Prade [7, 8], and then Radzikowska and Kerre[21] provided general constructive approaches based on different kinds of fuzzy logical operators. This subsection will firstly introduce the general definition of fuzzy rough approximation and the fuzzy rough sets. Then the Dubois and Prade's fuzzy rough sets model is given as an instance of the general model.

While the Pawlak's TRS is based on the equivalence relation, as an extension, in the fuzzy rough sets the equivalence relation is replaced by a fuzzy equivalence relation defined as follows[21].

**Definition 10.** *A fuzzy binary relation R on universe U is called a fuzzy equivalence relation if R satisfies the reflexivity (*$R(x, x) = 1$*), symmetry (*$R(x, y) = R(y, x)$*) and the sup-min transitivity (*$R(x, y) \ge \sup \min_{z \in U} \{R(x, z), R(z, y)\}$*).*

The nonempty universe $U$ and the fuzzy equivalence relation $R$ defined on it compose of the *fuzzy approximation space* (FAS)$F_{AS} = (U, R)$. The corresponding equivalence class $[x]_R$ with $x \in U$ is defined as a fuzzy set $[x]_R(y) = R(x, y)$ on $U$. Based on fuzzy approximation space, the definition of fuzzy rough approximations (FRA) is given below.

**Definition 11.** *Given a border implicator I, t-norm T and the power set* $F(U)$ *of universe U, the fuzzy lower and upper approximations of a fuzzy set* $X \in F(U)$ *are defined as two fuzzy sets with respect to I, T and the fuzzy approximation space* $F_{AS} = (U, R)$ *as*

$$\underline{F_{AS}}_I(X)(x) = \inf_{y \in U} I(R(x, y), X(y)) \tag{1}$$

$$\overline{F_{AS}}^T(X)(x) = \sup_{y \in U} T(R(x, y), X(y))$$

The fuzzy rough approximations $\underline{F_{AS}}_I(X)$ and $\overline{F_{AS}}^T(X)$ are called $I$-lower and $T$-upper fuzzy rough approximations of $X$ in $F_{AS}$.

**Definition 12.** *Given a FAS $F_{AS} = (U, R)$, a border implicator I and a t-norm T, a pair of fuzzy sets $(A_L, A_U)$ $(A_L, A_U) \in F(U) \times F(U)$ is called a (I, T)-fuzzy rough set in $F_{AS}$ iff $(A_L, A_U) = (\underline{F_{ASI}}(X), \overline{F_{AS}}^T(X))$ for some fuzzy sets $X \in F(U)$.*

As mentioned in papers [7, 8] that the Dubois and Prade's fuzzy rough sets (DP-FRS) can be thought as an instance of the general fuzzy rough sets for which the border implicator $I$ and the $t$-norm $T$ are specified as the $I_{KD}$ and $T_M$ respectively. In this case, the symbols for fuzzy rough sets can be simply denoted as a fuzzy sets pair $(R_*(X),\ R^*(X))$ for fuzzy equivalence relation $R$ and fuzzy set $X \in F(U)$. Here the $R_*(X)$ and $R^*(X)$ are defined as

$$\begin{aligned} R_*(X)(x) &= \inf_{y\in U} \max\{1 - R(x, y), X(y)\} \\ R^*(X)(x) &= \sup_{y\in U} \min\{R(x, y), X(y)\} \end{aligned} \tag{2}$$

Some properties of the DP-FRS that may be comparable with the properties P1-P6 of TRS are listed below:

- FP1) $R_*U = U$, $R^*\phi = \phi$;
- FP2) $R_*(X \cap Y) = R_*X \cap R_*Y$, $R^*(X \cup Y) = R^*X \cup R^*Y$;
- FP3) $R_*X^C = (R^*X)^C$, $R^*X^C = (R_*X)^C$, $here\ A^c(x) = 1 - A(x)$;
- FP4) $R_*X \subseteq X \subseteq R^*X$;
- FP5) $R_*(U - \{y\})(x) = R_*(U - \{x\})(y)$, $R^*x_1(y) = R^*y_1(x)$;
- FP6) $R_*X \subseteq R_*(R_*X)$, $R^*(R^*X) \subseteq R^*X$.

Here $X, Y \in F(U),\ x, y \in U$ and $x_\gamma(y) = \begin{cases} \gamma, & y = x \\ 0, & y \neq x \end{cases}$

For the properties of the general fuzzy rough sets and fuzzy rough approximations, we refer the readers to [21].

### 3.3 Axiomatic Approach of Fuzzy Rough Sets

By comparing with the constructive approach, we know that the axiomatic approach aims to investigate the mathematical characteristics of fuzzy rough sets rather than to develop methods for applications. The axiomatic approaches help us to gain much more insights into the mathematical structures of fuzzy approximation operators. In crisp rough set theory, the axiomatic approaches of approximation operators had been studied in details [5, 6, 16, 18, 26, 29]. The most important axiomatic studies were performed by Yao [31, 32, 33], Yao and Lin [34] in which various classes of rough sets algebras are characterized by different sets of axioms. The research of axiomatic approach has also been extended to approximation operators in fuzzy environment. There are many research works on the axiomatization on fuzzy rough sets. Here we only review two kinds of axiomatic approaches of fuzzy rough sets. One is proposed by Moris and Yakout [18], which is a pioneering work in studying a set of axioms on fuzzy rough sets. The other is proposed by Wu et al. [27, 28], in which various classes of fuzzy rough approximation operators is characterized by different sets of axioms.

In the paper [18], the fuzzy approximation operators are characterized as follows. Let $X$ denote the universe of discourse, $M$ denote the unit interval [0, 1]. The axiomatic characterizations of fuzzy rough sets defined by fuzzy equivalence relation are given below.

**Definition 13. (fuzzy upper approximation operator)** *An operator $\bar{A}$ $M^X$ (from $M^X$ to $M^X$) is said to be a fuzzy upper approximation operator on X if it satisfies the following axioms for all $\mu \in M^X$, $x, y \in X$ and all $\alpha \in M$ :*

- $(\bar{A}1)\bar{A}\mu \geq \mu$
- $(\bar{A}2)\bar{A}\ \bar{A}\mu = \bar{A}\mu$
- $(\bar{A}3)\bar{A}(\vee_{j\in J}\mu_j) = \vee_{j\in J}\bar{A}\mu_j$
- $(\bar{A}4)\overline{A}(1_y)(x) = \overline{A}(1_x)(y)$
- $(\bar{A}5)\bar{A}(\underline{\alpha}T\mu) = \underline{\alpha}T\bar{A}\mu$, T is a $t$-norm, $\underline{\alpha}(x) = \alpha$ for all $x \in X$.

**Definition 14. (fuzzy upper approximation operator)** *An operator $\bar{A}$ on $M^X$ (from $M^X$ to $M^X$) is said to be a fuzzy lower approximation operator on X if it satisfies the following axioms for all $\mu \in M^X$, $x, y \in X$ and all $\alpha \in M$ :*

- $(\underline{A}1)\underline{A}\mu \leq \mu$
- $(\underline{A}2)\underline{A}\ \underline{A}\mu = \underline{A}\mu$
- $(\underline{A}3)\bar{A}(\wedge_{j\in J}\mu_j) = \wedge_{j\in J}\bar{A}\mu_j$
- $(\underline{A}4)\underline{A}\vartheta_T\left|1_x, \alpha\right|(x) = \underline{A}\vartheta_T\left|1_y, \alpha\right|(y)$
- $(\underline{A}5)\underline{A}\vartheta_T\left|\underline{\alpha}, \mu\right| = \vartheta_T\left|\underline{\alpha}, \underline{A}\mu\right|$

Here $\vartheta_T\left|A, B\right|(x) = \vartheta_T(A(x), B(x))$.

However, fuzzy rough sets in [21] is restricted to fuzzy rough sets defined by fuzzy T-similarity relations. Thiele [23, 24, 25] has investigated axiomatic characterizations of fuzzy rough approximation operators and rough fuzzy approximation operators within modal logic for fuzzy diamond and box operators. The important axiomatic studies for fuzzy rough sets are done by Wu et al. [27, 28] who studied generalized rough set approximation operators in fuzzy environment, They examined many axioms on various classes of fuzzy rough approximation operators. The minimal axiom sets of fuzzy approximation operators guarantee the existence of certain types of fuzzy or crisp relations producing the same operators.

The fuzzy rough approximation operators defined by Wu et al. are introduced in the following.

In axiomatic approach, rough sets are axiomatized by abstract operators (the lower and upper approximation operators). For the case of fuzzy rough sets, the primitive notion is a pair of fuzzy sets operators $L, H : F(W) \to F(U)$, where $W$ and $U$ are two different universes. In their fuzzy rough sets, various classes of fuzzy approximation operators are characterized by different sets of axioms. The minimal axiom sets of fuzzy approximation operators guarantee the existence of certain types of fuzzy or crisp relations producing the same operators.

**Definition 15.** *Let $L, H : F(W) \rightarrow F(U)$ be two operators. They are referred to as dual operators if for all $A \in \boldsymbol{F}(W)$,*

- (FL1)$L(A) = \sim H(\sim A)$,
- (FU1)$H(A) = \sim L(\sim A)$.

**Definition 16.** *Let $L, H : F(W) \rightarrow F(U)$ be a pair of dual operators. If L satisfies axioms $L(\hat{\alpha} \vee A) = \hat{\alpha} \vee L(A)$ and $L(A \wedge B) = L(A) \wedge L(B)$ or equivalently, H satisfies axioms $H(\hat{\alpha} \wedge A) = \hat{\alpha} \wedge H(A)$ and $H(A \vee B) = H(A) \vee H(B)$, here $\hat{a}$ is just $\underline{\alpha}$, defined in Definition 13, then L and H are referred to as* ***fuzzy rough approximation operators.***

**Definition 17.** *Let L, H: F(W) → $F(U)$ be a pair of dual operators. If L satisfies axioms $L(1_{W-\{y\}}) \in \boldsymbol{F}(U), \forall y \in W, L(\hat{\alpha} \vee A) = \hat{\alpha} \vee L(A)$ and $L(A \wedge B) = L(A) \wedge L(B)$ or equivalently, H satisfies axioms $H(1_y) \in \boldsymbol{F}(U), \forall y \in W, H(\hat{\alpha} \wedge A) = \hat{\alpha} \wedge H(A)$ and $H(A \vee B) = H(A) \vee H(B)$, then L and H are referred to as* ***rough fuzzy approximation operators.***

Where $1_y$ denotes the fuzzy singleton with value 1 at y and 0 elsewhere. Similarly, $1_{W-\{y\}}$ denotes the fuzzy singleton with value 0 at y and 1 elsewhere.

For completeness, we summarize the main results of axiomatic characterizations of the fuzzy rough approximation operators corresponding to the different fuzzy binary relations in Theorem 1-Theorem 5 [28].

**Theorem 1.** *Suppose that $L, H : F(W) \rightarrow F(U)$ is a pair of dual fuzzy approximation operators, then there exists a serial fuzzy relation R from U to W such that $L(A) = \underline{R}(A)$ and $H(A) = \overline{R}(A)$ for all $A \in \boldsymbol{F}(W)$ iff L and H satisfy axioms:*

- (FL3.1) $L(\phi) = \phi$,
- (FU3.1) $H(W) = U$,
- (FLU3.1) $L(A) \subseteq H(A), \forall A \in \boldsymbol{F}(W)$.

**Theorem 2.** *Suppose that $L, H : F(U) \rightarrow F(U)$ is a pair of dual fuzzy rough approximation operators, then there exists a reflexive fuzzy relation R on U such that L(A) = $\underline{R}$ (A) and H(A) = $\overline{R}(A)$ for all $A \in \boldsymbol{F}(U)$ iff L and H satisfy axioms:*

- (FL3.2) $L(A) \subseteq A, \forall A \in \boldsymbol{F}(U)$,
- (FU3.2) $A \subseteq H(A), \forall A \in \boldsymbol{F}(U)$.

**Theorem 3.** *Suppose that $L, H : F(U) \rightarrow F(U)$ is a pair of dual fuzzy rough approximation operators, then there exists a symmetric fuzzy relation R on U such that L(A) = R (A) and H(A) = $\underline{R}(A)$ for all $A \in \boldsymbol{F}(U)$ iff L and H satisfy axioms:*

- (FL3.3) $L(1_{U-\{x\}})(y) = L(1_{U-\{x\}})(x), \forall (x, y) \in U \times U, \forall A \in \boldsymbol{F}(U)$,
- (FU3.3) $H(1_x)(y) = H(1_y)(x), \forall (x, y) \in U \times U$.

**Theorem 4.** *Suppose that L, H:F(U) → F(U) is a pair of dual fuzzy rough approximation operators, then there exists a transitive fuzzy relation R on U such that $L(A) = \underline{R}(A)$ and $H(A) = \overline{R}(A)$ for all $A \in \boldsymbol{F}(U)$ iff L and H satisfy axioms:*

- (FL3.4) $L(A) \subseteq L(L(A)), \forall A \in \boldsymbol{F}(U)$,
- (FU3.4) $H(H(A)) \subseteq H(A) \forall A \in \boldsymbol{F}(U)$.

**Theorem 5.** *Suppose that $L, H : F(U) \to F(U)$ is a pair of dual fuzzy rough approximation operators, then there exists an Euclidean crisp relation $R$ on $U$ such that $L(A) = \underline{R}(A)$ and $H(A) = \overline{R}(A)$ for all $A \in \boldsymbol{F}(U)$ $iff$ $L$ and $H$ satisfy axioms:*

- (FL3.5) $H(L(A)) \subseteq L(A), \forall A \in \boldsymbol{F}(U)$,
- (FU3.5) $H(A) \subseteq L(H(A)), \forall A \in \boldsymbol{F}(U)$.

The minimization and the independence of the axiom set for fuzzy approximation operators have also been investigated. Minimal axiom sets corresponding to various generalized fuzzy approximation operators are presented. For these cases, we refer the readers to papers [30, 35].

## 4 Attribute Reduction of Fuzzy Rough Sets

Nowadays a large amount of data are collected in real world. There is a clear need for reducing the large data in many applications such as pattern recognition, machine learning etc. It is well known that rough set theory has been proved to be a useful tool to deal with uncertainty and has been applied in data reduction, rule extraction and granularity computation. However, it works only with nominal values. In real world it is most often the case that there are both nominal and numerical values. To address this problem many researchers have developed techniques for attributes reduction with both nominal and numerical values using fuzzy rough sets. Roughly speaking, there are two types of fuzzy RSAR (Rough Set Attributes Reduction) approach. One is based on the notion of fuzzy lower approximation to enable attributes reduction. The other is based on the fuzzy significance of attributes to reduce attributes. In the following we will introduce these two types of methods of fuzzy RSAR.

### *4.1 Fuzzy RSAR Building on the Notion of Fuzzy Lower Approximation*

Recently, there are several approaches [2, 3,12, 13, 14, 22] of fuzzy RSAR, which build on the notion of fuzzy lower approximation. A pioneering work of Fuzzy RSAR is proposed in papers [12, 13, 14, 22]. A QUICKREDUCT algorithm to find the close-to-minimal attributes reduction is given. The key idea of this method to reduce attributes is to keep the dependency function after reduction invariant.

Before introducing the concept of dependency function, the notions of lower and upper approximations, different with the ones proposed by Dubois and Prade [7, 28], is given.

In subsection 3.2, the Dubois and Prade's notions of fuzzy rough sets have been given based on the fuzzy equivalence relation. From the literature [7], the fuzzy lower and upper approximations based on the fuzzy partition ${}^{U}/_{P}$ are given as

$$m_{\underline{P}X}(F_i) = \inf_{x \in U} \max\{1 - m_{F_i}(x), m_X(x)\} \quad (3)$$
$$m_{\overline{P}X}(F_i) = \sup_{x \in U} \min\{m_{F_i}(x), m_X(x)\}$$

Where $^U/_P$ is the collection of fuzzy sets $m_{[x]_R}(y) = m_X(x, y), y \in U$ which makes a fuzzy partition of $U$. These fuzzy rough sets are just equivalent to the ones in [7, 8]. The denotation $F_i$, called a fuzzy equivalence class, belongs to $^U/_P$. These definitions diverge a little from the crisp lower and upper approximations, as the membership of individual objects are not explicitly available [12]. As a result, the lower and upper approximations are redefined in paper [12] as follows.

**Definition 18.** *The fuzzy lower and upper approximations are defined as follows.*

$$\mu_{\underline{P}X}(x) = \sup_{F \in ^U/_P} \min\{\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\} \quad (4)$$
$$\mu_{\bar{P}X}(x) = \sup_{F \in ^U/_P} \min\{\mu_F(x), \sup_{y \in U} \min\{\mu_F(y), \mu_X(y)\}$$

We need to point out that the notions of lower and upper approximations proposed by Dubois and Prade (formula FR3) and ones proposed by Jensen and Shen (formula FR4) are both defined on the fuzzy partition. The former describes the relation between the fuzzy equivalence class and fuzzy approximations while the latter describes the relation between the individual objects and the fuzzy approximations.

After redefining the fuzzy lower and upper approximations in paper [12], the fuzzy positive region is introduced as follows.

**Definition 19.** *The membership of an object $x \in U$, belonging to the fuzzy positive region of $Q$ relative to $P$ can be defined by $\mu_{POS_p(Q)}(x) = \sup_{X \in U/Q} \mu_X(x)$, here $Q$ is a fuzzy similarity relation as a decision attribute.*

Using the above definition 19, the dependency function can be defined as follows.

$$\gamma'_P(Q) = \frac{\left|\mu_{POS_p(Q)}(x)\right|}{|U|} = \frac{\sum_{x \in U} \mu_{POS_p(Q)}(x)}{|U|}$$

Here $\left|\mu_{pos_P(Q)}(x)\right|$ mean the cardinal of the fuzzy set having membership function $\mu_{pos_P(Q)}(\bullet)$

By keeping the dependency function invariant after reduction, a QUICKREDUCT algorithm found in paper [12] is given in Fig. 2. The algorithm is depicted below.

This algorithm has been tested with some practical data sets such as web categorization and is claimed to perform well [12, 22]. However, the time complexity of the algorithm in papers [12, 13, 14, 22] increases exponentially with the inputting attributes because of the way of calculating the fuzzy equivalence classes. This makes the algorithm infeasible on large dimensional problems. Therefore the concept of

**Fuzzy RSAR algorithm 1:**

**Input:** U—the universe, C—the condition attributes set, D—the decision attributes set. $\gamma'_B(D)$ is the dependency function of $B \subseteq C$ with respect to D.

**Output:** $R$—the close-to-minimal attributes reduct.

**Step 1:** Initialize R=$\phi$, $\gamma'_{best} = 0$, $\gamma'_{prev} = 0$

**Step 2: Do**

$T = R$

$\gamma'_{prev} = \gamma'_{best}$

For each $a \in (C - R)$

If $\gamma'_{R\cup\{a\}}(D) > \gamma'_{\tau}(D)$

$T = R \cup \{a\}$

$\gamma'_{best} = \gamma'_{\tau}(D)$

$R = T$

Until $\gamma'_{best} = \gamma'_{prev}$

**Step 3: Output** $R$

**Fig. 2** The QUICKREDUCT algorithm based on fuzzy lower approximation

fuzzy rough sets in papers [2, 3] is put forward on a compact computational domain, which is then utilized to improve the computational efficiency.

The definition of fuzzy lower approximation on a compact domain is given as follows.

**Definition 20.** *Given an arbitrary fuzzy set* $\mu_A(x) : U \to [0,1]$, *and* $F_{ik} \in U/P$ *The fuzzy lower approximation on compact computational domain is defined by*

$$\mu_{\underline{A}}(F_{ik}) = \begin{cases} \inf\limits_{x \in D_{\underline{A}}(F_{ik})} \max\{\mu_{\overline{F_{ik}}}(x), \mu_A(x)\} & ; D_{\underline{A}}(F_{ik}) \neq 0 \\ 1 & ; D_{\underline{A}}(F_{ik}) = 0 \end{cases}$$

Where $D_{\underline{A}}(F_{ik}) \subset U$ is a compact computational domain for lower approximation defined as $D_{\underline{A}}(F_{ik}) = \{x \in U \mid \mu_{F_{ik}}(x) \neq 0 \wedge \mu_A(x) \neq 1\}$. Here $\wedge$ is a logical AND connective.

All these papers [2, 3, 12, 13, 14, 22] have never proposed a clear definition of attributes reduction with fuzzy rough sets and the structure of reduction is never discussed. Thus a method of attributes reduction with solid mathematical foundation such as clear structure of reduction and precise reduction process with mathematical support based on fuzzy rough sets is still expected.

## *4.2 Fuzzy RSAR Based on Fuzzy Attribute Significance*

There are also some methods of fuzzy RSAR based on fuzzy attribute significance [9, 10, 11]. These methods firstly propose one kind of information measure of indiscernibility relation and then based on these information measures the fuzzy attributes significance, i.e. the information increment, is measurable. Subsequently, the algorithm to find the reduction based on the fuzzy attribute significance is constructed. Owing to the fact that the concepts of fuzzy attribute significance are the key concepts of these, we focus on the introduction of the notions which relate to the notion of fuzzy attributes significance in this subsection.

In the fuzzy RSAR methods proposed by Hu and Yu [9, 10], a novel interpretation of Yager's entropy is proposed as the measure of the discernibility power of attributes. The Yager's entropy of an indiscernibility relation is defined as follows.

**Definition 21.** *Given a finite set of objects $X = \{x_1, x_2, \cdots, x_n\}$, P is the probability distribution on X and R is the fuzzy indiscernibility or equivalence relation on $X$. The Yager's entropy is defined as:*

$$H_p(R) = \sum\nolimits_{xX} -P(x)\log_2(\sum\nolimits_{y\in Y} P(y)R(x,y))$$

Generally, we assume that P is uniformly distributed on $X$ and $p(x) = {}^1/_{|X|}, X \in X[11]$. Thus the Yager's entropy can also be represented as:

$$H_P(R) = -\frac{1}{n}\sum\nolimits_{i=1}^{n}\log_2\frac{\sum_{j=1}^{n} R(x_i, x_j)}{n} = -\frac{1}{n}\sum\nolimits_{i=1}^{n}\log_2\frac{|[x_i]_R|}{n}$$

Where $[x_i]_R$ is one of the fuzzy equivalence classes of $U/R$.

Some extensions of Yager' entropies, such as joint entropy, condition entropy and relative entropy, are also proposed by Hu and Yu [9, 10] which are necessary to introduce the fuzzy attribute significance.

**Definition 22.** *Given a finite set of objects $X = \{x_1, x_2, \cdots, x_n\}$. Let $R_1$ and $R_2$ be two indiscernibility relations on X. P is the probability distribution on X and $p(x) = {}^1/_{|X|}, x \in X$.*

(1) The **joint entropy** of $R_1$ and $R_2$ is defined as:

$$H_P(R_1R_2) = -\frac{1}{n}\sum\nolimits_{i=1}^{n}\log_2\frac{\left|[x_i]_{R_1}\cap[x_i]_{R_2}\right|}{n}$$

(2) The **conditional entropy** of $R_2$ with respect to $R_1$ is defined as:

$$H_P(R_1 \mid R_2) = -\frac{1}{n}\sum\nolimits_{i=1}^{n}\log_2\frac{\left|[x_i]_{R_1}\cap[x_i]_{R_2}\right|}{\left|[x_i]_{R_2}\right|}$$

(3) Let $R = R_1$ È $R_2$, then the **relative entropy** of $R_1$ with respect to $R_2$ is defined as

$$\begin{aligned} H_P(R_1; R_2) = H_P(R_2) - H_P(R) &= H_P(R_2) - H_P(R_1 \cup R_2) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{\left|[x_i]_{R_2}\right|}{\left|[x_i]_{R_1}\right|\cup\left|[x_i]_{R_2}\right|} \end{aligned}$$

**The fuzzy RSAR algorithm 2:**
**Input:** U—the universe, C—the condition attributes set, D—the decision attributes set
**Output:** $R$—the close-to-minimal attributes reduct.
**Step 1: Initialize** R=$\phi$
**Step 2:** $a \in A$, **compute** the equivalence relation.
**Step 3:** For each $_{a\in(C-R)}$
    **Compute** $^{-}H_i = SIG(a|R; D)$
  Add the significant attribute $q$ to $R$:
    $R = R \cup \{q\}$
**Step 4:** For current $q$ and $R$,
    If SIG($q|R$; $D$) $\geq 0$, **goto** step 3; otherwise **goto** step 5.
**Step 5:** Output $R$.

**Fig. 3** The fuzzy RSAR algorithm based on the fuzzy attributes significance

Then conditional entropy and relative entropy are used to measure the information increment, which is the fuzzy attribute significance. The definition of fuzzy significance of attributes in papers [9, 10] is defined as follows.

**Definition 23.** *Given a fuzzy information system* $(U, A, V, f)$*,* $A = C \cup d$*, where C is the condition attributes set and d is the decision attribute.* $B \subseteq C,\ \forall a \in B$*. The fuzzy significance of attribute a in attribute set B relative to d is defined as*

$$
\begin{aligned}
SIG(a\,|B-a\,;d) &= H_P(a\,|B-a;\,d) = H_P(B-a;\,d) - H_P(B;\,d)\\
&= -H_P((B-a)\cup d) + H_P(B\cup d)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2\left|[x_i]_d \cup [x_i]_{B-a}\right| + \frac{1}{n}\sum_{i=1}^{n}\log_2|[x_i]_d \cup [x_i]_B|
\end{aligned}
$$

Here the subset of attributes corresponds to a fuzzy indiscernibility relation. For example, the attribute set $B$ corresponds to a fuzzy indiscernibility relation $B$.

The fuzzy RSAR algorithm based on fuzzy attributes significance is shown in Fig. 3.

## 5 Application of Fuzzy Rough Sets: One Illustrative Example of Fuzzy FRAR

Fuzzy rough sets have been applied in attributes reduction, rule extraction and granularity computation, such as medical time series, case generation, stock price mining, descriptive dimensionality reduction and so on. Attributes reduction is one important application of fuzzy rough sets. In this section the method of fuzzy FRAR proposed by Shen and Jensen [12, 13, 14, 22] is selected to demonstrate the process of fuzzy attributes reduction. We have already mentioned the basic concepts of

**Table 1** A training set of attribute reduction

| U | A | | B | | C | | Q |
|---|---|---|---|---|---|---|---|
| | FA1 | FA2 | FB1 | FB2 | FC1 | FC2 | |
| 1 | 0.8 | 0.2 | 0.6 | 0.4 | 1 | 0 | No |
| 1 | 0.8 | 0.2 | 0.6 | 0.4 | 1 | 0.6 | No |
| 2 | 0.8 | 0.2 | 0 | 0.6 | 0.2 | 0.8 | Yes |
| 3 | 0.6 | 0.4 | 0.8 | 0.2 | 0.6 | 0.4 | No |
| 4 | 0 | 0.4 | 0.6 | 0.4 | 0 | 1 | Yes |
| 5 | 0 | 0.6 | 0.6 | 0.4 | 0 | 1 | Yes |
| 6 | 0 | 0.6 | 0 | 1 | 0 | 1 | No |

attributes reduction and the algorithm of this method is given in Sect. 4.1. Here we cite one example in [12] to demonstrate the process of fuzzy FRAR of Shen and Jensen's method.

*Example 1.* Table 1 is a fuzzy information system which is adopted from [8]. In Table 1 U ={1, 2, 3, 4, 5, 6} is a set of fuzzified data. A, B and C are condition attributes and Q is the decision attribute. {FA1, FA2}, {FB1, FB2}, {FC1, FC2}, which are the fuzzy sets defined on the universe $U$, describe the corresponding fuzzy attributes A, B, C, respectively.

Firstly, we need to point out that the family $\Phi = \{F_1, F_2, \cdots, F_n\}$ of normal fuzzy sets defined on the universe of $U$ can be viewed as a weak fuzzy partition if they satisfy the following two conditions.

$$\inf_{x \in U} \max_i (\mu_{F_i}(x)) > 0$$

$$\forall i, j, \ \sup_{x \in U} \min_i (\mu_{F_i}(x), \mu_{F_j}(x)) < 1$$

$\{F_1, F_2, \cdots, F_n\}$ can play the role of a collection of fuzzy equivalence classes [8].

For instance, the family of fuzzy sets {FA1, FA2} in Table 1 can be viewed as a fuzzy partion of attribute A on $U$. Here {FA1, FA2} can play the role of a collection of fuzzy equivalence classes. In Shen and Jensen method {FA1, FA2}, {FB1, FB2}, {FC1, FC2} are viewed as the collections of fuzzy equivalence classes of attributes A, B, C on $U$, respectively.

The following are the detailed illustrative process to find the fuzzy attributes reduct using Shen and Jensen's method.

1. From the Table 1, the following fuzzy equivalence classes are obtained:
   $U/A = \{FA1, FA2\}, U/B = \{FB1, FB2\}$,

   $U/C = \{FC1, FC2\}, U/Q = \{\{1, 3, 6\}, \{2, 4, 5\}\}$.

2. According to the definition of fuzzy lower approximation, the first step is to compute the fuzzy lower approximation of the sets of A, B and C, respectively. For simplicity, only the attribute A is considered here. For the first decision class {1,3,6}, the fuzzy lower approximation requires $\mu_{\underline{A}\{1,3,6\}}(x)$ to be calculated as:

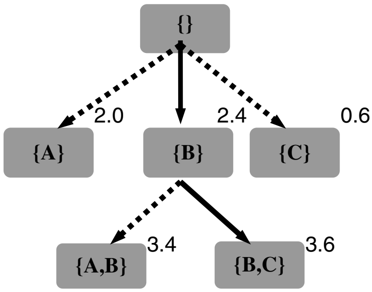



**Fig. 4** Path taken by the QUICKREDUCT algorithm

$$\mu_{\underline{A}\{1,3,6\}}(x) = \sup_{F \in U/A} \min\{\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_{\underline{A}\{1,3,6\}}(y)\}\}$$

Calculating the $A$ - fuzzy lower approximation of {1,3,6} for every object gives:

$\mu_{\underline{A}\{1,3,6\}}(1) = 0.2,\quad \mu_{\underline{A}\{1,3,6\}}(2) = 0.2,\quad \mu_{\underline{A}\{1,3,6\}}(3) = 0.4$

$\mu_{\underline{A}\{1,3,6\}}(4) = 0.4,\quad \mu_{\underline{A}\{1,3,6\}}(5) = 0.4,\quad \mu_{\underline{A}\{1,3,6\}}(6) = 0.4$

3. The corresponding values for { 2,4,5} can also be determined in this way. Using these values, the fuzzy positive region for each object can be calculated via using $\mu_{POS_A(Q)}(x) = \sup_{X \in U/Q}$ .
   The results are

   $\mu_{POS_A Q}(1) = 0.2,\quad \mu_{POS_A Q}(2) = 0.2,\quad \mu_{POS_A Q}(3) = 0.4$
   $\mu_{POS_A Q}(4) = 0.4,\quad \mu_{POS_A Q}(5) = 0.4,\quad \mu_{POS_A Q}(6) = 0.4$

4. It is coincident here that $\mu_{POS_A Q}(\mathrm{x}) = \mu_{\underline{A}\{1,3,6\}}(\mathrm{x})$

   for this example. The next step is to calculate the degree of dependency of Q on A:

$$\gamma'_A(Q) = 2/6$$

By calculating the dependency degree of B and C, we have:

$$\gamma'_B(Q) = 2.4/6,, \gamma'_C(Q) = 1.6/6.$$

From these results, it can be seen that the attribute $B$ will cause the greatest increase in dependency degree. This attribute is chosen and added to the potential reduct. The process iterates and the two dependency degrees are calculated as follows.

$$\gamma'_{\{A,B\}}(Q) = 3.4/6, \gamma'_{\{B,C\}}(Q) = 3.2/6$$

Adding attribute A to the reduct candidate causes the largest increase in dependency degree. So the new candidate becomes $\{A, B\}$. Lastly, the attribute $C$ is added to the reduct candidate and causes no increase in dependency degree:

$$\gamma'_{\{A,B,c\}}(Q) = 3.4/6.$$

The algorithm stops and outputs the reduct: {A, B, C}.
The steps taken by the QUICKREDUCT algorithm can be seen in Fig. 4.

## 6 Conclusion and Future Work

Rough set theory deals with the approximation of sets under indiscernibility [19, 20]. Fuzzy set theory handles the vague concepts by allowing partial membership [36]. They are different but complementary generalizations of classical set theory. Recently, many researches have been proposed to hybridize rough set theory and fuzzy set theory to fuzzy rough sets. Roughly speaking, there are two types of well developed fuzzy rough set approaches. One is the constructive approach, which generalizes the fuzzy relations on the universe and the lower and upper approximation operators are constructed via these fuzzy relations. The constructive approach is suitable for the applications of rough sets. The other is the axiomatic approach, which treats the lower and upper approximations as the primitive notions and focuses on algebra systems of rough set theory. The axiomatic is appropriate for studying the structure of rough set algebra. One important application of rough set theory is that of attribute reduction in databases. Some studies have been reported on the attributes reduction of fuzzy rough sets and found to have achieved good performance in some applications. From the above introduction of fuzzy rough sets, it is easy to see that fuzzy rough sets have a solid mathematical foundation and work well in some applications. However, there are still some problems that need to be addressed in the future. They are:

- The framework that works well in both constructive and axiomatic systems and also performs well in real applications is expected. For example, a fuzzy rough set model that has the formal structure of knowledge reduction and performs well in real problems is expected.
- So far rough sets are only suitable for static data. A fuzzy rough set dealing with problems that can be handled instantly is expected.
- There are many kinds of uncertain information in the real world, such as roughness, fuzziness, randomness etc. A framework that hybridizes all these uncertainties in a model is expected.

**Acknowledgment** This work is supported by the Hong Kong Research Grant Council CERG research grants No. B-Q826 and B-Q943.

## References

1. Bautista R., Millan M. and Diaz J. F. (1999) An efficient implementation to calculate relative core and reducts. NAFIPS' 99, pp. 791–794.
2. Bhatt R. B., Gopal M. (2005) On fuzzy rough sets approach to feature selection, Pattern recognition Letters, Vol. 26, no.7, pp. 965–975.

3. Bhatt R. B., Gopal M. (2005) On the compact computational domain of fuzzy-rough sets, pattern recognition letter, Vol. 26 no.11, pp. 1632–1640.
4. Boixader D., Jacas J., and Recasens J. (1999) Upper and lower approximations of fuzzy sets, International Journal General System, Vol. 29, no. 4, pp. 555–568.
5. Comer S. (1991) An algebraic approach to the approximation of information, Fundamenta Informaticae Vol. 14, pp. 492–502.
6. Comer S. (1993) On connections between information systems, rough sets, and algebraic logic, in: C. Rauszer (Ed.), Algebraic Methods in Logic and Computer Science, Vol. 28, Banach Center Publisher, Polish Academy of Sciences, pp. 117–127.
7. Dubois D., Prade H. (1990) Rough fuzzy sets and fuzzy rough sets, International Journal General System, Vol. 17, no. 2–3, pp. 191–209.
8. Dubois D., Prade H. (1992) Putting rough sets and fuzzy sets together, in: R. Slowinski (Ed.), Intelligent Decision Support, Kluwer Academic Publishers, Dordrecht, pp. 203–232.
9. Hu Q. H., Yu D. R. (2005) Entropies of fuzzy indiscernibility relation and its operations, international journal of uncertainty, fuzziness and knowledge-based systems, Vol. 12, no. 5, pp. 575–589.
10. Hu Q. H., Yu D. R., Xie Z. X. (2006) Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognition Letters Vol. 27, pp. 414–423.
11. Hu Q. H., Yu D. R., Xie Z. X., Liu J.F. (2006) Fuzzy probabilistic approximation spaces and their information measures, IEEE Transactions on fuzzy systems, Vol. 14, no. 2, pp. 191–201.
12. Jensen R., Shen Q. (2004) Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets and Systems, Vol. 141, pp. 469–485.
13. Jensen R., Shen Q. (2004) Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Transaction on Knowledge and Data Engineering, Vol. 16, no. 12, pp. 1457–1471.
14. Jensen R., Shen Q. (2005) Fuzzy-rough data reduction with ant colony optimization, fuzzy sets and systems Vol. 149, pp. 5–20.
15. Klir G. J., Yuan B. (1995) Fuzzy Logic: Theory and Applications, Prentice-Hall, Englewood CliIs, NJ.
16. Lin T. Y., Liu Q. (1994) Rough approximate operators: axiomatic rough set theory, in: W. Ziarko (Ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer, Berlin, pp. 256–260.
17. Mi J. S., Zhang W. X. (2004) An axiomatic characterization of a fuzzy generalization of rough sets, Information Sciences Vol. 160, pp. 235–249.
18. Morsi N. N., Yakout M. M. (1998) Axiomatics for fuzzy rough set, Fuzzy Sets Systems, Vol. 100, pp. 327–342.
19. Pawlak Z. (1982) Rough Sets. International Journal of Computer and Information Sciences. Vol. 11, pp. 341–356.
20. Pawlak Z. (1991) Rough Sets: Theoretical Aspects of Reasoning about Data [M]. Dordrecht: Kluwer Academic Publishers.
21. Radzikowska M., Kerre E. E. (2002) A comparative study of fuzzy rough sets, Fuzzy Sets and Systems, Vol. 126, pp. 137–155.
22. Shen Q., Jensen R., (2004) Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, Pattern Recognition, Vol. 37, pp. 1351–1363.
23. Thiele H. (2000) On axiomatic characterisation of fuzzy approximation operators I, the fuzzy rough set based case, RSCTC 2000, Banff ParkLodge, Bariff, Canada, 19–19 October 2000, Conference Proceedings, pp. 239–247.
24. Thiele H. (2001) On axiomatic characterisation of fuzzy approximation operators II, the rough fuzzy set based case, Proceedings of the 31st IEEE International Symposium on Multiple-Valued Logic, pp. 330–335.
25. Thiele H. (2001) On axiomatic characterization of fuzzy approximation operators III, the fuzzy diamond and fuzzy box cases, The 10th IEEE International Conference on Fuzzy Systems Vol. 2, pp. 1148–1151.
26. Wiweger R. (1989) On topological rough sets, Bulletin of Polish Academy of Sciences: Mathematics Vol. 37, pp. 89–93.

27. Wu W. Z., Zhang W. X. (2003) Generalized fuzzy rough sets. Information Sciences. Vol. 15, pp. 263–282.
28. Wu W. Z., Zhang W. X. (2004) Constructive and axiomatic approaches of fuzzy approximation operators, Information Sciences Vol. 159, pp. 233–254.
29. Wybraniec-Skardowska U. (1989) On a generalization of approximation space, Bulletin of the Polish Academy of Sciences: Mathematics Vol. 37, pp. 51–61.
30. Yang X. P. (2005) The Minimization of Axiom Sets Characterizing Generalized Fuzzy Rough Approximation Operators, fuzzy systems and knowledge discovery second international conference, FSKD 2005, Changsha, China, August 27–29,2005, Proceeding, Part I, pp. 1303–1308.
31. Yao Y. Y. (1996) Two views of the theory of rough sets in finite universes, International Journal of Approximate Reasoning Vol. 15, 1996, pp. 291–317.
32. Yao Y. Y. (1998) Constructive and algebraic methods of the theory of rough sets, Journal of Information Sciences Vol. 109, pp. 21–47.
33. Yao Y. Y. (1998) Generalized rough set model, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery 1. Methodology and Applications, Physica-Verlag, Heidelberg, pp. 286–318.
34. Yao Y. Y., Lin T. Y. (1996) Generalization of rough sets using modal logic, Intelligent Automation and Soft Computing, an International Journal Vol. 2, pp. 103–120.
35. Yeung D. S., Chen D. G., Tsang E. C. C., Lee J. W. T., Wang X. Z. (2005) On the Generalization of Fuzzy Rough Sets, IEEE Transactions on Fuzzy Systems, Vol. 13, no. 3, pp. 343–361.
36. Zadeh L. A. (1965) Fuzzy sets, Information Control, Vol. 8, 1965, pp. 338–353.
37. Zakowski W. (1982) On a concept of rough sets, Demonstratio Mathematica XV 1129–1133.

# An Overview of Computing with Words using Label Semantics

**Jonathan Lawry**

**Abstract** This chapter will give an overview of the label semantics framework for computing with words. Label semantics is an alternative methodology that models linguistic labels in terms of label descriptions, appropriateness measures and mass assignments. It has a clear operational semantics and is straightforward to integrate with other uncertainty formalisms. In the chapter we will introduce the theory of label semantics, discuss its compatibility with a functional calculus and show how it can be used to infer both probabilistic and possibilistic information from linguistic assertions.

## 1 Introduction to Computing with Words

The principle aim of the *computing with words* paradigm as proposed by Zadeh [21] is to increase the use of natural language for information and knowledge processing in computer systems. In practice this will require the development of a formal representation framework based on some restricted subset of natural language. Zadeh has suggested a form of *precisiated natural language* [18] based on the theory of generalised constraints and linguistic variables. *Label semantics*, introduced by Lawry [7, 8, 9, 10], provides an alternative representation for computing and modelling with words, which takes a somewhat different perspective than Zadeh on the processing of natural language information. This paper provides an overview of computing with words based on label semantics and contrasts it with Zadeh's approach. In this section we briefly describe Zadeh's methodology.

Zadeh's approach is based fundamentally on the notion of a linguistic variable [17, 18, 19] where a semantic rule links natural language labels to an underlying graded vague concept as represented by a fuzzy set on the domain of discourse.

**Definition 1.** *Linguistic Variable*
*A linguistic variable is a quintuple $\langle L, T(L), S, \Omega, M\rangle$ in which $L$ is the name of the variable, $\Omega$ is a universe of discourse, $T(L)$ is a term set of words to describe elements of $\Omega$, $S$ is a set of syntactic rules for generating new elements of $T(L)$ from existing words and $M$ a set of semantic rules which identify a fuzzy subset of $\Omega$ for each word in $T(L)$, corresponding to its fuzzy extension.*

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

The term set $T(L)$ is generated recursively from an initial set of labels by application of the syntactic rules $S$ which include both logical operators ($\wedge, \vee, \neg$ etc) and linguistic hedges such as *quite, very* etc. The semantic rules $M$ are based on the standard truth-functional fuzzy logic calculus so that given mappings for the initial labels it then holds that $\forall \theta_1, \theta_2 \in T(L)$ $M(\neg\theta_1) = 1 - M(\theta_1)$, $M(\theta_1 \wedge \theta_2) = \min(M(\theta_1), M(\theta_2))$ and $M(\theta_1 \vee \theta_2) = \max(M(\theta_1), M(\theta_2))$. Zadeh, also defines semantic rules for linguistic hedges so that, for example, if $\theta \in T(L)$ then $M(very\ \theta) = M(\theta)^2$ and $M(quite\ \theta) = M(\theta)^{0.5}$. As with the logical connectives, the semantic rules for hedges can then be applied recursively so that, for example, $M(very^n\ \theta) = M(\theta)^{2n}$ where $very^n$ denotes $n$ occurrences of the hedge $very$.

Words $\theta \in T(L)$ describe elements of the underlying universe $\Omega$ according to a system of generalised constraints [22], whereby an expression '$x$ is $\theta$' referring to some unknown element $x \in \Omega$ conveys the information that the value of $x$ has a possibility distribution $M(\theta)$ (see Zadeh [20]). In this context possibility distributions can be viewed an upper probability of $x$ given $\theta$ [15].

Fuzzy logic can be criticised for the lack of a clear operational interpretation of membership functions. In other words, it is difficult to understand the exact meaning of the numerical values $M(\theta)(x)$. This weakens model transparency and makes knowledge elicitation much more difficult. In fact there have been several attempts to provide an operational semantics for fuzzy logic (see [2] for an overview), many of which give interesting and useful interpretations of fuzzy sets. However, none of them are compatible with a truth functional calculus. As part of our discussion of labels semantics, in the following section we present a calculus which, though while not truth-functional, is functional in a slightly weaker sense.

## 2 Label Semantics

In contrast to fuzzy sets and linguistic variables, *label semantics* encodes the meaning of linguistic labels according to how they are used by a population of communicating agents to convey information. From this perspective, the focus is on the decision making process an intelligent agent must go through in order to identify which labels or expressions can actually be used to describe an object or value. In other words, in order to make an assertion describing an object in terms of some set of linguistic labels, an agent must first identify which of these labels are appropriate or assertible in this context. Given the way that individuals learn language through an ongoing process of interaction with the other communicating agents and with the environment, then we can expect there to be considerable uncertainty associated with any decisions of this kind. Furthermore, there is a subtle assumption central to the label semantic model, that such decisions regarding appropriateness or assertibility are meaningful. For instance, the fuzzy logic view is that vague descriptions like 'John is *tall*' are generally only partially true and hence it is not meaningful to consider which of a set of given labels can truthfully be used to described John's height. However, we contest that the efficacy of natural language as a means of conveying information between members of a population lies in shared conventions governing the appropriate use of words which are, at least loosely, adhere to by individuals within the population. We explore this idea in more detail in the following subsection.

## 2.1 The Epistemic Stance

It cannot be denied that in their use of linguistic labels humans posses a mechanism for deciding whether or not to make assertions (e.g. John is *tall*) or to agree to a classification (e.g. Yes, that is a tree). Further, although the concepts concerned are vague this underlying decision process is fundamentally crisp (bivalent). For instance, you are either willing to assert that $x$ is a *tree* given your current knowledge, or you are not. In other words, either tree is an appropriate label to describe $x$ or it is not. As humans we are continually faced with making such crisp decisions regarding vague concepts as part of our every day use of language. Of course, we may be uncertain about labels and even express these doubts (e.g. I'm not sure whether you would call that a tree or a bush, or both) but the underlying decision is crisp.

Given this decision problem, we suggest that it is useful for agents to adopt what might be called an *epistemic stance* as follows:

> *Each individual agent in the population assumes the existence of a set of labelling conventions, valid across the whole population, governing what linguistic labels and expression can be appropriately used to describe particular instances.*

Of course, such linguistic conventions do not need to be imposed by some outside authority like the Oxford English Dictionary or the Academia Lengua Espanola, but instead would emerge as a result of interactions between agents each adopting the epistemic stance. Hence, label semantics does not attempt to link label symbols to fuzzy set concept extensions but rather to quantify an agent's subjective belief that a label $L$ is appropriate to describe an object $x$ and hence whether or not it is reasonable to assert that '$x$ is $L$'. In this respect it is close to the 'anti-representational' view of vague concepts proposed by Rohit Parikh [11] which focuses on the notion of assertibility rather than that of truth; a view that is shared by Alice Kyburg [5]. There is also some similarity between the label semantics view and the epistemic theory of vague concepts as proposed by Williamson [16]. In this theory it is assumed that there is a true but unknown (or only partially known) definition of a vague concept and hence that the subsequent uncertainty regarding concept boundaries is purely epistemic in nature. However, while in label semantics the uncertainty regarding the appropriateness of labels is also assumed to be epistemic, there is not assumed to be some objectively correct definition of a vague concept. Instead it is assumed that the actual rules or conventions for label use emerge from the interaction between individuals each adopting the epistemic stance.

## 2.2 Appropriateness Measures

Label semantics proposes two fundamental and inter-related measures of the appropriateness of labels as descriptions of an object or value. Given a finite set of labels $LA$ from which can be generated a set of expressions $LE$ through recursive applications of logical connectives, the measure of appropriateness of an expression $\theta \in LE$ as a description of instance $x$ is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective belief that $\theta$ can be used to describe $x$ based on his/her (partial) knowledge of the current labelling conventions of the population. From an alternative perspective, when faced with an object to describe, an agent may consider each label in $LA$ and attempt to identify the subset of labels that are appropriate to use. Let this set be denoted by $\mathcal{D}_x$. In the face of their uncertainty regarding labelling conventions the agent will also be uncertain as to the composition of $\mathcal{D}_x$, and in label semantics this is quantified by a probability mass function $m_x : 2^{LA} \rightarrow [0, 1]$ on subsets of labels. The relationship between these two measures will be described below.

Unlike linguistic variables, which allow for the generation of new label symbols using a syntactic rule, label semantics assumes a finite set of labels $LA$. These are the basic or core labels to describe elements in an underlying domain of discourse $\Omega$. Based on $LA$, the set of label expressions $LE$ is then generated by recursive application of the standard logic connectives as follows:

**Definition 2.** *Label Expressions*
*The set of label expressions of $LA$, $LE$, is defined recursively as follows:*

- *If $L \in LA$ then $L \in LE$*
- *If $\theta, \varphi \in LE$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi, \theta \rightarrow \varphi \in LE$*

So for example, if $\Omega$ is the set of all possible $rgb$ values and $LA$ is the set of basic colour labels such as, *red, yellow, green, blue, orange* etc then $LE$ contains those compound expressions such as *red & yellow*, *not blue nor orange* etc. Typically we assume that the overlapping vagueness of the labels $LA$ means that for some elements of $\Omega$ close to concept boundaries it is appropriate within language conventions to use a number of labels. For instance, some colours can be appropriately described both as red and as yellow.

Hence, in label semantics we preserve a clear distinction between the basic labels $LA$ and the compound label expressions $LE$ generated from them. This distinction is fundamentally important for the subsequent development of a calculus for measures of appropriateness. Also, note that while we allow logical combinations of label expression, linguistic hedges are not included in the definition of $LE$. We argue that the semantics of such hedges seem far from clear and the rather arbitrary definitions given in [17, 18, 19] appear inadequate. Indeed, in our view, it is far from apparent that there should be a simple functional relationship between the meaning of a word and the meaning of a new word generated from it by applying a hedge. In other words, we would claim that while hedges are a simple syntactic device for generating new terms there is no equally simple semantic device for generating the associated new meanings. Furthermore, it is unclear if hedges can really be applied in a recursive manner. For example, we might describe John as *very tall* or even *very*

*very tall* but repeat application of the same hedge soon renders the expression meaningless. Also, there would seem to be significant restrictions on those expressions to which a particular hedge can be applied. For example, we would never say *very quite tall* or *very (tall and medium)*. Hence, while label semantics permits hedges to be used in the definition of labels in $LA$ it is only as a syntactic device for naming labels.

A mass assignment $m_x$ on sets of labels then quantifies the agent's belief that any particular subset of labels contains all and only the labels with which it is appropriate to describe $x$.

**Definition 3.** *Mass Assignment on Labels*
*$\forall x \in \Omega$ a mass assignment on labels is a function $m_x : 2^{LA} \to [0, 1]$ such that* $\sum_{S \subseteq LA} m_x(S) = 1$

Now depending on labelling conventions there may be certain combinations of labels which cannot all be appropriate to describe any object. For example, *small* and *large* cannot both be appropriate. This restricts the possible values of $\mathcal{D}_x$ to the following set of focal elements:

**Definition 4.** *Set of Focal Elements*
*Given labels $LA$ together with associated mass assignment $m_x : \forall x \in \Omega$, the set of focal elements for $LA$ is given by:*

$$\mathcal{F} = \{S \subseteq LA : \exists x \in \Omega,\ m_x(S) > 0\}$$

The appropriateness measure, $\mu_\theta(x)$, and the mass $m_x$ are then related to each other on the basis that asserting '$x$ is $\theta$' provides direct constraints on $\mathcal{D}_x$. For example, asserting '$x$ is $L_1 \wedge L_2$', for labels $L_1, L_2 \in LA$ is taken as conveying the information that both $L_1$ and $L_2$ are appropriate to describe $x$ so that $\{L_1, L_2\} \subseteq \mathcal{D}_x$. Similarly, '$x$ is $\neg L$' implies that $L$ is not appropriate to describe $x$ so $L \notin \mathcal{D}_x$. In general we can recursively define a mapping $\lambda : LE \to 2^{2^{LA}}$ from expressions to sets of subsets of labels, such that the assertion '$x$ is $\theta$' directly implies the constraint $\mathcal{D}_x \in \lambda(\theta)$ and where $\lambda(\theta)$ is dependent on the logical structure of $\theta$. For example, if $LA = \{low,\ medium,\ high\}$ then $\lambda(medium \wedge \neg high) = \{\{low, medium\}, \{medium\}\}$ corresponding to those sets of labels which include $medium$ but do not include $high$. Hence, the description $\mathcal{D}_x$ provides an alternative to Zadeh's linguistic variables in which the imprecise constraint '$x$ is $\theta$' on $x$, is represented by the precise constraint $\mathcal{D}_x \in \lambda(\theta)$, on $\mathcal{D}_x$.

**Definition 5.** *$\lambda$-mapping*
*$\lambda : LE \to 2^{2^{LA}}$ is defined recursively as follows: $\forall \theta,\ \varphi \in LE$*

- $\forall L_i \in LA\ \lambda(L_i) = \{T \subseteq LA : L_i \in T\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$
- $\lambda(\neg\theta) = \lambda(\theta)^c$
- $\lambda(\theta \to \varphi) = \lambda(\neg\theta) \cup \lambda(\varphi)$

Based on the λ mapping we then define $\mu_\theta(x)$ as the sum of $m_x$ over those set of labels in $\lambda(\theta)$.

**Definition 6.** *Appropriateness Measure*

$$\forall \theta \in LE,\ \forall x \in \Omega\ \mu_\theta(x) = \sum_{S \in \lambda(\theta)} m_x(S)$$

Note that in label semantics there is no requirement for the mass associated with the empty set to be zero. Instead, $m_x(\emptyset)$ quantifies the agent's belief that none of the labels are appropriate to describe $x$. We might observe that this phenomena occurs frequently in natural language, especially when labelling perceptions generated along some continuum. For example, we occasionally encounter colours for which none of our available colour descriptors seem appropriate. Hence, the value $m_x(\emptyset)$ is an indicator of the describability of $x$ in terms of the labels $LA$.

Given definition 6 it can be shown that appropriateness measures have the following general properties [9, 10]:

**Theorem 1.** *General Properties of Appropriateness Measures* $\forall \theta, \varphi \in LE$ *then the following properties hold:*

- *If* $\theta \models \varphi$ *then* $\forall x \in \Omega\ \mu_\theta(x) \leq \mu_\varphi(x)$
- *If* $\theta \equiv \varphi$ *then* $\forall x \in \Omega\ \mu_\theta(x) = \mu_\varphi(x)$
- *If* $\theta$ *is a tautology then* $\forall x \in \Omega\ \mu_\theta(x) = 1$
- *If* $\theta$ *is a contradiction then* $\forall x \in \Omega\ \mu_\theta(x) = 0$
- $\forall x \in \Omega\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$

Note that for implication $\mu_{\theta \to \varphi}(x)$ quantifies the agent's belief that if $\theta$ is appropriate to describe a given $x$ then so is $\varphi$. In particular, for two basic labels $L, L' \in LA$ $\mu_{L \to L'}(x)$ is an aggregation of those set values for $\mathcal{D}_x$ for which $L' \in \mathcal{D}_x$ whenever $L \in \mathcal{D}_x$. Such judgements are especially relevant when the values for $m_x$ are allocated so as to be consistent with an initial ordering on label appropriateness; this is an idea that we will discuss further in the sequel.

## *2.3 Functionality of Appropriateness Measures*

From definition 6 we see that in order to be able to evaluate the appropriateness measure of any given expression $\theta \in LE$ as a label for $x \in \Omega$ we must potentially know the value of $m_x$ for all subsets of $LA$. Hence, we are, in principle, required to specify of order $2^{|LA|} - 1$ functions $m_x$. For large basic label sets this is clearly computationally infeasible. One solution to this problem would be to make additional assumptions about the definition of the mass assignment $m_x$ so that there exists a functional relationship between the appropriateness measure for the basic labels (i.e. $\mu_L(x) : L \in LA$) and $m_x$. This would result in a functional calculus for appropriateness measures according to which the appropriateness of any compound expression could be determined directly from the appropriateness of the basic labels in the following sense:

**Definition 7.** *Functional Measures*
*A measure* $\mu$ *on* $LE \times \Omega$ *is said to be functional if* $\forall \theta \in LE$ *there exists a function* $f_\theta : [0,1]^{|LA|} \to [0,1]$ *such that* $\forall x \in \Omega$ $\mu_\theta(x) = f_\theta(\mu_L(x) : L \in LA)$

Now fuzzy logic is clearly functional in the sense of definition 7 but it also satisfies the stronger property of truth-functionality. Truth-functionality defines the mapping $f_\theta$ to be a recursive combination of functions representing each connective, as determined by the logical structure of the expression $\theta$. More formally:

**Definition 8.** *Truth-Functional Measures*
*A measure* $\mu$ *on* $LE \times \Omega$ *is said to truth-functional if there exists mappings* $f_\neg : [0,1] \to [0,1]$, $f_\wedge : [0,1]^2 \to [0,1]$, $f_\vee : [0,1]^2 \to [0,1]$ *and* $f_\to : [0,1]^2 \to [0,1]$ *such that* $\forall \theta, \varphi \in LE$*:*

- $\forall x \in \Omega$ $\mu_{\neg\theta}(x) = f_\neg(\mu_\theta(x))$
- $\forall x \in \Omega$ $\mu_{\theta\wedge\varphi}(x) = f_\wedge(\mu_\theta(x), \mu_\varphi(x))$
- $\forall x \in \Omega$ $\mu_{\theta\vee\varphi}(x) = f_\vee(\mu_\theta(x), \mu_\varphi(x))$
- $\forall x \in \Omega$ $\mu_{\theta\to\varphi}(x) = f_\to(\mu_\theta(x), \mu_\varphi(x))$

Now from theorem 1 it follows that appropriateness measures must satisfy the laws of excluded middle and idempotence. Hence, by the following theorem due to Dubois and Prade [1] they cannot be truth-functional except in the trivial case where all appropriateness values are either 0 or 1.

**Theorem 2.** *Dubois and Prade [1]*
*If* $\mu$ *is a truth-functional measure and satisfies both idempotence and the law of excluded middle then* $\forall \theta \in LE$, $\forall x \in \Omega$, $\mu_\theta(x) \in \{0,1\}$.

However, theorem 2 does not apply to all functional measures, only those which are truth-functional, hence it may still be possible to define a functional calculus for appropriateness measures consistent with both definitions 6 and 7. To investigate this possibility further we consider the relationship between appropriateness measures of compound expressions and those of the basic labels, imposed by definition 6.

From definition 5 the lambda mapping for a basic label $L \in LA$ is given by $\lambda(L) = \{S \subseteq LA : L \in S\}$ and hence the mass assignment $m_x$ must satisfy the following constraint imposed by the appropriateness measures for the basic labels $\mu_L(x) : L \in LA$:

$$\forall x \in \Omega,\ \forall L \in LA\ \mu_L(x) = \sum_{S \subseteq LA : L \in S} m_x(S)$$

This constraint, however, is not sufficient to identify a unique mass assignment $m_x$ given values for $\mu_L(x) : L \in LA$. Indeed, there are in general an infinite set of mass assignments satisfying the above equation for a given set of basic label appropriateness values. Hence, in this context, the assumption of a functional calculus for appropriateness measures is equivalent to the assumption of a selection function which identifies a unique assignment from this set [8, 10].

**Definition 9.** *Mass Selection Function*
*Let $\mathcal{M}$ be the set of all mass assignments on $2^{LA}$. Then a mass selection function (msf) is a function $\Delta\colon [0,1]^{|LA|} \to \mathcal{M}$ such that if $\forall x \in \Omega\ \Delta(\mu_L(x)\colon L \in LA) = m_x$ then*

$$\forall x \in \Omega\ \forall L \in LA \quad \sum_{S \subseteq LA: L \in S} m_x(S) = \mu_L(x)$$

Now since the value of $\mu_\theta(x)$ for any expression $\theta \in LE$ can be evaluated directly from $m_x$, then given a mass selection function $\Delta$ we have a functional method for determining $\mu_\theta(x)$ from the basic label appropriateness values, where $f_\theta$ in definition 7 is given by:

$$f_\theta(\mu_L(x)\colon L \in LA) = \sum_{F \in \lambda(\theta)} \Delta(\mu_L(x)\colon L \in LA)(F)$$

In this paper we shall consider the consonant and the independent mass selection functions as defined below:

**Definition 10.** *Consonant Mass Selection Function*
*Given non-zero appropriateness measures on basic labels $\mu_{L_i}: i = 1, \ldots, n$ ordered such that $\mu_{L_i}(x) \geq \mu_{L_{i+1}}(x)$ for $i = 1, \ldots, n$ then the consonant mass selection function identifies the mass assignment,*

$$m_x(\{L_1, \ldots, L_n\}) = \mu_{L_n}(x)$$
$$m_x(\{L_1, \ldots, L_i\}) = \mu_{L_i}(x) - \mu_{L_{i+1}}(x) \text{ for } i = 1, \ldots, n$$
$$m_x(\emptyset) = 1 - \mu_{L_1}(x)$$

**Definition 11.** *Independent Mass Selection Function*
*Given appropriateness measures on basic labels $\mu_L(x)\colon L \in LA$ then the independent mass selection function identifies the mass assignment,*

$$\forall S \subseteq LA\ m_x(S) = \prod_{L \in S} \mu_L(x) \times \prod_{L \notin S} (1 - \mu_L(x))$$

The consonant msf corresponds to the assumption that for each $x \in \Omega$ an agent first identifies a total ordering on the appropriateness of labels. They then evaluate their belief values $m_x$ about which labels are appropriate to describe $x$ in such a way so as to be consistent with this ordering. More formally, let $\preceq_x$ denote the appropriateness ordering on $LA$ for element $x$ so that $L_1 \preceq_x L_2$ means that $L_2$ is at least as appropriate as $L_1$ for describing $x$. When evaluating $m_x(S)$ for $S \subseteq LA$ the agent then makes the assumption that the mass value is non-zero only if for every label $L \in S$ it also holds that $L' \in S$ for every $L' \in LA$ for which $L \preceq_x L'$.

The independent msf simply assumes that when judging the appropriateness of a label an agent does not take into account the level of appropriateness of any other label. Although this may seem difficult to justify, it could be reasonable in cases where labels relate to different facets of the object. For example, the appropriateness of the label *thin* might well be assumed to be independent of the appropriateness of the label *rich*.

The following theorems show that the consonant and independent msf result in familiar combination rules for conjunctions and disjunctions of basic labels (see [7, 8, 9, 10]).

**Theorem 3.** *If* $\forall x \in \Omega$, *$m_x$ is determined from $\mu_L(x) : L \in LA$ according to the consonant msf then for labels $L_1, \ldots, L_n \in LA$ we have that $\forall x \in \Omega$:*

$$\mu_{L_1 \wedge L_2 \wedge \ldots \wedge L_n}(x) = \min\left(\mu_{L_1}(x), \ldots, \mu_{L_n}(x)\right)$$
$$\mu_{L_1 \vee L_2 \vee \ldots \vee L_n}(x) = \max\left(\mu_{L_1}(x), \ldots, \mu_{L_n}(x)\right)$$

**Theorem 4.** *If* $\forall x \in \Omega$, *$m_x$ is determined from $\mu_L(x) : L \in LA$ according to the independent msf then for labels $L_1, \ldots, L_n \in LA$ we have that $\forall x \in \Omega$:*

$$\mu_{L_1 \wedge L_2 \wedge \ldots \wedge L_n}(x) = \prod_{i=1}^{n} \mu_{L_i}(x)$$
$$\mu_{L_1 \vee L_2 \vee \ldots \vee L_n}(x) = \sum_{\emptyset \neq S \subseteq \{L_1, \ldots, L_n\}} (-1)^{|S|-1} \prod_{L \in S} \mu_L(x)$$

Notice that for both theorem 3 and 4 the rule does not apply generally but only to basic labels. This contrasts with fuzzy logic based on the min and max or the product rule respectively where the combination rules can be applied to any expression in a truth-functional manner. In fact a recent result due to Tang and Zheng [14] does show that theorem 3 can be extended to cover all expression involving only $\wedge$ and $\vee$ and no other connectives.

**Theorem 5.** *Let $LE^{\wedge,\vee} \subseteq LE$ denote those expressions generated recursively from $LA$ using only the connectives $\wedge$ and $\vee$. If $\forall x \in \Omega$, $m_x$ is determined from $\mu_L(x)$ : $L \in LA$ according to the consonant msf then $\forall \theta, \varphi \in LE^{\wedge,\vee}$, $\forall x \in \Omega$ it holds that:*

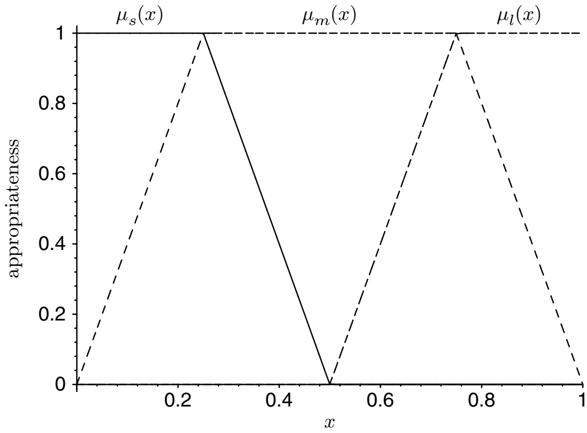
$$\mu_{\theta \wedge \varphi}(x) = \min\left(\mu_\theta(x), \mu_\varphi(x)\right) \text{ and } \mu_{\theta \vee \varphi}(x) = \max\left(\mu_\theta(x), \mu_\varphi(x)\right)$$

*Example 1.* Let $\Omega = [0, 1]$ and $LA = \{small(s), medium(m), large(l)\}$. Assume that the appropriateness measures for the basic labels are defined as trapezoidal functions of $x$ as given in Fig. 1. Applying the consonant msf generates the mass assignments shown in Fig. 2 for the focal elements $\mathcal{F} = \{\{s\}, \{s, m\}, \{m\}, \{m, l\}, \{l\}\}$.

## 2.4 Label Semantics and Random Sets

In this section we explore connections between label semantics and random sets and contrast these with the well known links between fuzzy sets and random sets.

**Fig. 1** Appropriateness measures for labels *small*, *medium*, and *large*




Firstly, note that $\mathcal{D}_x$ can be viewed as a random set into the power set of labels. From this perspective we have from definitions 5 and 6 that for basic labels $L \in LA$, the appropriateness measure $\mu_L(x)$ corresponds to the single point coverage value of $L$ for random set $\mathcal{D}_x$. That is:

$$\mu_L(x) = P\left(L \in \mathcal{D}_x\right) = \sum_{S:L \in S} m_x\left(S\right)$$

Now this has some similarity to the random set interpretation of fuzzy sets, where a fuzzy concept $\theta$ is modelled by a random set $R_\theta$ into the power set of the underlying universe $\Omega$. (See [3] and [4] by Goodman and Nguyen). The fuzzy set membership function, $M(\theta)$, is then taken to be the single point coverage function for $R_\theta$ as given by:

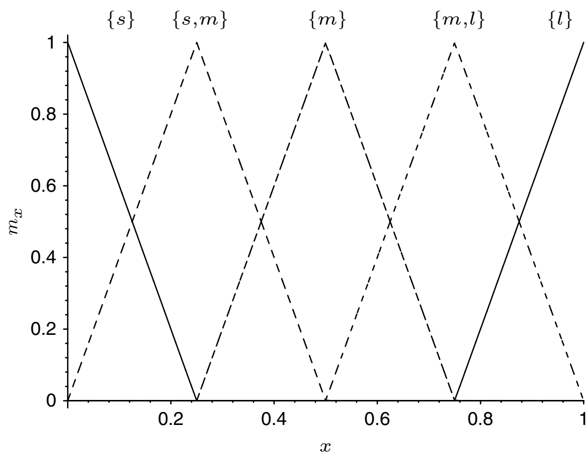



**Fig. 2** Mass assignment values $m_x$ for the focal elements as generated from the appropriateness measures in Fig. 1 according to the consonant msf

$$M(\theta)(x) = P(x \in R_\theta)$$

Hence, a fundamental difference between the two approaches is that labels semantics considers random sets defined on sets of labels, while the approach of Goodman and Nguyen defines random sets on sets of values from the underlying universe $\Omega$. This apparently subtle distinction means that the resulting calculi are quite different. In particular, the calculus for membership functions resulting from Goodman and Nguyen's approach is not functional, although combinations using min and product can be justified in certain special cases. Also, from a purely practical viewpoint label semantics allows for finite mathematics to be applied even when the underlying universe $\Omega$ is infinite, provided that the basic label set $LA$ is always assumed to be finite. This considerably simplifies the resulting calculus making its application to practical problems much more straightforward.

## 3 A Model of Assertions

In the previous sections we have introduced appropriateness measures and their underlying mass assignments as a mechanism by which an agent can assess the level of appropriateness of label expressions for describing some instance $x \in \Omega$. However, when faced with an object $x$ about which an agent wishes to convey information to other individuals in the population he/she is faced with a much more specific decision problem. What expression $\theta \in LE$ do they choose to assert in order to describe $x$ and how do they use their appropriateness measure to guide that choice?

In principle an agent can assert any appropriate expression in $LE$. However, in practice there is only a small subset of expressions which are really assertible. In particular, an agent may tend not to assert expressions which are logically equivalent to simpler expressions (i.e those involving fewer connectives). For example, neither $\neg\neg L$ nor $(L \wedge L') \vee (L \wedge \neg L')$ is likely to be asserted instead of $L$. Also there is some evidence to suggest that as humans we tend to not use negative statements as descriptions if at all possible. One is much more likely to describe the colour of this paper as *white* rather than *not red*, even though both expressions are appropriate. On the other hand we may use purely negative statements in situations where none of our label descriptors are appropriate. This can certainly occur if we are labelling elements of a continuum. For example, we may encounter colours for which none of our colour descriptors are appropriate. Overall, this suggests that while purely negative expressions may be assertible there is likely to be an a priori propensity to use positive expressions.

We now introduce a model of the decision process by which an agent identifies a particular assertion to describe an object, taking account of the measure of appropriateness as defined in the previous section. Let $AS \subseteq LE$ denote the finite set of permitted assertions. Let $\mathcal{A}_x$ be the assertion selected by the agent for describing instance $x$. Now each set of appropriate labels $\mathcal{D}_x = F$ identifies a set of possible values for $\mathcal{A}_x$ corresponding to those expression $\theta \in AS$ for which $F \in \lambda(\theta)$ (i.e. those expressions consistent with $F$). Hence, the mass assignment $m_x$ on $2^{LA}$ naturally generates a mass assignment on sets of possible assertions ($2^{AS}$) as follows:

**Definition 12.** *Mass Assignment on Assertions*
$m_x^{\mathcal{A}} : 2^{AS} \to [0, 1]$ *is defined such that: if* $\forall G \subseteq AS$

$$m_x^{\mathcal{A}}(G) = \sum_{F \in \mathcal{F} : \mathcal{C}(F) = G} m_x(F)$$

*where* $\mathcal{C} : \mathcal{F} \to 2^{AS}$ *and:*

$$\forall F \in \mathcal{F}\ \mathcal{C}(F) = \{\theta \in AS : F \in \lambda(\theta)\}$$

*Hence,* $\mathcal{C}(F)$ *identifies those assertions which are consistent with the knowledge that* $F$ *is the set of appropriate labels.*

Providing $AS$ is sufficiently large so that $\forall F \in \mathcal{F}, \exists \theta \in AS$ such that $F \in \lambda(\theta)$ (i.e. so that $m_x^{\mathcal{A}}(\emptyset) = 0$), then from definition 12 we can define a belief and plausibility measure on $\mathcal{A}_x$ in the normal manner, so that:

$$\forall S \subseteq AS\ Bel(\mathcal{A}_x \in S|x) = Bel_x^{\mathcal{A}}(S) = \sum_{G \subseteq S} m_x^{\mathcal{A}}(G)$$

$$Pl(\mathcal{A}_x \in S|x) = Pl_x^{\mathcal{A}}(S) = \sum_{G : G \cap S \neq \emptyset} m_x^{\mathcal{A}}(G)$$

This plausibility measure on assertions can be related directly to appropriateness measures according to the following theorem:

**Theorem 6.**

$$\forall \theta \in AS\ Pl_x^{\mathcal{A}}(\{\theta\}) = \mu_\theta(x)$$

*Proof.*

$$Pl_x^{\mathcal{A}}(\{\theta\}) = \sum_{G : \theta \in G} m_x^{\mathcal{A}}(G) = \sum_{F : F \in \lambda(\theta)} m_x(F) = \mu_\theta(x)$$

Notice, that $Bel_x^{\mathcal{A}}(\theta) = m_x^{\mathcal{A}}(\{\theta\})$ which corresponds to the level of belief that $\theta$ is the only assertible expression which is appropriate to describe $x$. Now given the mass assignment $m_x^{\mathcal{A}}$ and a prior distribution on the assertible expressions $AS$ we can define a probability distribution for $\mathcal{A}_x$ in accordance with Shafer [12] as follows:

**Definition 13.** *Probability of an Assertion*

$$\forall x \in \Omega,\ \forall \theta \in AS\ P\left(\mathcal{A}_x = \theta | x\right) = P_x^{\mathcal{A}}\left(\theta\right) = P\left(\theta\right) \sum_{S \subseteq AS: \theta \in S} \frac{m_x^{\mathcal{A}}(S)}{P(S)}$$
$$= P\left(\theta\right) \sum_{F \in \mathcal{F}: F \in \lambda(\theta)} \frac{m_x\left(F\right)}{P\left(\theta \in AS : F \in \lambda(\theta)\right)}$$

Notice that in the case that the prior probability $P$ on $AS$ is uniform then $P_x^{\mathcal{A}}$ is the pignistic distribution of $Bel_x^{\mathcal{A}}$ as defined by Smets [13]. Also, from Shafer [12] and theorem 6, we know that:

$$\forall \theta \in AS\ m_x^{\mathcal{A}}\left(\{\theta\}\right) \leq P_x^{\mathcal{A}}\left(\theta\right) \leq \mu_\theta\left(x\right)$$

In the following we investigate the form of $P_x^{\mathcal{A}}$ for different $AS$. In order to do this we recall the definition of logical atoms and describe how they are related to the focal set $\mathcal{F}$.

**Definition 14.** *Logical Atoms*

- *Let $ATT$ be the logical atoms of $LE$ corresponding to those expressions for which:*

$$\alpha = \bigwedge_{L \in LA} \pm L$$

*where $+L$ denotes L and $-L$ denote $\neg L$.*

- *For each focal element $F \in \mathcal{F}$ there is an atom $\alpha_F \in ATT$ equivalent to the statement $\mathcal{D}_x = F$ defined as follows:*

$$\alpha_F = \left(\bigwedge_{L ! \in ! F} L\right) \wedge \left(\bigwedge_{L \in LA - F} \neg L\right)$$

**Theorem 7.** *If $AS = \{\alpha_F : F \in \mathcal{F}\}$ then $\forall \alpha_F \in AS,\ P_x^{\mathcal{A}}\left(\alpha_F\right) = m_x\left(F\right)$*

*Proof.*

$$\forall F \in \mathcal{F}\ \mathcal{C}\left(F\right) = \{\alpha_F\} \Rightarrow m_x^{\mathcal{A}}\left(\{\alpha_F\}\right) = m_x\left(F\right) \text{ therefore}$$
$$P_x^{\mathcal{A}}\left(\alpha_F\right) = P\left(\alpha_F\right) \frac{m_x\left(\{\alpha_F\}\right)}{P\left(\alpha_F\right)} = m_x^{\mathcal{A}}\left(\{\alpha_F\}\right) = m_x\left(F\right)$$

**Theorem 8.** *If $AS = LA$ then $P_x^{\mathcal{A}}\left(L\right) = P\left(L\right) \sum_{F: L \in F} \frac{m_x(F)}{P(F)}$*

*Proof.* Follows trivially since $\forall F \in \mathcal{F}\ \mathcal{C}\left(F\right) = F$

Theorem 8 effectively corresponds to the basic model of assertibility as presented in Lawry's earlier work on computing with words, including for example [6].

**Theorem 9.** *If* $AS = LA \cup \{\neg L : L \in LA\}$ *then assuming a uniform prior distribution on* $AS$ *it holds that:*

$$P_x^{\mathcal{A}}(L) = \frac{\mu_L(x)}{|LA|} \text{ and } P_x^{\mathcal{A}}(\neg L) = \frac{\mu_{\neg L}(x)}{|LA|}$$

*Proof.*

$$\forall F \in \mathcal{F}\ \mathcal{C}(F) = F \cup \{\neg L : L \in LA - F\} \Rightarrow$$
$$P_x^{\mathcal{A}}(L) = \sum_{F:L\in F} \frac{m_x(F)}{|F \cup \{\neg L : L \in LA - F\}|} = \sum_{F:L\in F} \frac{m_x(F)}{|LA|}$$
$$= \frac{\sum_{F:L\in F} m_x(F)}{|LA|} = \frac{\mu_L(x)}{|L|}$$

Similarly for $P_x^{\mathcal{A}}(\neg L)$.

**Theorem 10.** *If* $AS = LA \cup \{\alpha_F : F \in \mathcal{F}\}$ *then*

$$\forall F \in \mathcal{F}\ P_x^{\mathcal{A}}(\alpha_F) = \left(\frac{P(\alpha_F)}{P(F) + P(\alpha_F)}\right) m_x(F) \text{ and}$$
$$\forall L \in LA\ P_x^{\mathcal{A}}(L) = P(L) \sum_{F:L\in F} \frac{m_x(F)}{P(F) + P(\alpha_F)}$$

*Proof.* Follows trivially since $\mathcal{C}(F) = F \cup \{\alpha_F\}$

*Example 2.* Let $LA = \{small(s), medium(m), large(l)\}$ and for $x \in \Omega$ let $m_x(\{s, m\}) = 0.6$, $m_x(\{m\}) = 0.4$ then consider the following cases:

- Case 1: Take $AS = \{s, m, l, s \wedge \neg m \wedge \neg l, s \wedge m \wedge \neg l, m \wedge \neg s \wedge \neg l, m \wedge l \wedge \neg s, l \wedge \neg m \wedge \neg s\}$ together with a uniform prior probability distribution on assertions then:

$$\mathcal{C}(\{s, m\}) = \{s, m, s \wedge m \wedge \neg l\},\ \mathcal{C}(\{m\}) = \{m, m \wedge \neg s \wedge \neg l\} \text{ therefore}$$
$$P_x^{\mathcal{A}}(s \wedge m \wedge \neg l) = \frac{m_x(\{s, m\})}{3} = \frac{0.6}{3} = 0.2$$
$$P_x^{\mathcal{A}}(m \wedge \neg s \wedge \neg l) = \frac{m_x(\{m\})}{2} = \frac{0.4}{2} = 0.2$$
$$P_x^{\mathcal{A}}(s) = \frac{m_x(\{s, m\})}{3} = \frac{0.6}{3} = 0.2$$
$$P_x^{\mathcal{A}}(\{m\}) = \frac{m_x\{s, m\}}{3} + \frac{m_x(\{m\})}{2} = 0.2 + 0.2 = 0.4$$

Figures 3 and 4 show values for $P_x^{\mathcal{A}}$ in this case as a function of $x$, for the labels and logical atoms respectively where the basic labels are defined according to the appropriateness measures shown in Fig. 1.

- Case 2: Take $AS = \{s, m, l, \neg s, \neg m, \neg l\}$ together with a uniform prior probability distribution on assertions then:

$$\mathcal{C}(\{s, m\}) = \{s, m, \neg l\}, \ \mathcal{C}(\{m\}) = \{m, \neg s, \neg l\} \text{ therefore}$$

$$P_x^{\mathcal{A}}(s) = \frac{m_x(\{s, m\})}{3} = \frac{\mu_s(x)}{3} = \frac{0.6}{3} = 0.2$$

$$P_x^{\mathcal{A}}(\neg s) = \frac{m_x(\{m\})}{3} = \frac{0.4}{3} = \frac{\mu_{\neg s}(x)}{3} = 0.1333$$

$$P_x^{\mathcal{A}}(m) = \frac{m_x(s, m)}{3} + \frac{m_x(m)}{3} = \frac{1}{3} = \frac{\mu_m(x)}{3} = 0.3333$$

$$P_x^{\mathcal{A}}(\neg m) = 0 = \frac{\mu_{\neg m}(x)}{3}$$

$$P_x^{\mathcal{A}}(l) = 0$$

$$P_x^{\mathcal{A}}\{\neg l\} = \frac{1}{3} = 0.3333$$

Figures 5 and 6 show values for $P_x^{\mathcal{A}}$ in this case as a function of $x$, for the labels and negated labels respectively where the basic labels are defined according to the appropriateness measures shown in Fig. 1.

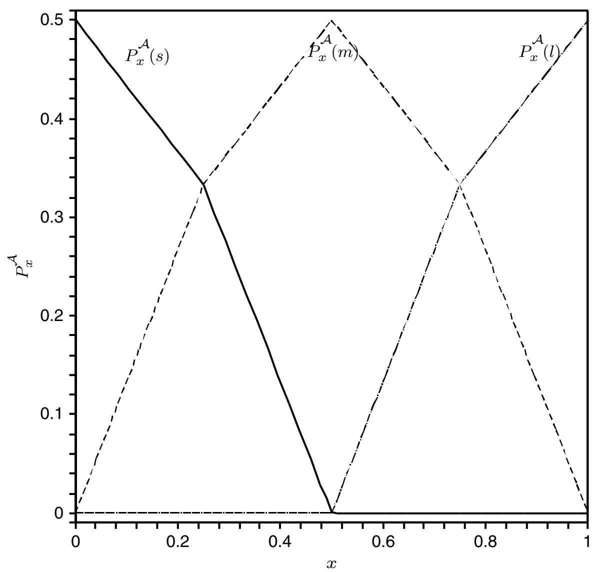



**Fig. 3** The probability distribution $P_x^{\mathcal{A}}$ in case 1, for the labels *small*, *medium*, and *large* as defined in Fig. 1

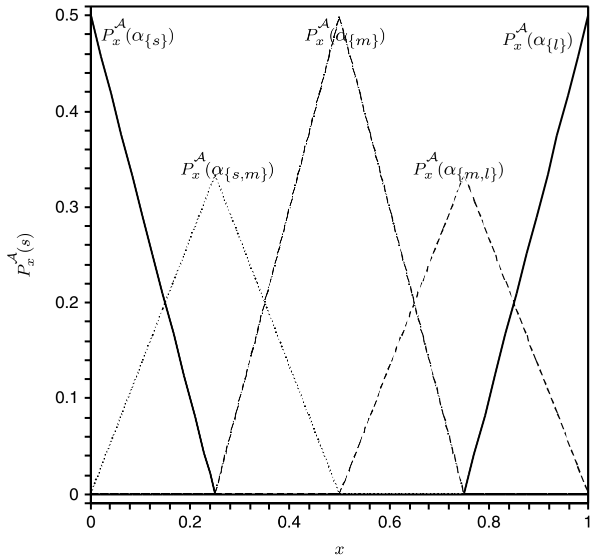



**Fig. 4** The probability distribution $P_x^{\mathcal{A}}$ in case 1, for the atoms $\alpha_{\{s\}} = s \wedge \neg m \wedge \neg l$, $\alpha_{\{s,m\}} = s \wedge m \wedge \neg l$, $\alpha_{\{m\}} = m \wedge \neg s \wedge \neg l$, $\alpha_{\{m,l\}} = m \wedge l \wedge \neg s$, and $\alpha_{\{l\}} = l \wedge \neg s \wedge \neg m$

- Case 3: Take $AS = \{s, m, l\}$ together with a uniform prior probability distribution on assertions then:

$$\mathcal{C}(\{s, m\}) = \{s, m\}, \ \mathcal{C}(\{m\}) = \{m\} \text{ therefore}$$
$$P_x^{\mathcal{A}}(s) = \frac{m_x(\{s, m\})}{2} = \frac{0.6}{2} = 0.3$$
$$P_x^{\mathcal{A}}(m) = \frac{m_x(\{s, m\})}{2} + m_x(\{m\}) = 0.3 + 0.4 = 0.7$$

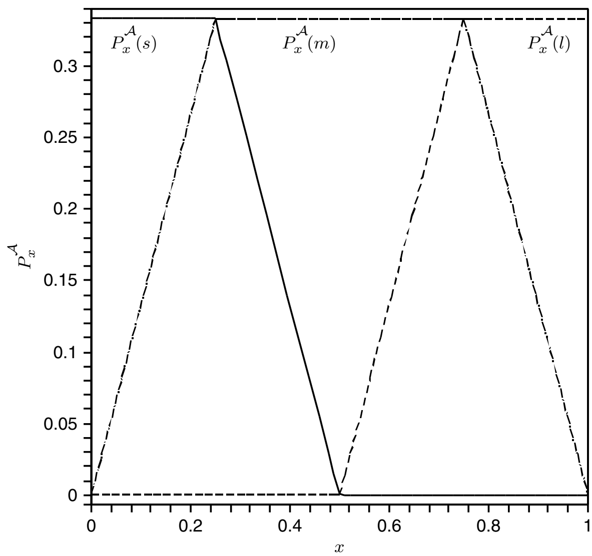



**Fig. 5** The probability distribution $P_x^{\mathcal{A}}$ in case 2, for the labels *small*, *medium*, and *large* as defined in Fig. 1

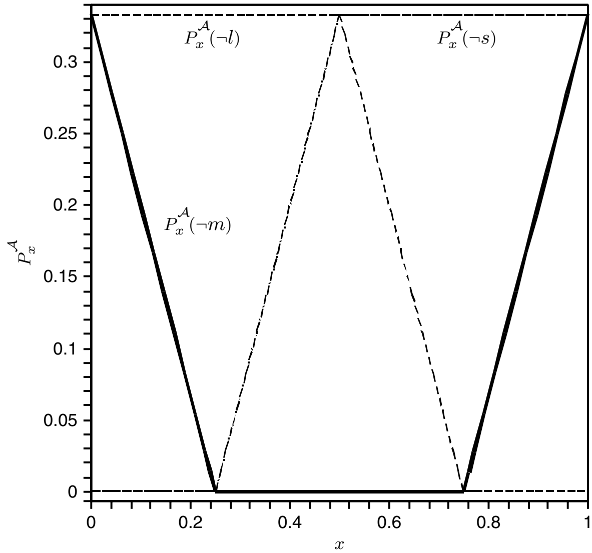



**Fig. 6** The probability distribution $P_x^{\mathcal{A}}$ in case 2, for the negated labels $\neg s$, $\neg m$ and $\neg l$

Figures 7 and shows the values for $P_x^{\mathcal{A}}$ in this case as a function of $x$, for the labels defined according to the appropriateness measures shown in Fig. 1.

In example 2 we have assumed a uniform prior distribution on $AS$. However, this is somewhat unrealistic in view of the discussion at the beginning of this section. For instance, if in communications positive assertions are favoured over negative then we would expect to have a prior $P$ for which $P(L) > P(\neg L)$. Also, if one of the aims of an agent is to be as informative as possible then they would tend, a priori,

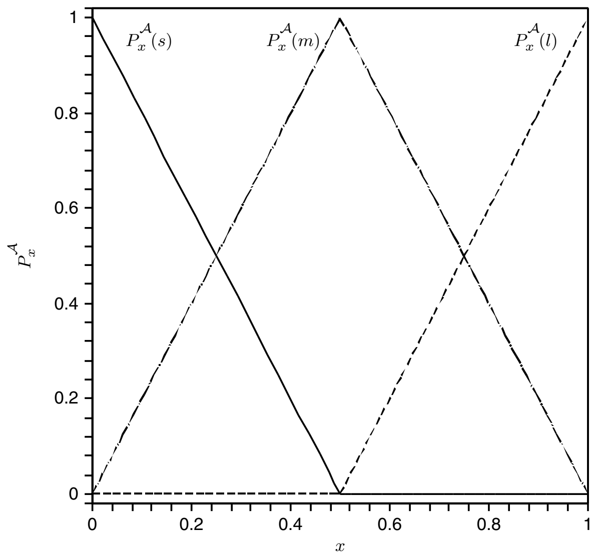



**Fig. 7** The probability distribution $P_x^{\mathcal{A}}$ in case 3, for the labels $s$, $m$ and $l$

to favour more precise over more general expressions. In this case, for $\theta, \varphi \in AS$ such that $\theta \models \varphi$ we would expect that $P(\theta) \geq P(\varphi)$. In the case that a precise prior probability distribution is specified then an agent could be expected to select his/her assertion from the pareto set of maximally probable assertions as given by:

$$\left\{\theta \in AS : \forall \varphi \in AS,\ P_x^{\mathcal{A}}(\varphi) \leq P_x^{\mathcal{A}}(\theta)\right\}$$

If necessary, further restrictions might then be imposed in order to select a unique assertion, perhaps, for example, by minimizing the number of connectives or by some other interpretability criterion.

On the other hand, the constraints described above are unlikely to be sufficient to identify a unique prior distribution on $AS$. This suggests that an imprecise probability model may be more appropriate. In this case, one approach would be to define the above pareto set in terms of upper probabilities. For example, if the prior $P$ is allowed to range across all potential distributions on $AS$ then, according to theorem 6, this would correspond to:

$$\left\{\theta \in AS : \forall \varphi \in AS,\ \mu_\varphi(x) \leq \mu_\theta(x)\right\}$$

## 4 Information from Linguistic Expressions

One of the principle roles of vague linguistic descriptors in communication is to convey useful information in a robust and flexible way. If we learn that ‘John is *tall*’ then this tells us something about John’s height, which would then be useful to us if we were ask to try to identify John from a crowd of people, each with different heights. Generally, such descriptions across a range of attributes provide useful guides in complex search problems [11]. But what is the nature of the information conveyed in assertions of this kind and how can we model the associated uncertainty? In this section we investigate the problem from a label semantics perspective.

Consider a scenario in which, agent $A$ makes the assertion ‘$x$ is $\theta$’ to agent $B$. We now consider several models of the information conveyed to $B$ by this assertion which are consistent with the label semantics framework. The study is simplified by making the following assumptions:

- $B$ assumes $A$ is telling the truth.
- $B$ assumes $A$ uses the expression $\theta$ in the same way as him/her. (i.e. both $A$ and $B$ have the same appropriateness measure $\mu_\theta(x)$)

While these assumptions might be questioned they do allow us to focus specifically on the linguistic uncertainty in the information that $A$ conveys to $B$. Also, $B$ might justify the second assumption to him or herself along the lines that he/she believes their current model of the conventions of language use (as modelled by their appropriateness measure), which have been learnt from experience, is a reasonable approximation of the ‘actual’ conventions employed across the population.

Evidence for this might be based on $B$'s impression that he/she has both understood other agents and been understood by them in pervious exchanges of this kind.

## 4.1 Bayesian Models

Here we present two Bayesian models of $B$'s uncertainty based on subtly different interpretations of the assertion '$x$ is $\theta$'. Both models require that $B$ holds prior beliefs regarding the underlying universe $\Omega$ in this case represented by a prior density function $f$ [1].

(i) *$B$ interprets $A$'s assertion to mean that $\theta$ is appropriate to describe $x$.* Let $P(\theta|x)$ denote $B$'s subjective probability that $\theta$ is appropriate given $x$, corresponding to $\mu_\theta(x)$ in label semantics. Applying Bayes' theorem $B$ then obtains a posterior density function on $\Omega$ given by:

$$\forall x \in \Omega \;\; f(x|\theta) = \frac{P(\theta|x) f(x)}{\int_\Omega P(\theta|x) f(x) dx} = \frac{\mu_\theta(x) f(x)}{\int_\Omega \mu_\theta(x) f(x) dx}$$

(ii) *$B$ interprets $A$'s assertion to mean that $\theta$ is assertible as a description of $x$.* In order for $B$ to evaluate a posterior density in this case, he/she must assume that $B$ shares their view on what expressions are assertible (i.e they must have the same $AS$) and also have the same prior distribution on assertible expressions. In summary, $A$ and $B$ should have the same probability distribution $P_x^{\mathcal{A}}$. In this case:

$$\forall x \in \Omega \;\; f(x|\theta) = \frac{P_x^{\mathcal{A}}(\theta) \, f(x)}{\int_\Omega P_x^{\mathcal{A}}(\theta) \, f(x) dx}$$

## 4.2 Possibilistic Models

In this sub-section we present a possibilistic model of $B$'s uncertainty based on a looser interpretation of $A$'s assertion. Specifically, $B$ interprets '$x$ is $\theta$' to mean simply that $\theta$ is *highly* appropriate as a description of $x$. This is formalised by the constraint $\mu_\theta(x) \geq \alpha$ for some unknown threshold value $\alpha$ between 0 and 1. This restricts $x$ to a set of *possible* values with appropriateness of $\theta$ at least $\alpha$ (i.e. $\theta_\alpha = \{x \in \Omega : \mu_\theta(x) \geq \alpha\}$). However, since $\alpha$ is unknown we define a measure of the possibility of $x$ as the maximal value of $\alpha$ for which the constraint holds:

$$\pi(x|\theta) = \sup\{\alpha \in [0,1] : x \in \theta_\alpha\} = \mu_\theta(x)$$

[1] We assume here that $\Omega$ is a compact interval of the real line.

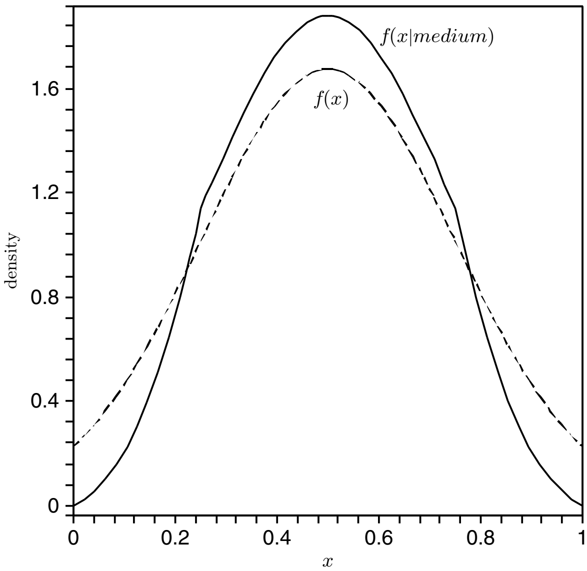



**Fig. 8** The posterior distribution obtained using Bayesian method (i)

A subtle variation of this model can be obtain if $B$ interprets $A$'s assertion as $\theta$ is *highly* assertible, formally represented by the constraint $P_x^{\mathcal{A}}(x) \geq \alpha$. In this case the set of possible values is given by $\theta_\alpha^{\mathcal{A}} = \{x \in \Omega : P_x^{\mathcal{A}}(x) \geq \alpha\}$ resulting in the following possibility distribution:

$$\pi^{\mathcal{A}}(x|\theta) = \sup \left\{\alpha \in [0, 1] : x \in \theta_\alpha^{\mathcal{A}}\right\} = P_x^{\mathcal{A}}(\theta)$$

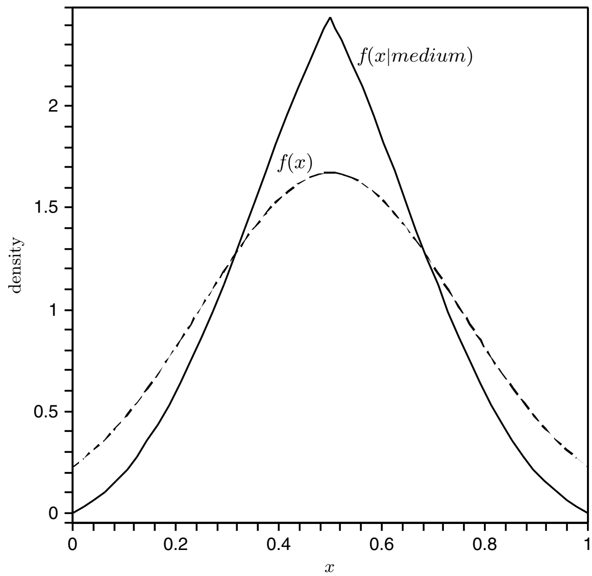



**Fig. 9** The posterior distribution obtained using Bayesian method (ii)

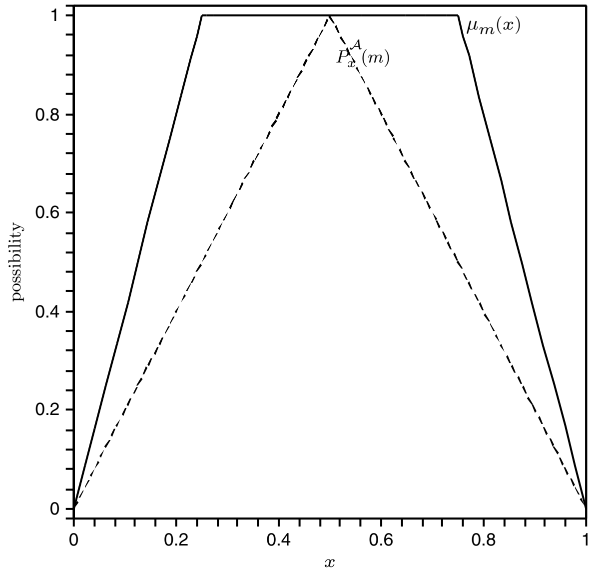



**Fig. 10** Possibility measures $\pi(x|medium) = \mu_{medium}(x)$ and $\pi^{\mathcal{A}}(x|medium) = P_x^{\mathcal{A}}(medium)$

Now trivially from theorem 6 we have that $\pi^{\mathcal{A}}(x|\theta) \leq \pi(x|\theta)$ so that the first model clearly generates the most general distribution. This also suggests that in cases where $B$ is not prepared to make assumptions concerning $A$'s prior distribution on assertions, then the distribution $\pi(x|\theta)$ provides a good possibilistic model of his/her uncertainty.

*Example 3.* Consider the labels $LA = \{small, medium, large\}$ defined as in Fig. 1, to describe elements in $\Omega = [0, 1]$. Suppose $A$ asserts that '$x$ is $medium$' and suppose that $B$ has a prior distribution $f$ on $\Omega$ corresponding to a normal distribution $\mathcal{N}(0.5, 0.25)$ normalised so as to give a density of 1 on [0, 1]. Figure 8 then shows $B$'s posterior density based on Bayesian method (i) where the assertion is interpreted as meaning that $medium$ is appropriate to describe $x$. Figure 9 shows $B$'s posterior density based on Bayesian method (ii) where the assertion is interpreted as meaning that $medium$ is assertible given $x$. This is based on case 3 of example 2 where $AS = \{small, medium, large\}$ and where there is a uniform prior distribution on assertions. The two types of possibility measure conditional on the assertion '$x$ is $medium$' are then also shown in Fig. 10.

## 5 Conclusions

In this paper we have presented an overview of label semantics as a formal framework for modelling and computing with words. In contrast to Zadeh's methodology this approach is based on measures of an agent's subjective belief that a linguistic expression is appropriate to describe a particular object or value. Central to label semantics is the *epistemic stance* according to which individual agents believe in the existence of a set of linguistic conventions, shared across the population, governing what labels are appropriate to use in a given context. In this framework appropriateness measures can be functional but never truth-functional, and certain well known operators can be applied for combining basic labels both conjunctively and disjunctively.

A model of the uncertainty associated with selecting a unique assertion for describing an object, has been proposed and a number of special cases have been investigated. These correspond to different a priori assumptions about what expression are assertible. In context it is shown that the appropriateness measure of an expression is an upper probability that it will be asserted.

Finally, we have considered label semantics based models of the information that is conveyed by linguistic assertions of the form '$x$ is $\theta$'. In particular, we have proposed both Bayesian and possibilistic posterior distributions resulting from conditioning on such assertions. These arise from different interpretations of the assertion within the label semantics framework.

## References


1. D. Dubois, H. Prade, (1988), 'An Introduction to Possibility and Fuzzy Logics' in *Non-Standard Logics for Automated Reasoning* (eds. P. Smets et al), Academic Press, pp. 742–755.
2. D. Dubois, H. Prade, (1997), 'The Three Semantics of Fuzzy Sets' *Fuzzy Sets and Systems* Vol. 90, pp. 141–150.
3. I.R. Goodman, (1982), 'Fuzzy Sets as Equivalence Classes of Random Sets' in *Fuzzy Set and Possibility Theory* (ed. R. Yager), pp. 327–342.
4. I.R. Goodman, H.T. Nguyen ,(1985), *Uncertainty Models for Knowledge Based Systems* North Holland.
5. A. Kyburg, (2000), 'When Vague Sentences Inform: A Model of Assertability', *Synthese*, Vol. 124, pp. 175–192.
6. J. Lawry, (2001), 'A Methodology for Computing with Words', *International Journal of Approximate Reasoning*, Vol. 28, pp. 51–89.
7. J. Lawry, (2001), 'Label Semantics: A Formal Framework for Modelling with Words', *Proceedings of The European Conference on Symbolic and Quantitative Approaches to Reasoning Under Uncertainty, Lecture Notes in Artificial Intelligence*, (Eds. S. Benferhat, P. Besnard), Vol. 2143, pp. 374–385.
8. J. Lawry, J. Recasens, (2003), 'A Random Set Model for Fuzzy Labels', *Proceedings of The European Conference on Symbolic and Quantitative Approaches to Reasoning Under Uncertainty, Lecture Notes in Artificial Intelligence* (Eds. T. Nielsen, N. Zhang), Vol. 2711, pp. 357–369.
9. J. Lawry, (2004), 'A Framework for Linguistic Modelling' *Artificial Intelligence*, Vol. 155, pp. 1–39.
10. J. Lawry, (2006), *Modelling and Reasoning with Vague Concepts*, Springer.
11. R. Parikh, (1996) 'Vague Predicates and Language Games', *Theoria (Spain)*, Vol. XI, No. 27, pp. 97–107.
12. G. Shafer, (1976), *A Mathematical Theory of Evidence*, Princeton University Press.
13. P Smets, (1990), 'Constructing the Pignistic Probability Function in a Context of Uncertainty', in *Uncertainty in Artificial Intelligence 5*, (Ed. M. Henrion), North Holland, Amsterdam, pp. 29–39.

14. Y. Tang, J. Zheng, (2006), 'Linguistic Modelling Based on Semantic Similarity Relation Among Linguistic Labels', *Fuzzy Sets and Systems*, in press.
15. P. Walley, G. de Cooman, (2001), 'A Behavioural Model of Linguistic Uncertainty', *Information Sciences*, Vol. 134, pp. 1–37.
16. T. Williamson, (1994), *Vagueness*, Routledge.
17. L.A. Zadeh 'The Concept of Linguistic Variable and its Application to Approximate Reasoning Part 1' *Information Sciences* Vol. 8 (1975) pp. 199–249.
18. L.A. Zadeh 'The Concept of Linguistic Variable and its Application to Approximate Reasoning Part 2' *Information Sciences* Vol. 8 (1975) pp. 301–357.
19. L.A. Zadeh 'The Concept of Linguistic Variable and its Application to Approximate Reasoning Part 3' *Information Sciences* Vol. 9 (1976) pp. 43–80.
20. L.A. Zadeh 'Fuzzy Sets as a Basis for a Theory of Possibility' *Fuzzy Sets and Systems* Vol. 1 (1978) pp. 3–28.
21. L.A. Zadeh 'Fuzzy Logic = Computing with Words' *IEEE Transactions on Fuzzy Systems* Vol. 4 (1996) pp. 103–111.
22. L.A. Zadeh, (2004), 'Precisiated Natural Language (PNL)', *AI Magazine*, Vol. 25, No. 3, pp. 74–91.

# On the Construction of Models Based on Fuzzy Measures and Integrals

**Vicenç Torra**

**Abstract** In this chapter, we review some families of fuzzy measures and their use in fuzzy integrals. We will also review the determination of fuzzy measures from examples in the case of the Choquet integral.

## 1 Introduction

Aggregation operators [2, 23] are functions that are used to combine data from information sources. Such information sources can be of any nature, either experts, sensors or agents (in a multi-agent system). Aggregation operators combine the data these sources supply to give a data of better quality.

Therefore, the data supplied by the sources is the basic object for aggregation operators, and it is the basis for computing the output. Nevertheless, another element plays a central role: the parameters of such operators. Most aggregation operators permit to consider some information about the importance of the sources. For example, in the case of the weighted mean, the weighting vector plays this role. Weights permit to establish which are the sources that are more reliable, or which is the expert with a higher degree of expertise.

Such information about the reliability or importance of the sources corresponds in some sense to the so-called background knowledge, using the jargon in artificial intelligence. That is, the weights are typically used to represent some knowledge on the sources that is already known at the time of implementing the system.

In the case of fuzzy integrals, fuzzy measures play this role. Fuzzy integrals are versatile tools for data aggregation. Formally speaking, fuzzy integrals integrate a function with respect to a fuzzy measure. The function is a mapping between the sources and the value that such sources supply. The fuzzy measure is used to represent the prior information on the sources.

In this chapter we will present a review on fuzzy measures for using them in fuzzy integrals. The structure of the paper is as follows. First, in Sect. 2 we will review some fuzzy integrals. Then, in Sect. 3, we will consider a few families of such fuzzy measures. Then, in Sect. 4 we will describe the problem of measure determination from examples. The paper finishes with some conclusions.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

## 2 Fuzzy Integrals

The use of fuzzy integrals for aggregation needs some formalization. First, we need to define the set of information sources, let it be $X = \{x_1, \ldots, x_N\}$. Then, we need to express the datum (a number) supplied by source $x_i$. This is represented with a function $f$ from $X$ into $\mathbb{R}$. Then, the aggregation of the data supplied by sources $X$ can be computed as the integral of the function $f$ with respect to a fuzzy measure.

A fuzzy measure is a set function on $X$. It assigns a value in [0, 1] to each subset $A$ of $X$. Fuzzy measures are monotonic with respect to set inclusion. They are formally defined as follows:

**Definition 1.** *[16, 17] A fuzzy measure $\mu$ is a set function $\mu : \wp(X) \rightarrow [0,1]$ that satisfies:*

- $\mu(\emptyset) = 0$, $\mu(X) = 1$
- $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$

*In this definition, boundary conditions are given first, and, then, monotonicity with respect to set inclusion.*

Fuzzy measures are used to represent the importance or relevance of a set, and play a role similar to the one of weights in the weighted mean. Several families of fuzzy measures are described in the next section.

The most widely used fuzzy integrals are the Choquet and the Sugeno integrals. They are defined as follows:

**Definition 2.** *[3] The* Choquet integral *of a function $f : X \rightarrow \mathbb{R}^+$ with respect to a fuzzy measure $\mu$ is defined by:*

$$(C)\int f d\mu = \sum_{i=1}^{N} [f(x_{s(i)}) - f(x_{s(i-1)})]\mu(A_{s(i)}) \tag{1}$$

*where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq \cdots \leq f(x_{s(N)}) \leq 1$, where $f(x_{s(0)}) = 0$ and where $A_{s(i)} = \{x_{s(i)}, \ldots, x_{s(N)}\}$.*

There exists a previous definition (see [24]) but for additive measures.

**Definition 3.** *[16, 17] The* Sugeno integral *of a function $f : X \rightarrow [0,1]$ with respect to a fuzzy measure $\mu$ is defined by:*

$$(S)\int f d\mu = \max_{i=1,N} \min(f(x_{s(i)}), \mu(A_{s(i)})) \tag{2}$$

*Here, $s$ and $A_{s(i)}$ are defined as above.*

Other fuzzy integrals have been defined in the literature. We can distinguish among them the ones that generalize Choquet and Sugeno integrals. Generalization is in the sense that a given particularization can reduce the integral to the former ones. The twofold integral defined below is an example of such expressions.

The t-conorm integral [12] and the general fuzzy integral [1] are other examples of fuzzy integrals.

**Definition 4.** *[21, 13] The* twofold integral *of a function $f : X \to [0,1]$ with respect to fuzzy measures $\mu_S$ and $\mu_C$ is defined by:*

$$TI_{\mu_S,\mu_C}(f) = \sum_{i=1}^{n} \Big( \big( \bigvee_{j=1}^{i} f(x_{s(j)}) \wedge \mu_S(A_{s(j)}) \big) \big( \mu_C(A_{s(i)}) - \mu_C(A_{s(i+1)}) \big) \Big) \quad (3)$$

*Here, $s$ and $A_{s(i)}$ are defined as above. Naturally, $A_{s(n+1)} = \emptyset$.*

When $\mu_C = \mu^*$, the twofold integral reduces to the Sugeno integral and when $\mu_S = \mu^*$, the twofold integral reduces to the Choquet integral. Here, $\mu^*$ represents ignorance and is defined by $\mu(\emptyset) = 0$ and $\mu^*(A) = 1$ when $A \neq \emptyset$.

## 3 Fuzzy Measures

There exist several families of fuzzy measures. The most remarkable ones are the additive measures. They correspond to probability measures. A fuzzy measure is additive when $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A \cap B = \emptyset$.

An important result that links the Choquet integral with the weighted mean is that the Choquet integral with respect to an additive fuzzy measure corresponds to a weighted mean with weights $\mu(\{x_i\})$.

Before presenting some of the families of fuzzy measures, we review the definition of the Möbius transform:

**Definition 5.** *[15] The Möbius transform $m$ of a fuzzy measure $\mu$ is defined for all subsets $A$ of $X$ by:*

$$m_\mu(A) := \sum_{B \subseteq A} (-1)^{|A|-|B|} \mu(B) \quad (4)$$

Then, given a Möbius transform $m$, the set function $\mu$ defined for all $A \subseteq X$ by:

$$\mu(A) = \sum_{B \subseteq A} m(B)$$

is a fuzzy measure (the fuzzy measure with Möbius transform $m$).

It can be proven that for additive fuzzy measures, the Möbius transform $m$ of a fuzzy measure $\mu$ satisfies $m(A) = 0$ for all $|A| \geq 2$.

The so-called $k$-additive fuzzy measures can be seen as a generalization of additive fuzzy measures. They are so in the sense that instead of having $m(A) = 0$ for all $|A| > 1$, they have $m(A) = 0$ for all $|A| > k$.

**Definition 6.** *[8] $\mu$ is a $k$-*order additive fuzzy measure *if its Möbius transform $m$ is such that $m(A) = 0$ for all $S \subseteq X$ such that $|S| > k$ and there exists at least one $S \subseteq X$ with $|S| = k$ such that $m(S) \neq 0$.*

This family of measures has the property that cover all the space of fuzzy measures. That is, for any fuzzy measure $\mu$, there is a $k$ such that $\mu$ is a $k$-order additive fuzzy measure.

$k$-additive fuzzy measures permit to express additive interactions for sets of up to $k$ sources (e.g. interaction of $k$ criteria in multi-criteria decision making applications).

Another family of measures are the symmetric fuzzy measures. They are measures where $\mu(A)$ only depends on the cardinality of the set. Such measures are defined by:

**Definition 7.** *A fuzzy measure $\mu$ is symmetric when for all $A, B \subseteq X$ we have that:*

$$\textit{if } |A| = |B| \textit{ then } \mu(A) = \mu(B)$$

An important result for such measures is that we can express them in terms of a weighting vector $w = (w_1, \ldots, w_{|X|})$ as follows:

$$\mu(A) = \sum_{i=1}^{|A|} w_i \tag{5}$$

Naturally, $w_i$ are positive and add to one (as $\mu(X) = 1$).

The Choquet integral with respect to a symmetric fuzzy measure corresponds to the OWA operator [26], and the Sugeno integral with respect to a symmetric fuzzy measure corresponds to the WMax operator (introduced in [4]).

Belief and plausibility measures are another example of fuzzy measures. They are formally defined as fuzzy measures that satisfy an additional condition. A fuzzy measure $Bel$ is a belief measure if the following condition is satisfied:

$$\begin{aligned} Bel(A_1 \cup ... \cup A_n) \geq \sum_j Bel(A_j) - \sum_{j<k} Bel(A_j \cap A_k) + ... + \\ (-1)^{n+1} Bel(A_1 \cap ... \cap A_n) \end{aligned} \tag{6}$$

Instead, a fuzzy measure $Pl$ is a plausibility if it satisfies:

$$\begin{aligned} Pl(A_1 \cap ... \cap A_n) \leq \sum_j Pl(A_j) - \sum_{j<k} Pl(A_j \cup A_k) + ... + \\ (-1)^{n+1} Pl(A_1 \cup ... \cup A_n) \end{aligned} \tag{7}$$

Equivalently, we can say that a fuzzy measure is a belief function when its Möbius transform is always positive. For these measures the Möbius transform is known as a *basic probability assignment*.

Another example of fuzzy measures are the so-called distorted probabilities. They are measures that can be expressed in terms of a probability and a function applied to it that distorts the probability. Formally speaking, the are defined as shown below:

**Definition 8.** *[6, 7] A fuzzy measure $\mu$ is a distorted probability if it is represented by a probability distribution $P$ on $(X, \wp(X))$ and a function $f$ that is non-decreasing with respect to a probability $P$.*

The Choquet integral with respect to a distorted probability results into the WOWA operator. The WOWA operator was introduced in [18].

Distorted probabilities were generalized into $m$-dimensional distorted probabilities. They are measures that can be decomposable into a set of probabilities and a function to combine them.

**Definition 9.** *[14] $\mu$ is an at most $m$-dimensional distorted probability if there exists a function $f$ on $[0, 1]^m$, a partition $\mathcal{P} := \{X_1, X_2, \cdots, X_m\}$ of $X$, and probabilities $P_i$ on $(X_i, \wp(X_i))$ such that*

$$\mu(A) = f(P_1(A \cap X_1), P_2(A \cap X_2), \cdots, P_m(A \cap X_m)) \tag{8}$$

*where $f$ is strictly increasing with respect to the i-th axis for all $i = 1, 2, \ldots, m$.*

*An at most $m$-dimensional distorted probability $\mu$ is said to be an $m$-*dimensional distorted probability *if it is not an at most $m - 1$ dimensional one.*

This family of measures is similar to $k$-order additive ones because it also covers all the space of fuzzy measures. Note that any fuzzy measure can be expresssed using the partition $X_i = \{x_i\}$ for all $i$ in $\{1, \ldots, |X|\}$.

Another generalization of additive fuzzy measures is obtained replacing *addition* by another operator. This is the case of $\perp$-decomposable fuzzy measures where the addition is replaced by a t-conorm. A t-conorm (see e.g., [9]) is a binary operator ($\perp : [0, 1] \times [0, 1] \rightarrow [0, 1]$) extensively used in fuzzy logic that satisfies: (i) symmetry ($\perp(x, y) = \perp(y, x)$); (ii) associativity ($\perp(\perp(x, y), z) = \perp(x, \perp(y, z))$); (iii) monotonicity ($\perp(x, y) \leq \perp(x', y')$ if $x \leq x'$ and $y \leq y'$); and (iv) that its neutral element is 0 ($\perp(x, 0) = x$ for all $x$).

The maximum, and the bounded sum ($\perp(x, y) = \min(1, x + y)$) are examples of t-conorms.

**Definition 10.** *(see e.g., [5, 25]) $\mu$ is a $\perp$-decomposable fuzzy measure if for all $A, B \subseteq X$ with $A \cap B = \emptyset$ we can express it as:*

$$\mu(A \cup B) = \mu(A) \perp \mu(B)$$

*for a given t-conorm $\perp$.*

Sugeno $\lambda$-measures are an example of $\perp$-decomposable fuzzy measures. Such measures are defined as follows:

**Definition 11.** *[17] $\mu$ is a Sugeno $\lambda$-measure if for some fixed $\lambda > -1$ we have that:*

$$\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B) \tag{9}$$

*for all* $A \cap B = \emptyset$

Hierarchically $\perp$-decomposable fuzzy measures [19] are a generalization of $\perp$-decomposable fuzzy measures. The generalization starts from a hierarchy of the elements in $X$. Then, a t-conorm is assigned to each node, and such t-conorm is used to combine the measures of subsets.

The generalization comes from the fact that in a $\perp$-decomposable fuzzy measure, a single t-conorm is used for all combination of pairs of measures $\mu(A)$ and $\mu(B)$. Instead, in a hierarchically decomposable one different t-conorms might be used for combining different subsets.

## 4 Parameter Determination

We consider in this section the problem of parameter determination for the Choquet integral. There are different approaches for solving this problem. One of them is based on learning parameters from examples. That is, we have a set of examples or cases. Each example consists on the values that the information sources give and the expected output of the aggregation for such inputs. So, formally speaking, for a set of sources $x_1, \ldots, x_N$ a set of $M$ examples consists on the inputs $f^j(x_i)$ for $j = 1, \ldots, M$ and $i = 1, \ldots, N$ and an outcome $o^j$ for $j = 1, \ldots, M$.

Then, the goal is to obtain a fuzzy measure $\mu$ such that

$$CI_\mu(f^j(x_1), \ldots, f^j(x_N))$$

is as much similar as possible to $o^j$ for all $j = 1, \ldots, M$. Usually, the goodness of the solution is evaluated using the Euclidean distance.

The most general case is to consider fuzzy measures with no additional constraints (we call such measures unconstrained fuzzy measures). That is, only satisfying the axioms on boundary conditions and on monotonicity.

Up to our knowledge, this problem was first considered by Mori and Murofushi in [11, 23]. They showed that this is a quadratic problem with linear constraints.

To solve the problem we need to represent the objective function and the constraints. Software exists that solve quadratic problems given the constraints. The formulation of the objective function and the constraints can either be using $\mu(A)$ for all non empty sets $A \subseteq X$ or using the Möbius transform of $\mu$. Both formulations are equivalent. We use the formulation based on the Möbius because it permits to represent in a simple way some constrained fuzzy measures.

We now outline some results related with the process of learning some constrained fuzzy measures for a Choquet integral:

**Additive fuzzy measures:** This problem can be solved easily adding $m(A) = 0$ for all $|A| > 1$ to the optimization problem for unconstrained fuzzy measures. Alternatively, we can remove the corresponding variables from the model. Note that the learning of additive fuzzy measures for a Choquet integral corresponds to the learning of a weighting vector in a weighted mean. Naturally, this latter problem is much easier solved from a computational point of view than the original one with the Choquet integral.

***k*-order additive fuzzy measures:** This situation corresponds to the quadratic problem with linear constraints requiring $m(A) = 0$ for all $A$ with a cardinality larger than $k$. As for the case of the additive fuzzy measures, we can just rewrite the constraints and the objective function removing such variables.

**Symmetric fuzzy measures:** Such measures can also be determined using a quadratic problem with linear constraints. Symmetry of the measure implies that the Möbius transform only depends on the cardinality of the measure. For this kind of measures, we can reformulate the problem so that only the $(w_1, \ldots, w_{|X|})$ in (5) are considered. Note that these $w_i$ are positive and they add to one. This is again a quadratic problem with linear constraints. The determination of a symmetric fuzzy measure for the Choquet integral corresponds to the determination of the weights for the OWA operator. Again, solving the problem for the OWA is much easier from a computational point of view than the general one for the Choquet integral.

**Belief measures:** This corresponds to the general optimization problem adding that the Möbius is always positive. So, we add a constraint for each subset $A$ of $X$.

**Distorted probabilities:** The determination of distorted probabilities cannot be formulated with a quadratic problem with linear constraints. A method based on the gradient descent has been used for this problem. The method bootstraps from the optimal solution of the weighted mean and the OWA. This method that is described in [20, 22] returns both the probability distribution and the distortion function. Note that this problem is equivalent to determine the weights for the WOWA operator.

***m*-dimensional distorted probabilities:** The optimization problem for learning such fuzzy measures leads to a complex optimization problem. The only approach considered so far is based on the gradient descent, as in the case of distorted probabilities. See [14] for details.

Additional details for such methods for parameter determination are explained in [23]. See also the chapter by Beliakov and Calvo in this monograph. They give details on methods for parameter determination. Among other operators, they also consider the case of the Choquet integral. Other fuzzy measures, not considered in this chapter, are also analyzed as e.g. Sugeno λ-measures.

## 5 Conclusions and Future Work

In this paper we have reviewed a few families of fuzzy measures. Fuzzy integrals are useful aggregation operators. Nevertheless, the requirement of $2^{|X|} - 2$ parameters makes their application difficult. So, fuzzy measures with a reduced number of parameters are useful in real applications. Research in this area will go on.

Additionally, families that permit to cover all the space of measures varying a parameter permit to ease the process of finding a compromise between generality and simplicity, or between $2^{|x|}-2$ parameters or only $|X|$. New families of measures will be developed satisfying this property.

Additionally, it is necessary to develop new algorithms to determine the measures, either from examples or helping an expert to define them. Such algorithms depend not only on the type of the measure but also on the fuzzy integral used. Research on the other types of fuzzy integrals (Sugeno, twofold [10], etc.) is also needed.

**Acknowledgment** This work was partly funded by the Spanish Ministry of Education and Science under project SEG2004-04352-C04-02 "PROPRIETAS". The author also acknowledges the comments from the referees.

## References

1. Benvenuti, P., Mesiar, R., Vivona, D., (2002), Monotone set functions-based integrals, in E. Pap (Ed.), Handbook of Measure Theory, North-Holland, 1329–1379.
2. Calvo, T., Mayor, G., Mesiar, R., (Eds.) (2002), Aggregation Operators, Physica-Verlag.
3. Choquet, G., (1953/54), Theory of capacities, Ann. Inst. Fourier, 5, 131–295.
4. Dubois, D., Modèles mathématiques de l'imprécis et de l'incertain, en vue d'applications aux techniques d'aide à la décision, Thèse d'Etat, Univ. de Grenoble, 1983.
5. Dubois, D., Prade, H., (1982), A class of fuzzy measures based on triangular norms - A general framework for the combination of uncertain information, Int. J. of General Systems, 8, 43–61.
6. Edwards, W., (1953), Probability-Preferences in Gambling, American Journal of Psychology 66, 349–364.
7. Edwards, W., (1962), Subjective probabilities inferred from decisions, Psychological review 69, 109–135.
8. Grabisch, M., (1997), $k$-order additive discrete fuzzy measures and their representation, Fuzzy sets and systems 92:2, 167–189.
9. Klement, E. P., Mesiar, R., Pap, E., (2000), Triangular Norms, Kluwer Academic Publisher.
10. Imai, H., Torra, V., (2003), On a modeling of decision making with a twofold integral, Proc. EUSFLAT 2003, 714–717.
11. Mori, T., Murofushi, T., (1989), An analysis of evaluation model using fuzzy measure and the Choquet integral, Proc. of the 5th Fuzzy System Symposium, 207–212, in Japanese.
12. Murofushi, T., Sugeno, M., (1991), Fuzzy t-conorm integral with respect to fuzzy measures: Generalization of Sugeno integral and Choquet integral, Fuzzy Sets and Systems, 42, 57–71.
13. Narukawa, Y., Torra, V., (2004), Twofold integral and Multi-step Choquet integral, Kybernetika, 40:1 39–50.
14. Narukawa, Y., Torra, V., (2005), Fuzzy measure and probability distributions: distorted probabilities, IEEE Trans. on Fuzzy Systems, 13:5, 617–629.
15. Rota, G.-C., (1964), On the Foundations of Combinatorial Theory. I. Theory of Möbius Functions, Z. Wahrscheinlichkeitstheorie 2, 340–368.
16. Sugeno, M., (1972), Fuzzy Measures and Fuzzy Integrals, Trans. of the Soc. of Instrument and Control Engineers, 8:2, 218–226.
17. Sugeno, M., (1974), Theory of fuzzy integrals and its applications, Ph. D. Dissertation, Tokyo Institute of Technology, Tokyo, Japan.
18. Torra, V., (1997), The Weighted OWA operator, Int. J. of Intel. Syst., 12, 153–166.

19. Torra, V., (1999), On hierarchically S-decomposable fuzzy measures, Int. J. of Intel. Syst., 14:9, 923–934.
20. Torra, V., (2000), Learning weights for Weighted OWA operators, Proc. of the IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation (IECON 2000), 2530–2535.
21. Torra, V., (2003), La integral doble o *twofold* integral: Una generalització de les integrals de Choquet i Sugeno, Butlletí de l'Associació Catalana d'Intel·ligencia Artificial 29 (2003) 13–19. (Preliminary version in English: Twofold integral: A generalization of Choquet and Sugeno integral, IIIA Technical Report TR-2003-08).
22. Torra, V., (2004), OWA operators in data modeling and reidentification, IEEE Trans. on Fuzzy Systems, 12:5, 652–660.
23. Torra, V., Narukawa, Y., (2006), Modeling decisions: Information fusion and aggregation operators, forthcoming.
24. Vitali, G., (1925), Sulla definizione di integrale delle funzioni di una variabile, Annali di Matematica Serie IV, Tomo II 111–121.
25. Weber, S., (1986), Two integrals and some modified versions – critical remarks, Fuzzy Sets and Systems 20, 97–105.
26. Yager, R. R., (1988), On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. on Systems, Man and Cybernetics, 18, 183–190.

# Interpolatory Type Construction of General Aggregation Operators

**Gleb Beliakov and Tomasa Calvo**

**Abstract** In this chapter we examine a number of methods to construct aggregation operators of interpolatory type for specific applications. The construction is based on the desired values of the aggregation operator at certain prototypical points, and on other desired properties, such as conjunctive, disjunctive or averaging behaviour, symmetry and marginals.

## 1 Introduction

Construction of aggregation operators is an important practical issue when building fuzzy systems. There exist a large number of families of aggregation operators, with a wide range of properties, and the choice of the operator suitable for a particular system is not simple. Overviews of the most important families of aggregation operators can be found in [11, 14].

Construction methods typically depart from the existing classes of operators, or from certain theoretical properties (such as functional equations [11]). In this article we consider alternative constructions which are based on the desired values of the operator at certain points (or subsets).

The main idea is to obtain the values of an aggregation operator at certain meaningful points (that would be easily assessable by the decision maker), and determine the aggregation operator on the whole of its domain by interpolation, or approximation. The reason behind is that it is extremely difficult, if not impossible, to extract such specific information as an algebraic form, or a full set of desired mathematical properties form the experts in the field when building a particular fuzzy system. On the other hand, it is much easier to obtain the numerical values of the aggregation operator for specific inputs. When working with the experts, this amounts to presenting them with a number of particular cases and asking for their judgement in each case – a rather simple and straightforward procedure [47]. Alternatively, it is often possible to observe and record the decision patterns of the experts and extract information from there. Finally, one can automate this process in online decision support or e-commerce systems and obtain the data by recording users' actions.

Besides the raw data, it is often possible to use certain general properties of the desired aggregation operator, such as commutativity, idempotency, and conjunctive

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

or disjunctive behaviour. These properties can be directly interpreted by the experts, but they are not sufficient by themselves to construct a specific family of aggregation operators. However, when they are used with the data, they help to construct a much tighter approximation to the desired aggregation function.

The methods discussed in this article will be referred to as pointwise, or interpolatory type constructions. These methods do not recover any specific algebraic formula, but produce an algorithm, capable of calculating the value of the aggregation operator for any input arguments. It is similar to black-box type constructions when using, e.g., neural networks, although the presented methods must also ensure consistency with the fundamental properties of aggregation operators. While such algorithms may not be suitable for manual calculations, for computer systems they are as good as having an algebraic formula in terms of efficiency and correctness of results.

This approach has recently attracted attention of a number of researchers. For instance, in [32, 36] the authors consider construction of aggregation operators based on their marginal values and diagonal sections. M. Grabisch [26] proposed piecewise linear interpolatory method for unipolar and bipolar operators, which recovers Choquet and Sugeno integrals. In statistical literature an issue of identifying copulas (which can be viewed as special cases of aggregation operators) was dealt with in [41], and recently in [20].

In this article we consider pointwise constructions of several important families of aggregation operators and also of general aggregation operators. In the first case we will rely on univariate generating functions which define $n$-variate aggregation operators by means of certain equations. We examine in detail associative aggregation operators (triangular norms, triangular conorms, uninorms and nullnorms) and other generated operators such as quasi-arithmetic means (Sect. 7). In the case of general aggregation operators, we examine tensor product monotone splines (Sect. 4) and optimal Lipschitz interpolation (Sect. 5).

## 2 Preliminaries

Overviews of aggregation operators and their properties are given in [11, 14, 15], and also in this volume [28]. We list only the relevant definitions.

**Definition 1.** An *aggregation operator* is a function $F\colon \bigcup_{n\in\mathbb{N}} [0,1]^n \to [0,1]$ such that:

(i) $F(x_1,\ldots,x_n) \le F(y_1,\ldots,y_n)$ whenever $x_i \le y_i$ for all $i \in \{1,\ldots,n\}$.
(ii) $F(t) = t$ for all $t \in [0,1]$.
(iii) $F(\underbrace{0,\ldots,0}_{n-times}) = 0$ *and* $F(\underbrace{1,\ldots,1}_{n-times}) = 1$

Each aggregation operator $F$ can be represented by a family of $n$-ary operators $f_n : [0,1]^n \to [0,1]$ given by $f_n = F|_{[0,1]^n}$ for $n = 1,2,3,\ldots$. It is possible use such a representation to formulate the properties of aggregation operators: a given property holds for $F$ if and only if it holds for all $f_n, n = 2, 3, \ldots$

Let $t, e, a \in [0, 1]$. $\mathrm{e}(t, i)$ will denote a vector whose components are all $e$ except the $i$-th component: $\mathrm{e}(t, i) = (e, \ldots, e, t, e, \ldots, e)$. $\mathrm{a}(x, i)$ will denote a vector whose $i$-th component is $a$: $\mathrm{a}(x, i) = (x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n)$.

- An aggregation operator $F$ is called averaging if all $f_n$ are bounded by

$$\min(x) = \min_{i=1,\ldots,n} x_i \le f_n(x) \le \max_{i=1,\ldots,n} x_i = \max(x).$$

- An aggregation operator is called idempotent if $f_n(t, t, \ldots, t) = t$ for any $t \in [0, 1]$ and any $n \in \mathbb{N}$. Note that since aggregation operators are monotone, idempotency is equivalent to the averaging behaviour.
- An aggregation operator $F$ is called conjunctive if all $f_n$ are bounded by

$$f_n(x) \le \min(x).$$

- An aggregation operator $F$ is called disjunctive if all $f_n$ are bounded by

$$\max(x) \le f_n(x).$$

- An aggregation operator $f_n$ is called symmetric (commutative) if $f_n(x) = f_n(x_P)$ for any $x \in [0, 1]^n$ and any $P$, where $x_P$ is a permutation of the components of $x$.
- A binary aggregation operator has a neutral element $e \in [0, 1]$ if $\forall t \in [0, 1]$, $f_2(t, e) = f_2(e, t) = t$. In the general case, $F$ has a neutral element $e$ if for any $n > 2$

$$f_n(x_1, \ldots, x_{i-1}, e, x_{i+1}, \ldots, x_n) = f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

for $e$ at any position within the vector $x$. Applying this formula iteratively, one obtains

$$\forall t \in [0, 1], i \in \{1, \ldots, n\} : f_n(\mathrm{e}(t, i)) = t. \tag{1}$$

- An aggregation operator $F$ has an annihilator $a \in [0, 1]$ if $\forall n, \forall x \in I^n$ : $f_n(\mathrm{a}(x, i)) = a$ for all $i \in \{1, \ldots, n\}$.

Throughout this paper the set of empirical data (dataset) will be denoted by $\mathcal{D} = \{(x^k, y^k)\}_{k=1}^K$. It consists of $K$ input-output tuples, $x^k \in [0, 1]^n$, $y^k \in [0, 1]$. We concentrate on pointwise construction of $n$-ary aggregation operators $f_n(x)$ from the dataset $\mathcal{D}$, possibly subject to some additional properties identified by the experts. Thus we require from the function $f_n$ the following minimal set of conditions: to be monotone increasing in each argument and satisfy $f_n(0, 0, \ldots, 0) = 0$ and $f_n(1, 1, \ldots, 1) = 1$. Let us denote the set of all monotone increasing functions of $n$ arguments by $Mon$. The generic mathematical problem we solve is

**Problem 1**
To find $f_n(x) \in Mon$, $f_n(x^k) = y^k, k = 1, \ldots, K$, such that $f_n(0, \ldots, 0) = 0$ and $f_n(1, \ldots, 1) = 1$.

Sometimes it will be impossible to find an exact solution to Problem 1 by a given method, perhaps due to inaccuracies in the data. In this case we will look for a solution that fits the data best by minimizing

$$\left( \sum_{k=1}^{K} |f_n(x^k) - y^k|^p \right)^{1/p}, p \geq 1, \tag{2}$$

subject to $f_n(x) \in Mon$, $f_n(0, \ldots, 0) = 0$, $f_n(1, \ldots, 1) = 1$ (plus other specific properties). We concentrate on the cases $p = 2$ (constrained least squares regression) and $p = 1$ (constrained least absolute deviation problem). The latter case is less sensitive to outliers and may present computational advantages.

There are few general methods of monotone multivariate interpolation that can be used for scattered data, i.e., when the data abscissae $x^k, k = 1, \ldots, K$ are not structured (e.g., lie on a rectangular grid). Scattered data are typical when the data comes from observations, therefore we concentrate on that case. We consider two groups of methods.

Methods in the first group reduce the approximation problem to the univariate case, in which monotone approximation methods are abundant. Methods that rely on approximation of additive generators and tensor product schemata belong to this group. Methods in the second group deal with the multivariate problem directly.

## 3 Associative and Generated Aggregation Operators

Methods in this group rely on the existence of a univariate function with certain properties, which generates multivariate aggregation operators through various equations. The approach is to identify the generating function, and then to construct the aggregation operator itself. The results in this section are taken from [4, 6, 9].

### *3.1 Triangular Norms and Conorms*

Triangular norms and conorms (t-norms and t-conorms) are associative and commutative aggregation operators with the neutral element $e = 1$ and $e = 0$ respectively. A comprehensive overview of these operators is given in [35], see also [11, 33]. Since t-norms and t-conorms are dual to each other ($T(x, y) = 1-C(1-x, 1-y)$), it will be sufficient to deal with t-norms only, as the results for t-conorms are obtained by duality. We concentrate on continuous t-norms. An important class of t-norms is called Archimedian t-norms. These are the ones satisfying

$$\forall(x, y) \in [0, 1]\ \exists n \in \mathbb{N} : f_n(x, \ldots, x) < y.$$

For the purposes of approximation it is sufficient to study Archimedian t-norms, because every continuous non-Archimedian t-norm can be approximated arbitrary well with continuous Archimedian t-norms [31, 30].

Continuous Archimedian t-norms $T$ possess continuous additive generators $g : [0, 1] \to [0, \infty]$, such that

$$T(x_1, x_2, \ldots, x_n) = g^{(-1)}\left(\sum_{i=1}^{n} g(x_i)\right), \tag{3}$$

where the pseudoinverse is defined by

$$g^{(-1)}(t) = \begin{cases} g^{-1}(t) & \text{if } t \le g(0), \\ 0 & \text{otherwise.} \end{cases}$$

Additive generators $g$ are strictly monotone decreasing, and satisfy $g(1) = 0$ and either $g(0) = \infty$ (strict t-norms) or $g(0) = a < \infty$ (nilpotent t-norms). They are defined up to an arbitrary positive multiplier, so with no loss of generality we put $g(0) = 1$ for nilpotent t-norms and $g(\varepsilon) = 1$ for strict t-norms, where $\varepsilon$ will be defined later.

It is possible to fit continuous Archimedian t-norms to data by approximating their additive generators on [0, 1]. This is because the convergence of a sequence of additive generators on [0, 1] is equivalent to the convergence of the corresponding t-norms [30, 31, 35]. Since neither the shape of $g$, nor its algebraic form is fixed, we will represent $g$ with a spline $S$, given as a linear combination of some basis functions $B_j$

$$S(t) = \sum_{j=1}^{J} c_j B_j(t). \tag{4}$$

We will use linear splines (i.e. piecewise linear continuous functions) since there are no requirements on smoothness of the additive generator. Of course, if smoothness is desired, splines of higher order can be used. The knots of the spline (where the linear pieces are joined together) will not be bound to the data, but chosen a priori. Such splines are called regression splines. Our goal is to determine spline coefficients $c_j$. In a suitably chosen basis $B_j$, the condition of monotonicity translates in simple negativity requirement [3, 2] (such a basis consists of linear combinations of standard B-splines).

Consider (3). For an observation $k$ write

$$g(x_1^k) + g(x_2^k) + \ldots + g(x_n^k) \approx g(y^k). \tag{5}$$

$K$ such equations must be satisfied in the least squares sense to fit the generator to the data. We note the difference between the classical approximation problem and the problem of fitting additive generators. In the classical setting the target values $y^k$ are given. In our case we have a set of equations (5) which have the unknown function $g$ on both sides of the equation.

Replacing $g$ with spline $S$ we obtain

$$\sum_{i=1}^{n}\Big(\sum_{j=1}^{J} c_j B_j(x_i^k)\Big) - \sum_{j=1}^{J} c_j B_j(y^k) \approx 0,\ k = 1, \dots, K. \tag{6}$$

By rearranging the terms, (6) becomes

$$\sum_{j=1}^{J} c_j \Big(\sum_{i=1}^{n} B_j(x_i^k) - B_j(y^k)\Big) \approx 0,\ k = 1, \dots, K. \tag{7}$$

In order to find the unknown coefficients $c_j$ we need to solve the system of linear equations (7), subject to restrictions of monotonicity ($c_j \leq 0$) and

$$\sum_{j=1}^{J} c_j B_j(1) = 0, \tag{8}$$

$$\sum_{j=1}^{J} c_j B_j(0) = 1 \text{ or } \sum_{j=1}^{J} c_j B_j(\varepsilon) = 1, \tag{9}$$

depending on whether we want a nilpotent or a strict t-norm.

Problem (7)-(9) is called the linear least squares problem with equality and inequality constraints (LSEI). It was studied in detail in [23, 24, 26], and one of the methods of its solution is based on active set methods used in quadratic programming. One implementation is available from `http://www.netlib.org` as *Algorithm 587*. The problem is formulated as a system of equations and inequalities

$$\mathrm{Ax} \approx \mathrm{b}, \mathrm{Cx} = \mathrm{c}, \mathrm{Dx} \geq \mathrm{d}, \tag{10}$$

where the first system is satisfied in the least squares sense, and the other two systems are satisfied exactly.

In our case, $\mathrm{D} = -\mathrm{I}$, $\mathrm{d} = 0$, matrix C is defined by (8)-(9), $\mathrm{b} = 0$, and the entries of A are given by

$$A_{kj} = \sum_{i=1}^{n} B_j(x_i^k) - B_j(y^k).$$

Solution of (10) yields the regression spline (11), which approximates the additive generator and defines the operator through (3). We should note that in principle, we can use any set of basis functions $B_j$ in (11), not necessarily B-splines, (e.g., generalized polynomials), and problem (7) will remain unchanged. However the condition of monotonicity will take a more complicated form. The advantage of using B-spline basis is the simplicity of such condition, expressed as non-positivity of spline coefficients c.

One important subclass of Archimedian t-norms, Archimedian copulas [35, 40], can be specified by introducing further constraints on the coefficients $c_j$, namely $c_{j+1} - c_j \geq 0$, $j = 1, \ldots, J-1$. These constraints force the spline $S$ to be convex, and this is precisely what is needed to characterize Archimedian copulas.

As far as the value of $\varepsilon > 0$ in condition (9) in concerned, we use the following method. We need to model the asymptotic behaviour of $g$ near 0 for strict t-norms. To model this behaviour we use "well-founded" generators [31],

$$g(x) = \begin{cases} \frac{1}{x} + S(\varepsilon) - \frac{1}{\varepsilon} & \text{if } x \leq \varepsilon; \\ S(x) & \text{if } x > \varepsilon. \end{cases} \tag{11}$$

We choose $\varepsilon$ in such a way that $\varepsilon < \min_{k=1,\ldots,K, i=1,\ldots,n}\{x_i^k, y^k\}|x_i^k, y^k > 0$, i.e. smaller than the smallest observed nonzero data value. In this case we have freedom of choosing $g$ on the interval $[0, \varepsilon]$, as these values of $g$ are never used in any computations with the data set. It is natural to fix a value of 1 at $\varepsilon$ as in (9), but other choices are also suitable.

*Example 1.* Let us take a (synthetic) data set $\mathcal{D}$ of $K = 20$ tuples $(x^k, y^k)$, $x^k \in [0,1]^3$, $y^k \in [0,1]$, and fit an additive generator of a strict triangular norm. Figure 1 shows a linear spline approximation to the model additive generator of the Dombi t-norm, used to generate the data,

$$g_\lambda^D = \left(\frac{1-x}{x}\right)^\lambda, \text{ with } \lambda = 0.2.$$

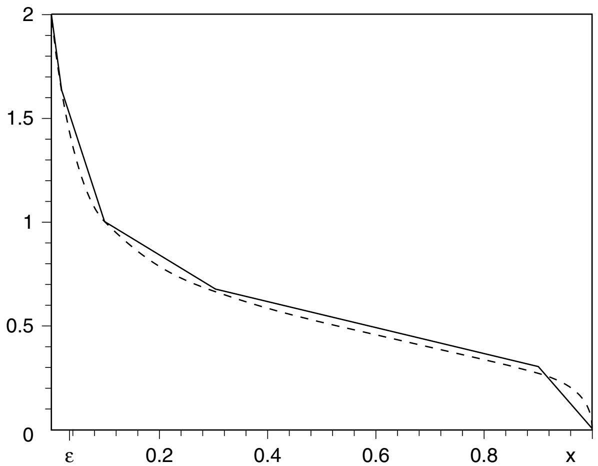



**Fig. 1** The additive generator of the Dombi t-norm (*dashed line*) and its linear spline approximation (*thin solid line*)

### 3.2 Uninorms and Nullnorms

Uninorms are a generalization of t-norms and t-conorms, obtained by allowing the neutral element $e : U(x, e) = x$ reside anywhere in [0, 1] [46]. Like t-norms, these functions are monotone, commutative and associative. On $[0, e]^n$ the uninorm coincides with some (scaled) t-norm, and on $[e, 1]^n$ it coincides with some (scaled) t-conorm. In the rest of the domain $min(x_1, \ldots, x_n) \leq U(x_1, \ldots, x_n) \leq max(x_1, \ldots, x_n)$, and it can be either conjunctive ($U(0, 1, \ldots, 1) = 0$) or disjunctive ($U(1, 0, \ldots, 0) = 1$).

We consider continuous *representable* uninorms [24, 34], that have an associated continuous additive generator $g : [0, 1] \rightarrow [-\infty, \infty], g(0) = -\infty, g(1) = \infty, g(e) = 0$, such that

$$U(x_1, x_2, \ldots, x_n) = g^{-1}\left(\sum_{i=1}^{n} g(x_i)\right). \tag{12}$$

The generator $g$ is strictly increasing, and the representable uninorm is linked to a pair of strict Archimedean t-norm and t-conorm. Continuity of $U$ on $[0, 1]^2 \setminus \{(1, 0), (0, 1)\}$ and strict monotonicity on $]0, 1[^2$ is equivalent to the existence of a strictly increasing additive generator [35].

The approach is similar to the case of t-norms and involves fitting additive generators. Consider first the case of a known (or fixed) neutral element $e$. We translate (12) into (5) for all data points ($k = 1, \ldots, K$). By representing $g$ with spline (11), we obtain (7). In addition we have the restrictions of monotonicity ($c_j > 0$), and boundary conditions

$$\sum_{j=1}^{J} c_j B_j(e) = 0, \sum_{j=1}^{J} c_j B_j(\varepsilon) = -1, \tag{13}$$

where $\varepsilon$ is determined as in the case of t-norms. We use well-founded t-norms representation (11) to model asymptotic behaviour of $g$ near 0 and 1.

In the case of unknown $e$ (so it must also be found from the data), we solve the least squares problem with respect to both $e$ and $c_j$. This is a bilevel optimization problem: at the outer level it is a nonlinear problem with respect to $e$, at the inner level, it is a linear constrained least squares problem

$$\min_{e}\left[\min_{c_j} \sum_{k=1}^{K}\left(\sum_{i=1}^{n} S(x_i^k) - S(y^k)\right)^2 \text{ s.t. (13) and } c_j > 0\right]. \tag{14}$$

The nonlinear problem may have multiple locally optimal solutions, so we use one of the global optimization methods (e.g., combination of grid search and local descent). At each iteration of the global method, we solve the problem with a fixed $e$, which is LSEI.

*Example 2.* Consider another synthetic data set $\mathcal{D}$ of $K = 60$ tuples $(x^k, y^k)$, $x^k \in [0, 1]^2$, $y^k \in [0, 1]$, in which the values $y^k = U(x^k)$ are obtained by using $3 - \Pi$ operator

$$U(x_1, x_2, \ldots, x_n) = \frac{\prod x_i}{\prod x_i + \prod(1 - x_i)},$$

with the convention 0/0=0 (for a conjunctive uninorm), and an additive generator $g(x) = \log(\frac{x}{1-x})$. We fit its additive generator together with the value $e$ as in (14). Figure 2 shows a linear spline approximation to the additive generator. The method correctly identified the value $e = 0.5$ (the computed value was $e = 0.498$).

Nullnorms, introduced in [14], are monotone, commutative and associative functions $[0, 1]^2 \to [0, 1]$, with boundary conditions $N(0, x) = x$ for $x \in [0, a]$ and $N(1, x) = x$ for $x \in [a, 1]$. The value $a \in [0, 1]$ is called the annihilator, so that $N(a, x) = a$ for $x \in [0, 1]$. The above conditions result in that a nullnorm is a (scaled) t-norm on $[a, 1] \times [a, 1]$, a t-conorm on $[0, a] \times [0, a]$, and takes a constant value $a$ elsewhere on the unit square.

Similarly to uninorms, let us consider a class of nullnorms linked to a pair of Archimedian t-norm and t-conorm. We also consider the pair of their respective additive generators $g_t$ and $g_c$. Such nullnorm can be represented as

$$N(x, y) = \begin{cases} g_c^{(-1)}\left(g_c(x) + g_c(y)\right), & \text{if } x, y < a \\ g_t^{(-1)}\left(g_t(x) + g_t(y)\right), & \text{if } x, y > a \\ a, & \text{if either } x \geq a, y \leq a, \text{ or } x \leq a, y \geq a. \end{cases}$$

Note that the situation with nullnorms is different from that with uninorms, where the form of $U(x, y)$ on $]e, 1] \times [0, e[$ and $[0, e[\times]e, 1]$ does depend on the additive generator (and hence the additive generator of a representable uninorm must

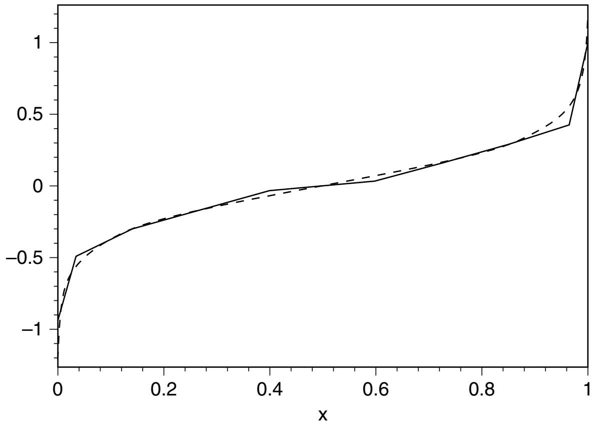



**Fig. 2** The additive generator of the $3 - \Pi$ operator (*dashed line*) and its linear spline approximation (*thin solid line*)

satisfy $g(0) = -\infty$, $g(1) = \infty$, among other conditions). In the case of nullnorms, $N(x, y)$ is uniquely defined on the above mentioned part of its domain, and no extra conditions on $g_t$, $g_c$ are necessary. Hence the participating t-norm and t-conorm can be either strict or nilpotent. So we are effectively dealing with two *independent* functions on disjoint parts of the domain. We can immediately translate the results on convergence of t-norms and their additive generators [30, 31] to nullnorms (by applying them separately on each part of the domain).

Let us now fit such nullnorm to the data by fitting its pair additive generators. Consider the case of a fixed $a$. Split the data into three groups: the data on $[0, a]^n$, the data on $]a, 1]^n$ and the data elsewhere. The third group of data points does not participate in fitting the generators, since for these data the value of $N$ is the constant $a$, so we discard this group. Now use two linear splines $S_1$, $S_2$ that will approximate two additive generators ($g_c$ on $[0, a]$ and $g_t$ on $[a, 1]$) and fit the splines independently of each other using the same technique as for t-norm/t-conorm in Sect. 3.1.

If the annihilator $a$ is not specified but has to be found from the data as well, we proceed as in the case of uninorms, by minimizing the error of approximation with respect to $a$. This is a nonlinear optimization problem, and we solve it using any global optimization method, at each iteration fitting the generators with a fixed $a$. Of course, the third group of data we discarded previously must also participate in the calculation of error as a function of $a$ (but not in fitting the generators).

### *3.3 Other Generated Aggregation Operators*

Consider now another class of aggregation operators similar to representable uninorms [38]. Let $g : [0, 1] \to [a, b]$, $-\infty < a < b < \infty$ be a monotone increasing function with zero $e \in [0, 1]$. Define

$$f_n(x_1, \ldots, x_n) = g^{(-1)}\left(g(x_1) + \ldots + g(x_n)\right), \tag{15}$$

where $g^{(-1)}$ denotes the pseudoinverse.

The function (15) is continuous on $[0, 1]^n$, but it is not associative. Further, on $[e, 1]^n$ it coincides with a (scaled) nilpotent t-conorm and on $[0, e]^n$ it coincides with a (scaled) nilpotent t-norm. As with uninorms, $e$ is its neutral element, and when $e = 1$ or $e = 0$ we obtain t-norms and t-conorms as limiting cases. The following example is taken from [11, 38].

Let $g(t) = t - \frac{1}{2}$. Then $F$ is an ordinal sum of Lukasiewicz t-norm and t-conorm, given by

$$F(x) = \max(0, \min(1, \frac{1}{2} + \sum_{i=1}^{n}(x_i - \frac{1}{2})))$$

Identification of the generating function $g$ is similar to that of representable uninorms, with the only difference that condition (13) becomes

$$\sum_{j=1}^{J} c_j B_j(e) = 0, \sum_{j=1}^{J} c_j B_j(0) = -1,$$

and no asymptotic behaviour as in (11) is needed. Identification of the neutral element is performed as in (14).

## *3.4 Generalized Means*

Consider the case of quasi-arithmetic means, given by [16, 11]

$$M(x_1, x_2, \ldots, x_n) = g^{(-1)}\left(\sum_{i=1}^{n} \frac{1}{n} g(x_i)\right). \tag{16}$$

$g$ denotes a monotone increasing function $[0, 1] \to [0, 1]$. As with t-norms, we will find $g$ from the data by fitting the equations

$$\frac{1}{n}\left(g(x_1^k) + g(x_2^k) + \ldots + g(x_n^k)\right) \approx g(y^k), k = 1, \ldots, K. \tag{17}$$

By representing generator $g$ with a linear spline (11), we obtain a system of equations and inequalities similar to (7)–(9), namely

$$\sum_{j=1}^{J} c_j\left(\frac{1}{n}\sum_{i=1}^{n} B_j(x_i^k) - B_j(y^k)\right) \approx 0\,, k = 1, \ldots, K, \tag{18}$$

$$\sum_{j=1}^{J} c_j B_j(0) = 0, \sum_{j=1}^{J} c_j B_j(1) = 1, \tag{19}$$

subject to $c_j \geq 0$. Once again we obtain LSEI problem (10), with the components of matrices and vectors defined in (18)–(19). The case of weighted quasi-arithmetic means is considered in detail in this volume in [10].

# 4 Tensor Product Schemata

## *4.1 General Aggregation Operators*

In tensor product schemata a multivariate function is approximated by a tensor product of univariate functions. A method of tensor product monotone regression splines applied to aggregation operators was described in [3, 2, 4, 13]. It consists in fitting the dataset $\mathcal{D}$ with a tensor product spline

$$S(x_1, \dots, x_n) = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \dots \sum_{j_n=1}^{J_n} c_{j_1 j_2 \dots j_n} B_{j_1}(x_1) B_{j_2}(x_2) \dots B_{j_n}(x_n).$$

The univariate basis functions are chosen to be linear combinations of standard B-splines, as in [3], which ensures that the conditions of monotonicity of $S$ are expressed as non-negativity of spline coefficients $c_{j_1 j_2 \dots j_n}$.

The computation of spline coefficients (there are $J_1 \times J_2 \times \dots \times J_n$ of them, where $J_i$ is the number of basis functions in respect to each variable) is performed by solving a quadratic programming problem

$$\text{minimize} \sum_{k=1}^{K} \left( \sum_{j_1=1}^{J_1} \dots \sum_{j_n=1}^{J_n} c_{j_1 j_2 \dots j_n} B_{j_1}(x_1^k) \dots B_{j_n}(x_n^k) - y^k \right)^2, \qquad (20)$$

subject to

$$\sum_{j_1=1}^{J_1} \dots \sum_{j_n=1}^{J_n} c_{j_1 j_2 \dots j_n} \geq 0,$$

and

$$S(0, \dots, 0) = \sum_{j_1=1}^{J_1} \dots \sum_{j_n=1}^{J_n} c_{j_1 j_2 \dots j_n} B_{j_1}(0) \dots B_{j_n}(0) = 0,$$

$$S(1, \dots, 1) = \sum_{j_1=1}^{J_1} \dots \sum_{j_n=1}^{J_n} c_{j_1 j_2 \dots j_n} B_{j_1}(1) \dots B_{j_n}(1) = 1.$$

This is a problem of type LSEI (10), however it involves very sparse matrices. For solving QP with sparse matrices we recommend OOQP sparse solver [19] (`http://www.cs.wisc.edu/~swright/ooqp/`).

In practice it is sufficient to use linear splines with 3-5 basis functions ($J_i = 3, 4, 5$), which gives good quality approximation for 2-5 variables. For more variables the method becomes impractical because the number of spline coefficients (and hence the sizes of all matrices) grows exponentially with $n$. On one hand it requires a large number of data, which is usually not available, on the other hand the required computing time also becomes too large.

*Example 3.* Consider a real data set $\mathcal{D}$ of 22 input-output pairs from [47], and its approximation with a bivariate tensor-product monotone spline $J_1 = J_2 = 4$. The resulting aggregation operator is plotted on Fig. 3.

Other tensor product monotone interpolation methods [17, 18, 19] can be applied to aggregation operators, although in most cases these methods are limited to two

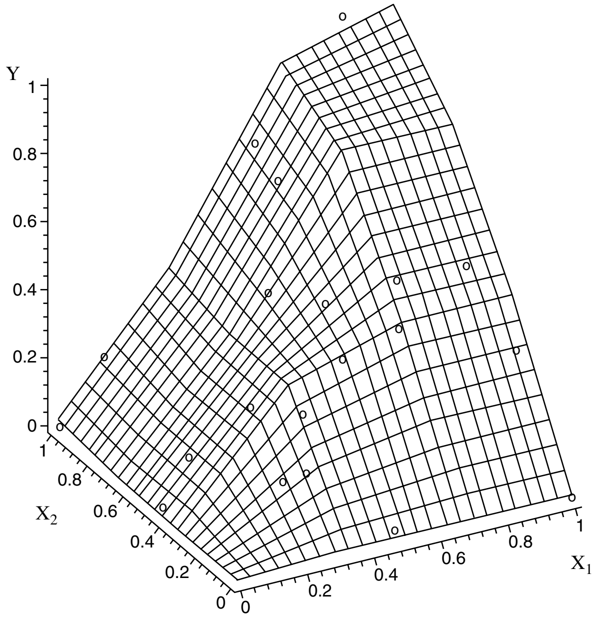



**Fig. 3** Tensor-product spline approximation of the data from [47], denoted with circles

variables. There are also alternative methods for approximation of scattered data based on triangulations [27, 44], in these methods the basis functions are determined by the data. However preservation of monotonicity becomes rather complicated, and the available methods are only suitable for bi-variate case.

## *4.2 Operators with Specific Properties*

It is important to incorporate other problem specific information into the construction of aggregation operators. Such information may be given in terms of boundary conditions, conjunctive, disjunctive or averaging behaviour, symmetry and so on. In this section we describe the method from [4, 2], which accommodates these conditions for tensor product monotone splines.

Conditions $f_n(t, 0, \ldots, 0) = t$ and $f_n(t, 1, \ldots, 1) = t$ are associated with the neutral element, and arise naturally in many applications. It is shown that for linear tensor product spline, it is sufficient to use the interpolation conditions $S(t_j, 0, \ldots, 0) = t_j$ or $S(t_j, 1, \ldots, 1) = t_j$ for $j = 1, \ldots, J_1$ respectively, where $t_j$ denote the knots of the spline. Similar conditions are imposed for other variables. These conditions are incorporated easily into the problem (20), (or (10)) as linear equalities.

Condition of idempotency $f_n(t, \ldots, t) = t$ is equivalent to the averaging behaviour of the aggregation operator. This condition can be enforced also by using a number of interpolation conditions $S(t_j, t_j, \ldots, t_j) = t_j$, $j = 1, \ldots, M$, but now $t_j$

are not the knots of the splines. The values of $t_j$ can be chosen with relative freedom with $M \geq n + 1$, see [2, 4].

Condition of symmetry is imposed by restricting the domain of the approximating function to the simplex defined by $x_1 \geq x_2 \geq \ldots \geq x_n$, $x_i \in [0, 1]$, and then extending $f_n$ to the rest of the domain by symmetry. In practical terms, it requires converting the dataset into $\mathcal{D}_{()} = \{(z^k, y^k)\}$, where vectors $z^k = x^k_{()}$ have their components arranged in decreasing order. When this dataset is used, the number of tensor products and spline coefficients is reduced by the factor of $1/n!$.

## 5 Optimal Lipschitz Interpolation

The method of monotone Lipschitz interpolation belongs to the second group of methods, in which the multivariate interpolation problem is tackled directly. This method is described in detail in [6, 7], and it was applied to aggregation operators in [8].

We will denote by $Lip(M, ||\cdot||)$ the set of all Lipschitz continuous functions on $[0, 1]^n$ with the Lipschitz constant in the norm $||\cdot||$ smaller or equal to $M$:

$$Lip(M, ||\cdot||) = \{f : [0, 1]^n \to [0, 1] | \, \forall x, y \in [0, 1]^n, \; |f(x) - f(y)| \leq M||x - y||\}.$$

Lipschitz-continuous aggregation operators are very important for applications, because they provide output values stable with respect to small changes of the arguments. Small changes in the arguments may be due to inaccuracies in the data, and one would expect that such inaccuracies do not affect drastically the behavior of the system. The concept of $p$-stable aggregation operators was proposed in [16]. These are precisely Lipschitz continuous operators whose Lipschitz constant $M$ in $l_p$ norm is one. Specific cases include 1-Lipschitz aggregation operators ($p = 1$) and kernel aggregation operators ($p = \infty$). Quasi-copulas arise as a special case of 1-Lipschitz operators, when the neutral element $e = 1$.

### *5.1 Optimal Interpolation*

Consider a dataset $\mathcal{D}$ and the requirement that the interpolating function is Lipschitz and monotone, $f_n \in Lip(M) \cap Mon$, and also satisfies the boundary conditions $f_n(0, \ldots, 0) = 0$, $f_n(1, \ldots, 1) = 1$. There are many functions from this class that interpolate the data. The information about the function $f_n$ as given above is not sufficient to find $f_n$, but it identifies a closed set $\mathcal{S}$ of functions consistent with this information. Whatever the algorithm used to determine $f_n$, its error cannot be smaller that the distance between the center of $\mathcal{S}$ and its farthest element, referred to as the radius of information [43]. The radius of information is the intrinsic error of the problem.

An optimal algorithm is the one whose error is the same as the intrinsic error of the approximation problem. Central algorithm is the algorithm whose solution is the center of the set $\mathcal{S}$, and it is always optimal [43]. However, the center of $\mathcal{S}$ may lie outside $\mathcal{S}$, in which case the central algorithm does not provide a solution consistent with the specified properties.

In the case of monotone Lipschitz functions, the centre of $\mathcal{S}$ is inside, and the central algorithm delivers an optimal solution consistent with the requirements. This solution minimizes the worst case error, i.e., solves the problem

$$\min_{g\in\mathcal{F}}\max_{f_n\in\mathcal{F}}\max_{x\in[0,1]^n}|f_n(x)-g(x)|$$
$$\text{subject to } g(x^k)=f_n(x^k)=y^k, k=1,\dots,K,$$

where $F = Lip(M)\cap Mon$.

In [6] an explicit solution to this problem is given. For each $x\in[0,1]^n$ tight upper and lower bounds on $f_n$ are identified: $\sigma_l(x)\le f_n(x)\le\sigma_u(x)$,

$$\begin{aligned}\sigma_u(x)&=\min_k\{y^k+M||(x-x^k)_+||\},\\ \sigma_l(x)&=\max_k\{y^k-M||(x^k-x)_+||\},\end{aligned}\tag{21}$$

where $z_+$ denotes the positive part of vector $z$: $z_+=(\bar{z}_1,\dots,\bar{z}_n)$, with

$$\bar{z}_i=\max\{z_i,0\}.$$

If the data set is infinite, say, $\mathcal{D}=\{(t,v(t))|t\in\Omega\subset[0,1]^n, v:\Omega\to[0,1]\}$ then the bounds translate into

$$\begin{aligned}\sigma_u(x)&=\inf_{t\in\Omega}\{v(t)+M||(x-t)_+||\},\\ \sigma_l(x)&=\sup_{t\in\Omega}\{v(t)-M||(t-x)_+||\}.\end{aligned}\tag{22}$$

The central optimal algorithm delivers an optimal interpolant

$$g(x)=\frac{1}{2}(\sigma_l(x)+\sigma_u(x)).\tag{23}$$

The functions $\sigma_l(x)$,$\sigma_u(x)$ and $g(x)\in Lip(M,||\cdot||)\cap Mon$.

Thus construction of a general aggregation operator from the dataset $\mathcal{D}$ involves application of (21),(23), which is straightforward. Note that if no empirical data is present (i.e., when the dataset contains only two compulsory data $((0,\dots,0),0),((1,\dots,1),1))$, we obtain general bounds on $p$-stable aggregation operators given by Yager t-norm and t-conorm

$$\begin{aligned}\sigma_l(x)&=T_Y(x)=\max\{0,1-||1-x||_p\},\\ \sigma_u(x)&=S_Y(x)=\min\{1,||x||_p\}.\end{aligned}$$

## *5.2 Application-specific Properties*

Our next goal is to incorporate other problem-specific information into construction of optimal aggregation operators. This information may come in form of the following generic conditions.

- Symmetry;
- Conjunctive and disjunctive behavior;
- Idempotency;
- Neutral element and annihilator;
- Given marginals;
- Given diagonal and opposite diagonal;
- Various combinations of the above.

Symmetry can be enforced in the same way as at the end of Sect. 4. Conjunctive and disjunctive behaviour imply additional bounds $f_n(x) \leq \min(x)$ or $\max(x) \leq f_n(x)$, and idempotency implies two bounds $\min(x) \leq f_n(x) \leq \max(x)$. These bounds are in addition to $\sigma_l, \sigma_u$; this can be formally expressed as

$$\underline{A}(x) \leq f_n(x) \leq \overline{A}(x),$$

$$\underline{A}(x) = \max\{\sigma_l(x), B_l(x)\}, \;\; \overline{A}(x) = \min\{\sigma_u(x), B_u(x)\}, \tag{24}$$

where $B_l$, $B_u$ are the additional lower and upper bounds arising from application-specific requirements. The optimal aggregation operator is then computed as

$$g(x) = \frac{1}{2}(\underline{A}(x) + \overline{A}(x)). \tag{25}$$

Thus incorporation of the additional properties of aggregation operators amounts to properly defining the extra bounds $B_l$, $B_u$, and then applying (24),(25). In the above mentioned cases we have

- Conjunctive behaviour: $B_l(x) = 0$, $B_u(x) = \min(x)$ ;
- Disjunctive behavior: $B_l(x) = \max(x)$, $B_u(x) = 1$;
- Idempotency: $B_l(x) = \min(x)$, $B_u(x) = \max(x)$.

In some applications the conjunctive, disjunctive or averaging behaviour must be restricted to a certain part of the domain. For example, in [45] Yager introduces the concept of "noble" reinforcement in the context of online recommender systems (such systems recommend certain products to internet customers based on their preferences and/or past purchases). Here it is important to use a symmetric disjunctive aggregation operator which positively reinforces only high values of the arguments. We wish to avoid a situation in which a number of weak "justifications" provide a strong recommendation when combined. Yager proposes to restrict the aggregation operator to $maximum$ for low vales of the arguments.

Such restrictions for Lipschitz aggregation operators were dealt with in [9], and they rely on using (22) for various choices of the domain $\Omega$.

Consider now the case of a given neutral element $e$, for which we examine condition (1) for all $i = 1, \ldots, n$. The bounds implied by this condition are

$$\begin{aligned} \forall x \in [0,1]^n : \quad & B_l(x) \le f_n(x) \le B_u(x), \\ \text{where } B_u(x) &= \min_{i=1,\ldots,n} B_u^i(x), \\ B_l(x) &= \max_{i=1,\ldots,n} B_l^i(x), \end{aligned} \tag{26}$$

where for a fixed $i$ the bounds are

$$\begin{aligned} B_u^i(x) &= \min_{t\in[0,1]} (t + M||(x - \mathrm{e}(t,i))_+||), \\ B_l^i(x) &= \max_{t\in[0,1]} (t - M||(\mathrm{e}(t,i) - x)_+||). \end{aligned} \tag{27}$$

The minimum and maximum in (27) have been found explicitly in [11].

Consider an aggregation operator which has an annihilator $a \in [0, 1]$. The existence of an annihilator does not imply conjunctive or disjunctive behaviour on any part of the domain, but together with monotonicity, it implies $f(x) = a$ on $[a, 1] \times [0, a]$ and $[0, a] \times [a, 1]$ (and their multivariate extensions).

Such restrictions are easily incorporated into the bounds by using

$$\begin{gathered} \max_{i=1,\ldots,n} B_l^i(x) \le f_n(x) \le \min_{i=1,\ldots,n} B_u^i(x), \\ B_l^i(x) = a - M(a - x_i)_+, \\ B_u^i(x) = a + M(x_i - a)_+. \end{gathered} \tag{28}$$

Let us denote by $\delta(t) = f_n(t, t, \ldots, t)$ the diagonal section of the $n$-ary aggregation operator $f_n$. If $f_n \in Lip(M) \cap Mon$, then $\delta \in Lip(Mn^{1/p})$. Also $\delta(t)$ is nondecreasing, and $\delta(0) = 0$, $\delta(1) = 1$. We denote by $\omega(t) = f(t, 1-t)$ the opposite diagonal section of a binary aggregation operator. We note that $\omega \in Lip(M)$. In the following we assume that the functions $\delta(t)$, $\omega(t)$ are given and they have the required Lipschitz properties. The goal is to determine the upper and lower bounds on Lipschitz aggregation operators with these diagonal and opposite diagonal sections.

In the special cases of bivariate 1-Lipschitz aggregation operators and quasicopulas, these bounds were studied in [32]. The general case is treated in [11]. From (21) it follows that

$$\begin{aligned} B_u(x) &= \min_{t\in[0,1]} (\delta(t) + M||((x_1 - t)_+, \ldots, (x_n - t)_+)||), \\ B_l(x) &= \max_{t\in[0,1]} (\delta(t) - M||((t - x_1)_+, \ldots, (t - x_n)_+)||). \end{aligned} \tag{29}$$

Computation of the functions $B_l$, $B_u$ requires solving a univariate optimization problem in (29). The objective function is not necessarily convex (hence a possibility of several locally optimal solutions), however it is Lipschitz, with the Lipschitz constant $Mn^{1/p}$, therefore we can apply Pijavsky-Shubert method [32] to calculate the global optimum.

For binary aggregation operators with given opposite diagonal $\omega(t) = f_2(t, 1-t)$ the bounds are computed as

$$B_u(x) = \min_{t \in [0,1]} (\omega(t) + M||((x_1 - t)_+, (t - (1 - x_2))_+)||),$$
$$B_l(x) = \max_{t \in [0,1]} (\omega(t) - M||((t - x_1)_+, (1 - x_2 - t)_+)||). \quad (30)$$

We notice that $\omega \in Lip(M)$ and so is the second term in the expression, hence the objective function is in $Lip(2M)$. We apply Pijavski-Shubert method with this Lipschitz parameter to calculate thevalues of the bounds for any $x$. The special case of 1-Lipschitz aggregation operators was treated in [32].

*Example 4.* The optimal aggregation operator with a given diagonal section $\delta(t) = \min(2t^2, 1)$, and no empirical data, is presented on Fig. 4. Note that each value has to be computed by solving a global optimization problem (29). The plot on Fig. 4 is done by computing the values of $f_n$ on a $50 \times 50$ mesh, performed in $< 1sec$ on a Pentium IV processor.

*Example 5.* The optimal aggregation operator with the opposite diagonal section $\omega(t) = t(1 - t)$ and neutral element $e = 1$ is presented on Fig. 4 (right).

## 5.3 Operators with Given Marginals

We consider the problem of obtaining the operator $F$ when certain functions are required to be its marginals. For some special cases of 1-Lipschitz aggregation oper-

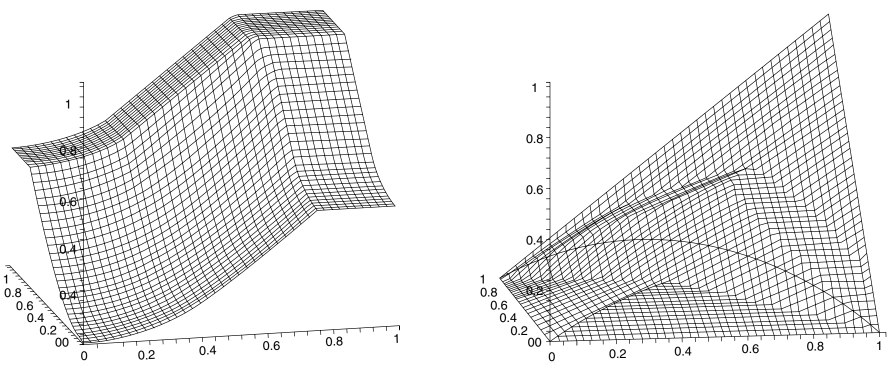



**Fig. 4** The optimal binary aggregation operators with a given diagonal section (**left**), and a given opposite diagonal section and neutral element (**right**)

ators this problem was treated in [36], the general case is presented below. Consider construction of a Lipschitz aggregation operator $f_2$ based on a given marginal $\gamma$, defined on some closed subset $\Omega$, for example $\Omega = \{x = (x_1, x_2) | 0 \le x_1 \le 1, x_2 = 0\}$. Let $\gamma \in Lip(M_\gamma)$. Then obviously the Lipschitz constant of $f$, $M \ge M_\gamma$. From (22) we obtain

$$\begin{aligned} B_u(x) &= \min_{t\in[0,1]} (\gamma(t) + M||((x_1 - t)_+, x_2)||) \\ &= \min_{t\in[0,x_1]} (\gamma(t) + M||((x_1 - t), x_2)||), \\ B_l(x) &= \max_{t\in[0,1]} (\gamma(t) - M||((t - x_1)_+, 0)||) = \gamma(x_1). \end{aligned} \tag{31}$$

If the marginal is given on $\Omega = \{x = (x_1, x_2) | 0 \le x_1 \le 1, x_2 = 1\}$, then the bounds are

$$\begin{aligned} B_u(x) &= \min_{t\in[0,1]} (\gamma(t) + M||((x_1 - t)_+, 0)||) = \gamma(x_1), \\ B_l(x) &= \max_{t\in[0,1]} (\gamma(t) - M||((t - x_1)_+, 1 - x_2)||) \\ &= \max_{t\in[x_1,1]} (\gamma(t) - M||((t - x_1), 1 - x_2)||). \end{aligned} \tag{32}$$

To solve the optimization problem in each case we apply Pijavski-Shubert method with the Lipschitz parameter $M$.

For the general multivariate case the equations are as follows. Let $\gamma_i(t), i = 1, \dots, n$ be a function from $Lip(M_\gamma)$ representing the $i$-th marginal

$$\forall x \in \Omega_i : f_n(x) = \gamma_i(x_i), \Omega_i = \{x \in [0, 1]^n | x_i \in [0, 1], x_j = 0, j \ne i\}$$

The bounds due to the $i$-th marginal are

$$\begin{aligned} B_u^i(x) &= \min_{t\in[0,x_i]} (\gamma_i(t) + M||(x_1, \dots, x_{i-1}, (x_i - t)_+, x_{i+1}, \dots, x_n)||) \\ B_l^i(x) &= \gamma_i(x_i), \end{aligned}$$

and altogether we have $B_l(x) = \max_{i=1,\dots,n} B_l^i(x)$, $B_u(x) = \min_{i=1,\dots,n} B_u^i(x)$.

The same technique is used for construction of $n$-variate aggregation operator from $m$-variate marginals, as exemplified below. Let $\gamma : [0, 1]^m \to [0, 1]$ denote a marginal of $f_n$: $\forall x \in \Omega : f_n(x) = \gamma(y)$, with

$$\Omega = \{x \in [0, 1]^n | x_1, \dots, x_m \in [0, 1], x_{m+1} = \dots = x_n = 0\}$$

and $y_i = x_i, i = 1, \dots, m$. Then the upper and lower bounds on $f_n(x), x \in [0, 1]^n \setminus \Omega$ are

$$\begin{aligned} B_u(x) &= \min_{z\in[0,x_1]\times\ldots\times[0,x_m]} (\gamma(z) + M||((x_1 - z_1)_+, \dots, (x_m - z_m)_+, x_{m+1}, \dots, x_n)||), \\ B_l(x) &= \gamma(x_1, \dots, x_m). \end{aligned}$$

Computation of the minimum in the expression for $B_u$ involves a nonconvex $m$-dimensional constrained optimization problem. There is a possibility of multiple locally optimal solutions, and the use of local descent algorithms will not deliver correct values. These algorithms are typically stuck in suboptimal solutions, which may be different for different $x$, therefore the resulting bounds may be discontinuous. The proper way of calculating $B_u$ is using deterministic global optimization methods. One such method, a multivariate generalization of Pijavski-Shubert algorithm known as the cutting angle method, is described in [5]. One should be aware that deterministic global optimization methods work reliably only in small dimension, $m < 10$. We do not expect $m$ to be greater than 3 in applications.

## 6 Conclusion

We outlined a number of approaches to pointwise construction of aggregation operators from empirical data. Some approaches involve identification of univariate generating functions whereas others deal with the multivariate case directly. The resulting constructions are numerically as efficient as having explicit algebraic formulae, however they provide a much greater flexibility, and do not require domain experts to come up with explicit formulae but rather specific cases. This greatly simplifies knowledge engineering process. At the same time it allows routine automatic construction of suitable aggregation operators from databases, tailoring aggregation procedures to specific cases (e.g., groups of customers), and also evolving them with time as more data becomes available. When new data are added, most methods require re-calculation of the entire set of fitted parameters; in contrast, the optimal interpolation methods from Sect. 5 are incremental.

An important feature of all mentioned approaches is preservation of semantically important application-specific properties. Many such properties can be stated in the form of interpolation conditions (finite or infinite number of such conditions). Incorporation of these conditions is an important part of the construction process.

A software package which implements a number of the mentioned techniques is available from `http://www.deakin.edu.au/~gleb/aotool.html`.

**Acknowledgment** The work of the second author was supported by the projects MTM2004-3175, BFM2003-05308 and TIC2002-11942-E from Ministerio de Educación y Ciencia, Spain, PRIB-2004-9250, Govern de les Illes Balears, and TIN2004-20503-E, Red Temática Nacional sobre Sistemas de Acceso a la Información en la Web basados en Soft-Computing. She is a member of Fundamentos Matemáticos para el Soft-Computing research group.

## References

1. G. Beliakov. Shape preserving approximation using least squares splines. *Approximation theory and applications*, 16: 80–98, 2000.
2. G. Beliakov. Monotone approximation of aggregation operators using least squares splines. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10: 659–676, 2002.

3. G. Beliakov. How to build aggregation operators from data? *Int. J. Intelligent Systems*, 18: 903–923, 2003.
4. G. Beliakov. The cutting angle method - a tool for constrained global optimization. *Optimization Methods and Software*, 19:137–151, 2004.
5. G. Beliakov. Fitting triangular norms to empirical data. In P.E. Klement and R. Mesiar, editors, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*, pp. 255–265. Elsevier, New York, 2005.
6. G. Beliakov. Monotonicity preserving approximation of multivariate scattered data. *BIT*, 45: 653–677, 2005.
7. G Beliakov. Interpolation of Lipschitz functions. *J. of Comp. and Applied Mathematics*, 196: 20–44, 2006.
8. G. Beliakov and T. Calvo. Identification of general and double aggregation operators using monotone smoothing. In E. Montseny and P. Sobrerillo, editors, *EUSFLAT 2005*, pp. 937–942, Barcelona, Spain, 2005. European Society for Fuzzy Logic and Technology.
9. G. Beliakov and T. Calvo. Construction of aggregation operators with noble reinforcement. submitted, 2006.
10. G. Beliakov and T. Calvo. Identification of weights in aggregation operators. In H. Bustince, F. Herrera, and J. Montero, editors, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, page in this volume. Springer, Heidelberg, 2007.
11. G. Beliakov, T. Calvo, and J. Lazaro. Pointwise construction of Lipschitz aggregation operators. In *Information Processing and Management of Uncertainty*, Paris, 2006.
12. G. Beliakov, R. Mesiar, and L. Valaskova. Fitting generated aggregation operators to empirical data. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12: 219–236, 2004.
13. G. Beliakov and J. Warren. Appropriate choice of aggregation operators in fuzzy decision support systems. *IEEE Trans. On Fuzzy Syst.*, 9: 773–784, 2001.
14. T. Calvo, B. De Baets, and J. Fodor. The functional equations of Frank and Alsina for uninorms and nullnorms. *Fuzzy Sets and Systems*, 120: 385–394, 2001.
15. T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Aggregation Operators. New Trends and Applications*, pp. 3–104. Physica-Verlag, Heidelberg, New York, 2002.
16. T. Calvo and R. Mesiar. Stability of aggregation operators. In *2nd Conf. of the Europ. Soc. for Fuzzy Logic and Technology*, pp. 475–478, Leicester, 2001.
17. R.E. Carlson and F.N. Fritsch. Monotone piecewise bicubic interpolation. *SIAM J. Numer. Anal.*, 22: 386–400, 1985.
18. P. Costantini and C. Manni. Monotonicity-preserving interpolation of nongridded data. *Computer Aided Geometric Design*, 13:467–497, 1996.
19. P. Costantini and C. Manni. A local shape-preserving interpolation scheme for scattered data. *Computer Aided Geometric Design*, 16: 385–405, 1999.
20. H. De Meyer and B. De Baets. Fitting piecewise linear copulas to data. In *4th Conf. of the Europ. Soc. for Fuzzy Logic and Technology*, Barcelona, Spain, 2005.
21. D. Dubois and H. Prade. A review of fuzzy set aggregation connectives. *Information Sciences*, 36: 85–121, 1985.
22. D. Dubois and H. Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142:143–161, 2004.
23. H. Dyckhoff and W. Pedrycz. Generalized means as model of compensative connectives. *Fuzzy Sets and Systems*, 14: 143–154, 1984.
24. J. Fodor, R. Yager, and A. Rybalov. Structure of uninorms. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5: 411–427, 1997.
25. E.M. Gertz and S. J. Wright. Object-oriented software for quadratic programming. *ACM Trans. Math. Software*, 29: 58–81, 2003.
26. M. Grabisch. The Choquet integral as a linear interpolator. In *10th Int. Conf. on Information Processing and Management of Uncertainty*, pp. 373–378, Perugia, Italy, 2004.
27. L. Han and L.L. Schumaker. Fitting monotone surfaces to scattered data using c1 piecewise cubics. *SIAM J. Numer. Anal.*, 34:569–585, 1997.

28. R.J. Hanson and K.H. Haskell. Algorithm 587. Two algorithms for the linearly constrained least squares problem. *ACM Trans. Math. Software*, 8: 323–333, 1982.
29. K.H. Haskell and R. Hanson. An algorithm for linear least squares problems with equality and nonnegativity constraints. *Math. Progr.*, 21: 98–118, 1981.
30. S. Jenei. On archimedian triangular norms. *Fuzzy Sets and Systems*, 99: 179–186, 1998.
31. S. Jenei and J. Fodor. On continuous triangular norms. *Fuzzy Sets and Systems*, 100: 273–282, 1998.
32. E.P. Klement and A. Kolesarova. Extension to copulas and quasi-copulas as special 1-Lipschitz aggregation operators. *Kybernetica*, 41: 329–348, 2005.
33. E.P. Klement and R. Mesiar, editors. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, New York, 2005.
34. E.P. Klement, R. Mesiar, and E. Pap. On the relationship of associative compensatory operators to triangular norms and conorms. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4: 129–144, 1996.
35. E.P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer, Dordrecht, 2000.
36. A. Kolesarova, E. Muel, and J. Mordelova. Construction of kernel aggregation operators from marginal values. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 Suppl.: 37–49, 2002.
37. C. Lawson and R. Hanson. *Solving Least Squares Problems*. SIAM, Philadelphia, 1995.
38. R. Mesiar and M. Komornikova. Triangular norm-based aggregation of evidence under fuzziness. In B. Bouchon-Meunier, editor, *Aggregation and Fusion of Imperfect Information*, pp. 11–35. Physica-Verlag, Heidelberg, 1998.
39. R. Mesiar, M. Komornikova, T. Calvo, and A. Kolesarova. A review of aggregation functions. In H. Bustince, F. Herrera, and J. Montero, editors, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, page in this volume. Springer, Heidelberg, 2007.
40. R.B. Nelsen. *An Introduction to Copulas*. Springer, Berlin, 1998.
41. R.B. Nelsen, J.J. Quesada Molina, J.A. Rodriguez Lallena, and M. Ubeda Flores. Best-possible bounds on sets of bivariate distribution functions. *J. of Multivariate Analysis*, 90: 348–358, 2004.
42. S.A. Pijavski. An algorithm for finding the absolute extremum of a function. *USSR Comput. Math. and Math. Phys.*, 2: 57–67, 1972.
43. J.F. Traub and H. Wozniakowski. *A General Theory of Optimal Algorithms*. Academic Press, New York, 1980.
44. K. Willemans and P. Dierckx. Smoothing scattered data with a monotone powell-sabin spline surface. *Numerical Algorithms*, 12: 215–232, 1996.
45. R. Yager. Noble reinforcement in disjunctive aggregation operators. *IEEE Trans. on Fuzzy Syst.*, 11: 754–767, 2003.
46. R. Yager and A. Rybalov. Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80: 111–120, 1996.
47. H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4: 37–51, 1980.

# A Review of Aggregation Functions

**Radko Mesiar, Anna Kolesárová, Tomasa Calvo and Magda Komorníková**

**Abstract** Several local and global properties of (extended) aggregation functions are discussed and their relationships are examined. Some special classes of averaging, conjunctive and disjunctive aggregation functions are reviewed. A special attention is paid to the weighted aggregation functions, including some construction methods.

## 1 Introduction

Aggregation of a finite number of observed values from a scale $I$ into a single output value from the same scale is an indispensable tool in each discipline based on data processing. The variability of spheres dealing with aggregation (fusion) techniques is so rich that we frequently meet the same results and techniques under different names. Nowadays, when aggregation theory becomes a well established field of mathematics, it is the time to unify the notations and terminology. This is one of the aims of the monograph on aggregation [26] which is currently under preparation by Grabisch, Marichal, Mesiar and Pap. Therefore in this chapter we will use the notations and terminology from [26]. The main aim of this chapter is to bring a review of some recent results in aggregation. To achieve the readability, we will also recall some older results whenever necessary. Note that comprehensive state-of-art overviews on aggregation can be found in [21] (dated to 1985) and in [7] (dated to 2002). The chapter is organized as follows. In the next section, basic notions, notations and properties are given, including the classification of aggregation functions. In Sect. 3, averaging aggregation functions are discussed. Section 4 is devoted to conjunctive aggregation functions, and by duality, also to disjunctive aggregation functions. Weighted aggregation functions are discussed in Sect. 5. Finally, some conclusions are given.

Note that though aggregation can be discussed on an arbitrary scale $I$ (equipped with linear order), we restrict our considerations to the real intervals. Moreover, a major attention will be put to the case $I = [0, 1]$.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

## 2 Basic Notions, Notations and Properties

Unless otherwise stated, the letter $I$ will denote a subinterval of the extended real line, $I = [a, b] \subseteq [-\infty, \infty]$. Aggregation on $I$ for a fixed number $n$ of inputs always means a processing of input data by a special $n$–ary function defined on $I^n$. Similarly, aggregation on $I$ for an arbitrary (but fixed) finite number of inputs can be seen as a data processing by a system of such functions. One of the crucial problems in that case is the relationship of the functions from the system differing in the number of inputs.

Note that to shorten some expressions, we will write x instead of $(x_1, \ldots, x_n)$.

**Definition 1.** (i) *An **n–ary aggregation function** is a function $A^{(n)} : I^n \to I$ that is non–decreasing in each place and fulfills the following boundary conditions*

$$\inf_{\mathrm{x} \in I^n} A^{(n)}(\mathbf{x}) = \inf I \quad \textit{and} \quad \sup_{\mathrm{x} \in I^n} A^{(n)}(\mathbf{x}) = \sup I.$$

(ii) *An **extended aggregation function** is a function $A : \bigcup_{n \in \mathbb{N}} I^n \to I$ such that for all $n > 1$, $A^{(n)} = A_{|I^n}$ is an n–ary aggregation function and $A^{(1)}$ is the identity on $I$.*

We first recall several examples of extended aggregation functions on $I$:

- The *sum* $\Sigma$,

$$\Sigma(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i,$$

  in the case of an interval $I$ with the left–end point $-\infty$ or 0, the right–end point 0 or $\infty$, and with the convention $(-\infty) + \infty = -\infty$ if necessary.
- The *product* $\Pi$,

$$\Pi(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i,$$

  if $I$ is an interval with the left–end point 0 or 1, the right–end point 1 or $\infty$ and with the convention $0 \,.\, \infty = 0$ if necessary.
- The *arithmetic mean* $M$,

$$M(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

  on an arbitrary interval $I$, and if $I = [-\infty, \infty]$, the convention $(-\infty) + \infty = -\infty$ is adopted.

- The *geometric mean* $G$,

$$G(x_1, \ldots, x_n) = \left(\prod_{i=1}^{n} x_i\right)^{1/n},$$

where $I \subseteq [0, \infty]$, and $0 \,.\, \infty = 0$ by convention.
- The *minimum* $Min$,

$$Min(x_1, \ldots, x_n) = \min\{x_1, \ldots, x_n\} = \bigwedge_{i=1}^{n} x_i .$$

- The *maximum* $Max$,

$$Max(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\} = \bigvee_{i=1}^{n} x_i .$$

In all above mentioned extended aggregation functions there is some relationship between aggregation functions $A^{(n)}$ and $A^{(m)}$ for all $n,\ m \in \mathbb{N}$. This is not guaranteed by Definition 1, in general. Before discussing this problem in more details, we recall some basic properties of ($n$–ary/extended) aggregation functions. Unless otherwise specified, a property of a discussed extended aggregation function $A : \bigcup\limits_{n\in\mathbb{N}} I^n \to I$ means that each $n$–ary aggregation function $A^{(n)} : I^n \to I$ possesses the mentioned property. Therefore we will define the next properties for $n$–ary aggregation functions only.

**Definition 2.** *For a fixed $n \in \mathbb{N} \setminus \{1\}$, let $A^{(n)} : I^n \to I$ be an $n$–ary aggregation function on $I$. Then $A^{(n)}$ is called:*

(i) ***symmetric*** *(anonymous) if for each permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ and each $\mathrm{x} \in I^n$*

$$A^{(n)}(\mathrm{x}) = A^{(n)}(x_{\sigma(1)}, \ldots, x_{\sigma(n)})\,;$$

(ii) ***idempotent*** *(unanimous) if for each $c \in I$*

$$A^{(n)}(c, \ldots, c) = c\,;$$

(iii) ***strictly monotone*** *if for all $x_i,\ y_i \in I,\ i \in \{1, \ldots, n\}$ such that $x_i \le y_i$ and $(x_1, \ldots, x_n) \neq (y_1, \ldots, y_n)$ it follows that*

$$A^{(n)}(x_1, \ldots, x_n) < A^{(n)}(y_1, \ldots, y_n)\,;$$

(iv) ***continuous*** *if for each $\mathbf{x}_0 \in I^n$,*

$$\lim_{\mathrm{x}\to\mathrm{x}_0} A^{(n)}(\mathbf{x}) = A^{(n)}(\mathbf{x_0}),$$

*i.e., if $A^{(n)}$ is a continuous function of $n$ variables in the usual sense;*

(v) ***1–Lipschitz**, if for all* $(x_1, \ldots, x_n),\ (y_1, \ldots, y_n) \in I^n$,

$$|A^{(n)}(x_1, \ldots, x_n) - A^{(n)}(y_1, \ldots, y_n)| \le \sum_{i=1}^{n} |x_i - y_i|;$$

(vi) ***bisymmetric** if for all* $n \times n$ *matrices* $X = (x_{ij})$, *with entries* $x_{ij} \in I$ *for all* $i,\ j \in \{1, \ldots, n\}$,

$$A^{(n)}\left(A^{(n)}(x_{11}, \ldots, x_{1n}), \ldots, A^{(n)}(x_{n1}, \ldots, x_{nn})\right)$$
$$= A^{(n)}\left(A^{(n)}(x_{11}, \ldots, x_{n1}), \ldots, A^{(n)}(x_{1n}, \ldots, x_{nn})\right).$$

We can equivalently say that, for example, an $n$–ary aggregation function $A^{(n)}$ is symmetric if and only if for all $\mathrm{x} \in I^n$ it holds

$$A^{(n)}(\mathrm{x}) = A^{(n)}(x_2, x_1, x_3, \ldots, x_n) = A^{(n)}(x_2, \ldots, x_n, x_1).$$

Similarly, the idempotency of $A^{(n)}$ is equivalent to the property

$$Min^{(n)} \le A^{(n)} \le Max^{(n)}.$$

**Definition 3.** *For a fixed* $n \in \mathbb{N} \setminus \{1\}$, *let* $A^{(n)} : I^n \to I$ *be an n–ary aggregation function on* $I$.

(i) *An element* $e \in I$ *is called* ***neutral element*** *of* $A^{(n)}$ *if for each* $i \in \{1, \ldots, n\}$ *and each* $x_i \in I$ *it holds that*

$$A^{(n)}(e, \ldots, e, x_i, e, \ldots, e) = x_i\ .$$

(ii) *An element* $a \in I$ *is called* ***annihilator*** *of* $A^{(n)}$ *if for all* $(x_1, \ldots, x_n) \in I^n$ *it holds that*
*if* $x_i = a$ *for some* $i \in \{1, \ldots, n\}$ *then* $A^{(n)}(x_1, \ldots, x_n) = a$.

For extended aggregation functions we can also introduce stronger versions of idempotency, neutral element and bisymmetry.

**Definition 4.** *Let* $A : \bigcup_{n\in\mathbb{N}} I^n \to I$ *be an extended aggregation function. Then*

(i) *A is* ***strongly idempotent*** *whenever*

$$A(\underbrace{\mathbf{x}, \ldots, \mathbf{x}}_{k-times}) = A(\mathrm{x})$$

*for all* $k \in \mathbb{N}$ *and* $\mathrm{x} \in \bigcup_{n\in\mathbb{N}} I^n$.

(ii) *An element $e \in I$ is said to be a* ***strong neutral element*** *of A if for each $n \in \mathbb{N}$, each $\mathrm{x} \in \bigcup_{n\in\mathbb{N}} I^n$ and $i \in \{1, \dots, n+1\}$ it holds*

$$A(\mathrm{x}) = A(x_1, \dots, x_{i-1}, e, x_i, \dots, x_n).$$

(iii) *A is* ***strongly bisymmetric*** *if for any $n \times m$ matrix $X = (x_{ij})$ with all entries $x_{ij} \in I$, it holds*

$$A^{(n)}(A^{(m)}(\mathbf{x}_{1.}), \dots, A^{(m)}(\mathbf{x}_{n.})) = A^{(m)}(A^{(n)}(\mathrm{x}_{.1}), \dots, A^{(n)}(\mathrm{x}_{.m})),$$

*where for all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$,*

$$\mathbf{x}_{i.} = (x_{i1}, \dots, x_{im}) \quad \text{and} \quad \mathbf{x}_{.j} = (x_{1j}, \dots, x_{nj}).$$

Classical properties linking different input arities of extended aggregation functions are:

- ***associativity***, that is, for each $n,\ m \in \mathbb{N}$, $\mathrm{x} \in I^n$, $\mathrm{y} \in I^m$

$$A^{(n+m)}(\mathbf{x}, \mathbf{y}) = A^{(2)}(A^{(n)}(\mathbf{x}), A^{(m)}(\mathbf{y}))\,;$$

- ***decomposability***, that is, for all integers $0 \le k \le n$, $n \in \mathbb{N}$, and all $\mathrm{x} \in I^n$

$$\begin{aligned} &A^{(n)}(x_1, \dots, x_k, x_{k+1}, \dots, x_n) \\ &= A^{(n)}(\underbrace{A^{(k)}(x_1, \dots, x_k)}_{k-times}, \underbrace{A^{(n-k)}(x_{k+1}, \dots, x_n)}_{(n-k)-times}). \end{aligned}$$

The associativity of an extended aggregation function $A$ is equivalent to the standard associativity of the corresponding binary aggregation function $A^{(2)}$,

$$A^{(2)}(x, A^{(2)}(y, z)) = A^{(2)}(A^{(2)}(x, y), z)$$

for all $x,\ y,\ z \in I$, and $A^{(n)}$ for $n > 2$, being the genuine $n$–ary extension of $A^{(2)}$ given by

$$A^{(n)}(x_1, \dots, x_n) = A^{(2)}\left(A^{(n-1)}(x_1, \dots, x_{n-1}), x_n\right)$$

defined by induction. Evidently, using this way, any binary aggregation function $A^{(2)}$ can be extended to an extended aggregation function $\overline{A^{(2)}} = A$. More generally, a huge class of extended aggregation functions can be constructed from a system $\mathcal{A} = (A_n^{(2)})_{n\in\mathbb{N}}$ of binary aggregation functions by induction. We define $A_{\mathcal{A}} = A$ as follows:

$$\begin{aligned} &A^{(1)}(x_1) = x_1, \\ &A^{(2)}(x_1, x_2) = A_1^{(2)}(x_1, x_2), \end{aligned}$$

$$\vdots$$
$$A^{(n)}(x_1, \ldots, x_n) = A^{(2)}_{n-1}\left(A^{(n-1)}(x_1, \ldots, x_{n-1}), x_n\right)$$
$$\vdots$$

Extended aggregation functions $A_{\mathcal{A}} = A$ were called recursive by Montero, see e.g. [36], compare also [18]. Evidently, each associative extended aggregation function is recursive but not vice–versa.

From the examples introduced above, the sum $\Sigma$ is symmetric, associative and bisymmetric. If $0 \in I$, then 0 is the strong neutral element of $\Sigma$, if $-\infty \in I$ then this element is the annihilator of $\Sigma$, and if $+\infty \in I$ and $-\infty \notin I$ then $+\infty$ is the annihilator of $\Sigma$. The extended aggregation function $\Sigma$ is 1–Lipschitz and strictly monotone if $I \subset \mathbb{R}$, continuous if $I \neq [-\infty, \infty]$.

The arithmetic mean $M$ is recursive, symmetric, strongly idempotent and bisymmetric on any interval $I$. It is 1–Lipschitz and strictly monotone if $I \subset \mathbb{R}$ and continuous if $I \neq [-\infty, \infty]$. It has an annihilator $a$ only if $I$ is an unbounded interval, namely, $a = -\infty$ if $-\infty \in I$; $a = \infty$, if $\infty \in I$ and $-\infty \notin I$.

Let the extended aggregation function $A : \bigcup_{n\in\mathbb{N}} I^n \to I$ be given by

$$A(x_1, \ldots, x_n) = \min\left(x_1, \prod_{i=2}^{n} x_i\right)$$

whenever $n > 1$. Evidently, $e = 1$ is the neutral element of $A$, but it is not a strong neutral element. Indeed, if we take $(x_1, x_2) = (0.5, 0.5)$ then, for $i = 1$ we have $A(1, x_1, x_2) = 0.25$, for $i = 2$ and $i = 3$ we have $A(x_1, 1, x_2) = A(x_1, x_2, 1) = 0.5$. Observe that $A$ is a quasi–copula, see Sect. 4, i.e., $A$ is 1-Lipschitz.

To simplify notation, if no confusion can arise, $n$–ary aggregation functions $A^{(n)}$ will simply be denoted by $A$ without stressing their arity.

The basic classification of aggregation functions takes into account the main fields of applications. Following Dubois and Prade [22], we will distinguish four classes of ($n$–ary/extended) aggregation functions:

- *conjunctive aggregation functions*: aggregation functions $A \leq Min$;
- *averaging aggregation functions*: aggregation functions $A$, $Min \leq A \leq Max$, or, equivalently, idempotent aggregation functions;
- *disjunctive aggregation functions*: aggregation functions $A \geq Max$;
- *mixed aggregation functions*: aggregation functions which do not belong to any of other three classes.

Observe that the interval $I$ may be crucial for the classification of a discussed aggregation function. For example, the product $\Pi$ is a conjunctive aggregation function on $[0, 1]$, disjunctive on $[1, \infty]$ and mixed on $[0, \infty]$.

For any decreasing one–to–one mapping $\varphi : I \to I$, $A : \bigcup_{n\in\mathbb{N}} I^n \to I$ is a conjunctive (disjunctive) extended aggregation function if and only if the function $A_\varphi : \bigcup_{n\in\mathbb{N}} I^n \to I$ given by

$$A_{\varphi}(x_1, \dots, x_n) = \varphi^{-1}\left(A(\varphi(x_1), \dots, \varphi(x_n)\right)$$

is a disjunctive (conjunctive) extended aggregation function. This duality allows to investigate, construct and discuss conjunctive aggregation functions only, and to transfer all the results by this duality to the disjunctive aggregation functions.

## 3 Averaging Aggregation Functions

We first recall the basic averaging aggregation functions, for more details we recommend [7]:

- The *arithmetic mean* $M$,

$$M(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i \, .$$

- *Quasi–arithmetic means* $M_f$, where $f : I \to [-\infty, \infty]$ is a continuous strictly monotone function and

$$M_f(x_1, \dots, x_n) = f^{-1}(M(f(x_1), \dots, f(x_n)),$$

  as, for example, the geometric, harmonic and quadratic means.

- *Weighted arithmetic means* $M_{\mathbf{w}}$, where $\mathbf{w} = (w_1, \dots, w_n)$, $w_i \geq 0$, $\sum\limits_{i=1}^{n} w_i = 1$ and

$$M_{\mathbf{w}}(x_1, \dots, x_n) = \sum_{i=1}^{n} w_i \, x_i,$$

  see also Sect. 5.
- *Weighted quasi–arithmetic means* $M_{f,\mathbf{w}}$,

$$M_{f,\mathbf{w}}(x_1, \dots, x_n) = f^{-1}\left(\sum_{i=1}^{n} w_i \, f(x_i)\right).$$

- *OWA* (*ordered weighted average*) *operator* $M'_{\mathbf{w}}$,

$$M'_{\mathbf{w}}(x_1, \dots, x_n) = M_{\mathbf{w}}(x'_1, \dots, x'_n) = \sum_{i=1}^{n} w_i \, x'_i,$$

  where $x'_i$ is the $i$–th order statistics from the sample $(x_1, \dots, x_n)$.

- *OWQA* (*ordered weighted quasi–arithmetic*) *operator* $M'_{f,\mathbf{w}}$,

$$M'_{f,\mathbf{w}}(x_1,\ldots,x_n) = M_{f,\mathbf{w}}(x'_1,\ldots,x'_n) = f^{-1}\left(\sum_{i=1}^{n} w_i\, f(x'_i)\right).$$

- *Idempotent uninorms*, [15].
- *Idempotent nullnorms*, i.e., $a$–medians, given for a fixed $a \in I$ by

$$Med_a(x_1,\ldots,x_n) = \mathrm{med}(x_1, a, x_2, a, x_3, a, \ldots, a, x_n).$$

- *Fuzzy integrals*, [29, 47].

Recall that for any 2–copula $C : [0,1]^2 \to [0,1]$ (for the definition of a copula see the next section) and for any fuzzy measure $m : \mathcal{P}(\{1,\ldots,n\}) \to [0,1]$, i.e., a non–decreasing set function such that $m(\emptyset) = 0$ and $m(\{1,\ldots,n\}) = 1$, we can define a fuzzy integral $F_{C,m} : [0,1]^n \to [0,1]$ by

$$F_{C,m}(x_1,\ldots,x_n) = \sum_{i=1}^{n} \left(C\left(x'_i, m(\{j \mid x_j \geq x'_i\})\right) - C\left(x'_{i-1}, m(\{j \mid x_j \geq x'_i\})\right)\right),$$

with the convention $x'_0 = 0$, where $x'_i$ is the $i$–th order statistics from the sample $(x_1,\ldots,x_n)$. Then $F_{\Pi,m}$ is the Choquet integral [12, 19, 38] and $F_{Min,m}$ is the Sugeno integral [43, 38]. Also observe that if $m$ is additive then $F_{\Pi,m} = M_{\mathbf{w}}$ is the weighted arithmetic mean with the weights given by $w_i = m(\{i\})$. Similarly, if $m$ is symmetric, i.e., $m(A) = h\left(\frac{card A}{n}\right)$ for some increasing function $h : [0,1] \to [0,1]$, then $F_{\Pi,m}$ is the OWA operator $M'_{\mathbf{w}}$ with the weights $w_i = h\left(\frac{i}{n}\right) - h\left(\frac{i-1}{n}\right)$.

Note that averaging aggregation functions are closed under composition, i.e., for any averaging (extended) aggregation functions $A, A_1, \ldots, A_n$ on $I$, also the function $D = A(A_1,\ldots,A_n) : \bigcup_{n\in\mathbb{N}} I^n \to I$, given by $D(\mathrm{x}) = A(A_1(\mathrm{x}),\ldots,A_n(\mathrm{x}))$, is an averaging extended aggregation function.

An interesting class of averaging aggregation functions are the *internal* aggregation functions characterized by $A(x_1,\ldots,x_n) \in \{x_1,\ldots,x_n\}$. Continuous internal aggregation functions are exactly lattice polynomials, whose prescription formula contains inputs $x_1,\ldots,x_n$, symbols for join $\vee$ and meet $\wedge$, i.e., $Max$ and $Min$ in infix form, and parentheses. Independently of the interval $I$, they have the same formula, and on any open interval $I$ they are the only aggregation functions invariant under any increasing $I \to I$ one–to–one transformation $\varphi$. On $[0,1]$, they are in a one–to–one correspondence with $\{0,1\}$– valued fuzzy measures (and then we can apply any fuzzy integral based on a copula $C$, e.g., the Choquet or Sugeno integrals). As an example we give all 18 ternary aggregation functions which are internal and continuous on any interval $I$:

$A^{(3)}(x_1, x_2, x_3) =$

$$\begin{array}{lll}
x_1; & x_2; & x_3; \\
x_1 \wedge x_2; & x_1 \wedge x_3; & x_2 \wedge x_3; \\
x_1 \vee x_2; & x_1 \vee x_3; & x_2 \vee x_3; \\
x_1 \wedge (x_2 \vee x_3); & x_2 \wedge (x_1 \vee x_3); & x_x \wedge (x_1 \vee x_2); \\
x_1 \vee (x_2 \wedge x_3); & x_2 \vee (x_1 \wedge x_3); & x_3 \vee (x_1 \wedge x_2); \\
x_1 \wedge x_2 \wedge x_3 = x_1'; & (x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_2 \wedge x_3) = x_2'; & x_1 \vee x_2 \vee x_3 = x_3'.
\end{array}$$

Another interesting and still not completely described family of averaging extended aggregation functions are the mixture operators $M^g : \bigcup_{n \in \mathbb{N}} I^n \to I$ given by

$$M^g(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} g(x_i)\, x_i}{\sum_{i=1}^{n} g(x_i)},$$

where $g : I \to [0, \infty[$ is a given weighting function [33, 46]. Evidently, mixture operators are idempotent and they generalize the arithmetic mean $M$, since $M = M^g$ for any constant weighting function $g$. Mixture operators are extended aggregation functions if and only if they are monotone, which is not a general case. For example, let $I = [0, b]$ and let $g : I \to ]0, \infty[$ be given by $g(x) = x + 1$. Then $M^g$ is an averaging extended aggregation function only if $b \in ]0, 1]$. Till now, only some sufficient conditions ensuring the monotonicity of mixture operators $M^g$ are known, as, for example, for a non–decreasing differentiable function $g$ the next two conditions:

(i) $g(x) \geq g'(x)\, l(I)$ for all $x \in I$, where $l(I)$ is the length of the interval $I$;
(ii) $g(x) \geq g'(x)\, (x - \inf I)$ for all $x \in I$.

Also other generalizations of mixture operators are interesting, as, for example, the *quasi–mixture operators* $M_f^g$, defined by

$$M_f^g(x_1, \ldots, x_n) = f^{-1}\left(\frac{\sum_{i=1}^{n} g(x_i)\, f(x_i)}{\sum_{i=1}^{n} g(x_i)}\right),$$

*generalized mixture operators* $M^{\mathbf{g}}$, where $\mathbf{g} = (g_1, \ldots, g_n)$ is a vector of weighting functions, defined by

$$M^{\mathbf{g}}(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} g_i(x_i)\, x_i}{\sum_{i=1}^{n} g_i(x_i)},$$

and *ordered generalized mixture operators* $M'^{\mathbf{g}}$,

$$M'^{\mathbf{g}}(x_1, \ldots, x_n) = M^{\mathbf{g}}(x'_1, \ldots, x'_n).$$

These operators can be seen as generalizations of the quasi–arithmetic means, weighted arithmetic means and OWA operators, respectively. In general, the monotonicity of such operators is not still clarified. Observe that these types of operators were already studied by Bajraktarevič, [3], especially the conditions under which two different couples $(f_1, g_1)$ and $(f_2, g_2)$ induce the same quasi–mixture operators, i.e., when $M_{f_1}^{g_1} = M_{f_2}^{g_2}$.

An interesting composition method of aggregation functions was recently proposed in [11]. For any extended aggregation functions $A$, $B$ and a binary aggregation function $C$ on $I$, we define $D = A_{B,C} : \bigcup_{n\in\mathbb{N}} I^n \to I$ by

$$D(x_1, \ldots, x_n) = A(C(x_1, B(x_1, \ldots, x_n)), \ldots, C(x_n, B(x_1, \ldots, x_n))).$$

Evidently, if all $A$, $B$, $C$ are idempotent then $D$ is also idempotent. As a special case of this method, consider $C = Min^{(2)}$, $A = F_{\Pi,m_1}$, i.e., the Choquet integral with respect to a fuzzy measure $m_1$ on $\{1, \ldots, n\}$, and $B = F_{Min,m_2}$, i.e., the Sugeno integral with respect to a fuzzy measure $m_2$ on $\{1, \ldots, n\}$. Then $D = A_{B,C}$ is the two–fold integral introduced by Narukawa and Torra in [48]. Observe that for $m_1$ equal to the strongest fuzzy measure $m^*$ given by

$$m^*(E) = \begin{cases} 0 & \text{if } E = \emptyset, \\ 1 & \text{otherwise,} \end{cases}$$

we get $A_{B,C} = F_{Min,m_2}$. Similarly, if $m_2 = m^*$, then $A_{B,C} = F_{\Pi,m_1}$. Thus the two–fold integral is an averaging aggregation function generalizing both the Choquet and Sugeno integrals.

## 4 Conjunctive Aggregation Functions

In this section we restrict our considerations to the interval $I = [0, 1]$ only. As the conjunctive aggregation functions are bounded from above by $Min$, the weakest extended aggregation function $A_w : \bigcup_{n\in\mathbb{N}} [0, 1]^n \to [0, 1]$ given by

$$A_w(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \prod_{i=1}^{n} x_i = 1, \\ 0 & \text{otherwise,} \end{cases}$$

is also the weakest conjunctive extended aggregation function, and, obviously, $Min$ is the strongest one. Evidently, $a = 0$ is the annihilator of any conjunctive aggregation function $A$. Depending on the field of applications, often some kind of neutrality for the element $e = 1$ is required [4, 14].

**Definition 5.** *An (n–ary) aggregation function A on* $[0, 1]$ *is called a (an n–)* ***semicopula*** *whenever* $e = 1$ *is its neutral element. An extended aggregation function A on* $[0, 1]$ *with the strong neutral element* $e = 1$ *is called a* ***conjunctor***.

Recall some distinguished classes of conjunctive aggregation functions.

- *Triangular norms* (t-norms for short) [41, 28] are associative symmetric conjunctors.
- *Quasi–copulas* [2, 27] are 1–Lipschitz conjunctive aggregation functions. Observe that each quasi–copula is necessarily a semicopula.
- *Copulas* [42, 37] are $n$–increasing semicopulas, where the $n$–increasing property means the non–negativity of all mixed $n$–th differences. For $n = 2$ this means that $A : [0, 1]^2 \to [0, 1]$ is 2–increasing if and only if for all $(x_1, x_2), (y_1, y_2) \in [0, 1]^2$ such that $x_1 \leq y_1$ and $x_2 \leq y_2$, it holds

$$(A(y_1, y_2) - A(y_1, x_2)) - (A(x_1, y_2) - A(x_1, x_2)) \geq 0.$$

Each copula is 1–Lipschitz, and thus a quasi–copula.

Observe that conjunctive aggregation functions, semicopulas, conjunctors, quasi–copulas and copulas are convex classes, which is not the case of triangular norms. Because of the existence of exhaustive monographs on t-norms [28] and copulas [37] we will not discuss these classes in detail. However, there are some new interesting results worth mentioning.

Recall that each 1–Lipschitz t-norm is an associative copula (as a binary function) and vice–versa. Thus each associative copula is an ordinal sum [28] of Archimedean 1–Lipschitz t-norms. These later are characterized by the convexity of their additive generator, that is, a strictly decreasing continuous function $t : [0, 1] \to [0, \infty]$, $t(1) = 0$.

A related problem concerning $k$–Lipschitz Archimedean t-norms, $k > 1$, was stated as an open problem in [1]. Mesiarová has recently characterized [35] $k$–Lipschitz Archimedean $t$-norms by the $k$–convexity of their additive generators. The $k$–convexity of an additive generator $t$ means that for all $0 < x < y < 1$ and $\varepsilon \in ]0, \min(1 - y, 1 - k\,x)]$ it holds

$$t(x + k\,\varepsilon) - t(x) \leq t(y + \varepsilon) - t(y).$$

Evidently, the 1–convexity reduces to the standard convexity.

The weakest 1–Lipschitz conjunctor is the Łukasiewicz t-norm $T_L$, in the framework of copulas also called the lower Fréchet–Hoeffding bound, which in the binary form is given by $T_L(x, y) = \max(x + y - 1, 0)$, $(x, y) \in [0, 1]^2$.

Note that the class of all $k$–Lipschitz t-norms for $k > 1$ has no weakest element, though there are several minimal $k$–Lipschitz t-norms. The weakest $k$–Lipschitz conjunctor in the binary form is given by $C_k(x, y) = \max(x + k\,y - k, k\,x + y - k, 0)$.

For each ternary conjunctive aggregation function $C : [0, 1]^3 \to [0, 1]$ we can introduce three binary functions $C_{12}, C_{23}, C_{13} : [0, 1]^2 \to [0, 1]$ given by

$$C_{12}(x, y) = C(x, y, 1), \quad C_{23}(x, y) = C(1, x, y), \quad C_{13}(x, y) = C(x, 1, y).$$

All functions $C_{12}, C_{23}, C_{13}$ are conjunctive. Evidently, if $C$ is the ternary form of some conjunctor, then $C_{12} = C_{23} = C_{13}$. In general these equalities fail even for semicopulas (quasi–copulas, copulas). An interesting problem is also the reverse compatibility problem, namely, under which conditions binary functions $A, B, D$ of some type are the marginal functions of a ternary conjunctive aggregation function $C$ of the same type. In the case of t-norms it is evident that $A = B = D$ are necessarily the binary forms and $C$ is the ternary form of the same t-norm. In the case of semicopulas (quasi–copulas), for any $A, B, D$ there is a ternary semicopula (quasi–copula) $C$, not necessarily unique, such that $C_{12} = A$, $C_{23} = B$, $C_{13} = D$, for example, $C : [0, 1]^3 \to [0, 1]$ given by

$$C(x, y, z) = Min(A(x, y), B(y, z), D(x, z)) . \tag{1}$$

However, for 2–copulas $A, B, D$ the ternary operation $C$ given by (1) need not be a copula, in general. This is, e.g., in the case $A = B = D = T_L$, when $C$ is a 3–quasi–copula but not a 3–copula. For any 2–copulas $A, B$, let $A * B = D : [0, 1]^2 \to [0, 1]$ be given by

$$D(x, y) = \int_0^1 \frac{\partial A(x, t)}{\partial t} \frac{\partial B(t, y)}{\partial t} \, dt .$$

Then $D$ is also a 2–copula [13], and $C : [0, 1]^3 \to [0, 1]$ given by

$$C(x, y, z) = \int_0^y \frac{\partial A(x, t)}{\partial t} \frac{\partial B(t, z)}{\partial t} \, dt$$

is a 3–copula and moreover, $C_{12} = A$, $C_{23} = B$, $C_{13} = D$, compare also [31]. For example, if $A = B = T_L$ then $A * B = Min^{(2)}$ and $(T_L, T_L, D)$ are marginal 2–copulas of a 3–copula $C : [0, 1]^2 \to [0, 1]$ if and only if $D = T_L * T_L = Min^{(2)}$ and $C(x, y, z) = \max(\min(x, z) + y - 1, 0)$.

Let $\mathcal{C} = \left(C_n^{(2)}\right)_{n \in \mathbb{N}}$ be a system of binary conjunctive aggregation functions. Then the recursive extended aggregation function $C = C_{\mathcal{C}}$,

$$\begin{aligned} C^{(n)}(x_1, \ldots, x_n) &= C_{n-1}^{(2)} \left(C^{(n-1)}(x_1, \ldots, x_{n-1}), x_n\right) \\ &= C_{n-1}^{(2)} \left(\ldots C_2^{(2)} \left(C_1^{(2)}(x_1, x_2), x_3\right), \ldots, x_n\right) \end{aligned}$$

is conjunctive. If all $C_n^{(2)}$, $n \in \mathbb{N}$, are semicopulas (quasi–copulas) then $C$ is an extended semicopula (quasi–copula). In the case of copulas, it is an open problem under which conditions $C^{(n)}$ is a copula and $C$ is an extended copula. In the case when $C = \overline{C_1^{(2)}}$, i.e., $C_n^{(2)} = C_1^{(2)}$ for all $n \in \mathbb{N}$, and $C_1^{(2)} : [0,1]^2 \to [0,1]$ is an Archimedean 2–copula, then $C$ is an extended copula, that is, an $n$–copula for each $n \in \mathbb{N}$, if and only if $C$ is generated by a decreasing bijection $t : [0,1] \to [0,\infty]$ whose inverse $t^{-1} : [0,\infty] \to [0,1]$ is totally monotone, that is, whose all derivatives at each point from $]0,\infty[$ exist and are non–negative [37]. Each such copula is necessarily bounded by the product, $C \geq \Pi$, which is an important example of an extended copula, reflecting the independence of random variables. To see a negative example, let $C_1^{(2)} = \Pi^{(2)}$ and $C_n^{(2)} = Min^{(2)}$ for all $n > 1$. Then $C$ is an extended quasi–copula but not an extended copula. Also note that not each extended copula is recursive. For example, the extended aggregation function $C : \bigcup_{n \in \mathbb{N}} [0,1]^n \to [0,1]$ given by

$$C(x_1, \ldots, x_n) = x_1\, Min(x_2, \ldots, x_n), \quad n \geq 2,$$

is an extended copula which is not recursive.

Finally, we introduce a useful proposition.

**Proposition 1.** *Let $\mathcal{F}$ be a class of (n–ary/extended) aggregation functions on* $[0,1]$ *and let $\mathcal{H}_{\mathcal{F}}$ be the set of all binary aggregation functions $D : [0,1]^2 \to [0,1]$ such that for all $A, B \in \mathcal{F}$ also $C = D(A,B)$ given by $C(\mathrm{x}) = D(A(\mathrm{x}), B(\mathrm{x}))$, is an element of $\mathcal{F}$. Then*

(i) *For the class $\mathcal{F} = \mathcal{A}$ of all (n–ary/extended) aggregation functions on* $[0,1]$ *it holds $\mathcal{H}_{\mathcal{A}} = \mathcal{A}^{(2)}$, that is, $D$ can be an arbitrary binary aggregation function.*

(ii) *For the class $\mathcal{F} = \mathcal{B}$ of all conjunctive (n–ary/extended) aggregation functions we have $\mathcal{H}_{\mathcal{B}} = \{D \in \mathcal{A}^{(2)} \mid D \leq Max^{(2)}\}$.*

(iii) *For the class $\mathcal{F} = \mathcal{S}$ of all (n–ary/extended) semicopulas we have*

$$\mathcal{H}_{\mathcal{S}} = \{D \in \mathcal{A}^{(2)} \mid D \text{ is idempotent}\}.$$

(iv) *For the class $\mathcal{F} = \mathcal{Q}$ of all quasi–copulas we have*

$$\mathcal{H}_{\mathcal{Q}} = \{D \in \mathcal{A}^{(2)} \mid \|D\|_\infty = 1\}.$$

*Note that the Chebyshev norm of a binary aggregation function $D$ is given by*

$$\|D\|_\infty = \sup \left( \frac{|D(x,y) - D(u,v)|}{\max(|x-u|, |y-v|)} \right),$$

*where the supremum is taken over all $(x,y)$, $(u,v) \in [0,1]^2$, $(x,y) \neq (u,v)$.*

(v) *For the class $\mathcal{F} = \mathcal{C}$ of all copulas we have*

$$\mathcal{H}_{\mathcal{C}} = \{D \in \mathcal{A}^{(2)} \mid D \textit{ is a weighted mean}\}.$$

(vi) *For the class $\mathcal{F} = \mathcal{T}$ of all t-norms we have*

$$\mathcal{H}_{\mathcal{T}} = \{P_F,\ P_L\},$$

*where $P_F(x, y) = x$ and $P_L(x, y) = y$ for all $(x, y) \in [0, 1]^2$.*
*Evidently,*

$$\mathcal{H}_{\mathcal{T}} \subset \mathcal{H}_{\mathcal{C}} \subset \mathcal{H}_{\mathcal{Q}} \subset \mathcal{H}_{\mathcal{S}} \subset \mathcal{H}_{\mathcal{B}} \subset \mathcal{H}_{\mathcal{A}}.$$

By duality, similar notions can be introduced and similar results can be obtained for disjunctive aggregation functions. For an ($n$-ary/extended) aggregation function $A$ on $I = [0, 1]$, the standard duality, here called simply duality, is related to the order reversing bijection $n : [0, 1] \to [0, 1]$, $n(x) = 1 - x$, the so–called standard negation on [0, 1]. Then an ($n$-ary/extended) aggregation function $A^d$ on [0, 1] is called the dual of $A$, if for all x it holds $A^d(\mathrm{x}) = 1 - A(1 - \mathrm{x})$.

For example, duals of t–norms are t–conorms, that is, associative symmetric aggregation functions with 0 as the strong neutral element. For binary 1–Lipschitz aggregation functions another type of duality was introduced, see [32], compare also [45]. For an aggregation function $A : [0, 1]^2 \to [0, 1]$ its reverse $A^* : [0, 1]^2 \to [0, 1]$ is given by $A^*(x, y) = x + y - A(x, y)$. Evidently $(A^*)^* = A$. An interesting problem is under which conditions $A^* = A^d$, i.e., for which binary 1–Lipschitz aggregation functions it holds

$$A(x, y) = x + y - 1 + A(1 - x, 1 - y) \quad \text{for all } (x, y) \in [0, 1]^2. \tag{2}$$

Restricting our considerations to the associative aggregation functions we end up with the famous Frank functional equation [25] and the only solutions to (2) are Frank's t–norms and the symmetric ordinal sums of Frank's t–norms, see [30].

## 5 Weighted Aggregation Functions

This section is devoted to a proposal how to introduce weights (importances) into aggregation. For an input vector $\mathbf{x} = (x_1, \ldots, x_n)$, the corresponding weights $w_1, \ldots, w_n$ can be understood as cardinalities of single inputs $x_1, \ldots, x_n$, respectively. We will deal with weighting vectors $\mathbf{w} = (w_1, \ldots, w_n)$, $w_i \in [0, \infty[\,, i \in \{1, \ldots, n\}$, and $\sum_{i=1}^{n} w_i > 0$. If $\sum_{i=1}^{n} w_i = 1$, $\mathbf{w}$ will be called a normal weighting vector.

For an extended aggregation function $A : \bigcup_{n\in\mathbb{N}} I^n \to I$, and a weighting vector $\mathbf{w} = (w_1, \ldots, w_n)$ (for some $n \in \mathbb{N}$), we will discuss an $n$–ary aggregation function

$A_{\mathbf{w}} : I^n \to I$, which will be called a *weighted aggregation function*. We expect the next quite natural properties of weighted aggregation functions, compare also [5].

(W1) If $\mathbf{w} = (1, \ldots, 1) = \mathbf{1}$ then

$$A_{\mathbf{1}}(x_1, \ldots, x_n) = A(x_1, \ldots, x_n)$$

for all $(x_1, \ldots, x_n) \in I^n$.

(W2) For any $(x_1, \ldots, x_n) \in I^n$ and any $\mathbf{w} = (w_1, \ldots, w_n)$,

$$A_{\mathrm{w}}(x_1, \ldots, x_n) = A_{\mathrm{w}^*}(x_{m_1}, \ldots, x_{m_k}),$$

where $\{m_1, \ldots, m_k\} = \{i \in \{1, \ldots, n\} \mid w_i > 0\}$, $m_1 < \ldots < m_k$, $\mathbf{w}^* = (w_{m_1}, \ldots, w_{m_k})$.

(W3) If $\mathbf{w}$ is a normal weighting vector then $A_{\mathbf{w}}$ is an idempotent aggregation function.

Observe that (W1) simply embeds the aggregation function $A$ into weighted aggregation functions. Further, due to (W2), a zero weight $w_i$ in a weighting vector $\mathbf{w}$ means that we can omit the corresponding score $x_i$ (and the weight $w_i = 0$) from aggregation. Finally, the property (W3) expresses the standard boundary condition for extended aggregation functions, namely, that the aggregation of a unique input $x$ results in $x$, $A(x) = x$. Then $A_{\mathbf{w}}(x_1, \ldots, x_n)$ with $\sum_{i=1}^{n} w_i = 1$ can be seen as the aggregation of $x$ with cardinality $\sum_{i=1}^{n} w_i = 1$, i.e., $A_{\mathbf{w}}(x, \ldots, x) = A(x) = x$, which is exactly the idempotency of the function $A_{\mathbf{w}}$.

The standard summation on $[0, +\infty]$ can be understood as a typical aggregation on $[0, +\infty]$. For a given weighting vector $\mathbf{w} = (w_1, \ldots, w_n)$, the weighted sum $\sum_{i=1}^{n} w_i\, x_i$ is simply the sum of inputs $x_i$ transformed by means of weights $w_i$ into new inputs $y_i = w_i\, x_i$. Note that the common multiplication of reals applied in the next transformation can be straightforwardly deduced from the original summation (and the standard order of real numbers), i.e., for $w \geq 0$, $x \in [0, +\infty]$

$$w \cdot x = \sup \left( y \in [0, +\infty] \mid \exists\, i, j \in \mathbb{N},\ \frac{i}{j} < w \text{ and } u \in [0, +\infty] \text{ such that } \sum_{k=1}^{j} u < x \text{ and } y = \sum_{k=1}^{i} u \right)$$

Recall that the weighted sum $\sum_{i=1}^{n} w_i\, x_i$ for weights $w_i$ such that $\sum_{i=1}^{n} w_i = 1$ is just the weighted arithmetic mean. The above discussed approach can be applied to any continuous symmetric associative aggregation function defined on $I = [0, c]$ with neutral element 0, as, for example, to any continuous t–conorm $S$. The weighted t–conorm $S_{\mathbf{w}} : [0, 1]^n \to [0, 1]$, where $n = \dim \mathbf{w}$, is simply defined as

$$S_{\mathbf{w}}(x_1, \dots, x_n) = S(w_1 \odot x_1, \dots, w_n \odot x_n) \tag{3}$$

where the transformed input data $w_i \odot x_i$ are obtained from the weights $w_i$ and the original inputs $x_i$ by means of a binary operation $\odot : [0, +\infty[ \times [0, 1] \to [0, 1]$,

$$w \odot x = \sup\,(y \in [0, 1] \mid \exists\, i, j \in \mathbb{N},\ \tfrac{i}{j} < w \text{ and } u \in [0, 1] \text{ such that } S(\underbrace{u, \dots, u}_{j-times}) < x \text{ and } y = S(\underbrace{u, \dots, u}_{i-times})).$$

Evidently, (4) is an appropriate modification of (3). Note that $0 \odot x = 0$ and $1 \odot x = x$ for all $x \in [0, 1]$. In the case when $S$ has unit multipliers, i.e., $S(x, y) = 1$ for some $x, y \in [0, 1[$ we should require $\sum_{i=1}^{n} w_i \geq 1$ to keep the boundary condition $S_{\mathbf{w}}(1, \dots, 1) = 1$. Obviously, the weighted t–conorm $S_{\mathbf{w}}$ for any continuous t–conorm $S$ fulfills axioms (W1), (W2), (W3). More details about weighted t–conorms can be found in [6], including several examples. Recall some facts:

- $Max_{\mathbf{w}}(x_1, \dots, x_n) = \max(x_i \mid w_i > 0)$, (due to $w \odot x = x$ if $w > 0$);
- $S_{\mathbf{w}}$ is lower semi–continuous (left continuous);
- $S_{\mathbf{w}}$ (with some nontrivial $w_i \notin \{0, 1\}$) is continuous if and only if either $S = Max$ or $S$ is a continuous Archimedean t–conorm;
- If $S$ is continuous Archimedean t–conorm with an additive generator $g : [0, 1] \to [0, +\infty]$, and $\mathbf{w}$ is a normal weighting vector, then $S_{\mathbf{w}}(x_1, \dots, x_n) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_i)\right)$, i.e., $S_{\mathbf{w}}$ is a weighted quasi–arithmetic mean (because $w \odot x = g^{-1}(w \cdot g(x))$ for $w \in [0, 1]$). It is either cancelative (if $S$ is a nilpotent t–conorm; e.g., the Yager t–conorm for $p = 2$, see [28], leads to the weighted quadratic mean) or it has annihilator $a = 1$ (if $S$ is a strict t–conorm).

Dual operators to t–conorms are t–norms [28]. Weighted t–norms can be defined in the spirit of (3) and (4), or, equivalently, by the duality, i.e.,

$$T_{\mathbf{w}}(x_1, \dots, x_n) = 1 - S_{\mathbf{w}}(1 - x_1, \dots, 1 - x_n), \tag{4}$$

where $T$ is an arbitrary continuous t–norm and $S = T^d$ is the corresponding dual t–conorm. Observe that (4) can be formally generalized exploiting an arbitrary strong negation $N$, i.e., a decreasing involutive mapping $N : [0, 1] \to [0, 1]$, and putting $T_{\mathbf{w}}(x_1, \dots, x_n) = N(S_{\mathbf{w}}(N(x_1), \dots, N(x_n)))$. However, independently of $N$, we will obtain the same weighted t–norm as by (4) whenever the same t–norm is obtained by the corresponding dualities.

Note that axioms (W1), (W2) and (W3) are also fulfilled for weighted t–norms. Similarly as in the case of weighted t–conorms we have the following facts:

- $Min_{\mathbf{w}}(x_1, \ldots, x_n) = \min(x_i \mid w_i > 0)$;
- $T_{\mathbf{W}}$ is upper semi–continuous (right continuous);
- $T_{\mathbf{W}}$ (with some nontrivial $w_i \notin \{0, 1\}$) is continuous if and only if either $T = \min$ or $T$ is a continuous Archimedean t–norm;
- If $T$ is a continuous Archimedean t–norm with an additive generator $f : [0, 1] \to [0, +\infty]$, and $\mathbf{w}$ is a normal weighting vector, then

$$T_{\mathbf{W}}(x_1, \ldots, x_n) = f^{-1}\left(\sum_{i=1}^{n} w_i f(x_i)\right),$$

  i.e., $T_{\mathbf{W}}$ is a weighted quasi–arithmetic mean. It is cancelative whenever $T$ is nilpotent and it has annihilator 0 whenever $T$ is a strict t–norm.

For example, for the product t–norm $\Pi$, the relevant normal weighted function $\Pi_{\mathbf{W}}$ is just the weighted geometric mean.

Observe that if $\sum_{i=1}^{n} w_i = n$, then for a continuous Archimedean t–norm $T$ generated by an additive generator $f$ the corresponding weighted operator is given by $T_{\mathbf{W}}(x_1, \ldots, x_n) = f^{(-1)}\left(\sum_{i=1}^{n} w_i f(x_i)\right)$ what is just a weighted generated t–norm as proposed by Dubois and Prade in [21].

Several aggregation functions can be built by means of t–norms and t–conorms, for example, nullnorms, uninorms, $\Gamma$–operators, etc. Their weighted versions are then built from the corresponding weighted t–norms and t–conorms. For more details we recommend [10].

The basic idea of quantitative weights as cardinalities can be straightforwardly illustrated on the example of the weighted mean arising from the arithmetic mean. In statistics, starting with integer weights $n_i$, which are simply frequencies of observations $x_i$, the weighted mean is

$$M_{\mathbf{n}}(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} n_i\, x_i}{\sum_{i=1}^{n} n_i},$$

where $\mathbf{n} = (n_1, \ldots, n_n)$. Because of the strong idempotency of the standard arithmetic mean, $M_{\mathbf{n}}$ can be easily generalized into the form

$$M_{\mathbf{W}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i\, x_i, \quad w_i \geq 0, \ \sum_{i=1}^{n} w_i = 1.$$

The previous property of the standard arithmetic mean we can apply on any symmetric strongly idempotent extended aggregation function $A$. The strong idempotency of a symmetric extended aggregation function $A$ allows to introduce integer and rational quantitative weights – simply looking at them as cardinalities. In fact, we repeat the standard approach applied to the arithmetic mean as mentioned above. Indeed, for inputs $x_1, \ldots, x_n \in I$ and integer weights $\mathbf{w} = (w_1, \ldots, w_n) \in (\mathbb{N} \cup \{0\})^n$, we put

$$A_{\mathbf{w}}(x_1, \ldots, x_n) = A(\underbrace{x_1, \ldots, x_1}_{w_1-times}, \underbrace{x_2, \ldots, x_2}_{w_2-times}, \ldots, \underbrace{x_n, \ldots, x_n}_{w_n-times}). \tag{5}$$

Obviously, if $\mathbf{k} = (k, \ldots, k)$, $k \in \mathbb{N}$, is a constant weighting vector, the symmetry and the strong idempotency of $A$ result in $A_{\mathrm{k}}(\mathbf{x}) = A(\mathbf{x})$. This fact allows to define consistently the weighted aggregation in the case of rational weights $w_i \in \mathbb{Q}^+$. In that case we find such an integer $k \in \mathbb{N}$ that $k\, w_i \in \mathbb{N} \cup \{0\}$ for all $i = 1, \ldots, n$, and we put

$$A_{\mathbf{w}}(\mathrm{x}) = A_{k\mathbf{w}}(\mathrm{x}). \tag{6}$$

The resulting fused value in (4) does not depend on the actual choice of $k \in \mathbb{N}$. Further, because of (4) and (3), $A_{\mathbf{w}} = A_{p\mathbf{w}}$ for each positive rational $p$ and each rational weighting vector $\mathbf{w} \in (\mathbb{Q}^+)^n$, $\mathbf{w} \neq (0, \ldots, 0)$. Therefore we can deal with normed (rational) weighting vectors only, that is, we may suppose that $\sum_i w_i = 1$. The last problem we need to solve, is the case when also irrational weights $w_i$ are admitted.

**Definition 6.** Let $A : \bigcup_{n \in \mathbb{N}} I^n \to I$ be a symmetric strongly idempotent extended aggregation function. For any non–zero weighting vector $\mathbf{w} = (w_1, \ldots, w_n) \in [0, \infty[^n$, the corresponding $n$–ary weighted function $A_{\mathbf{w}} : I^n \to I$ is defined as follows:

(i) If all weights $w_i$ are rational, we apply formulas (4) and (3).
(ii) If there is some irrational weight $w_i$, denote $\mathbf{w}^* = (w_1^*, \ldots, w_n^*)$ the corresponding normed weighting vector, that is, $\mathbf{w} = \left(\sum_i w_i\right) \mathbf{w}^*$.

For any $m \in \mathbb{N}$, $i \in \{1, \ldots, n\}$, let

$$w_i^{(m)} = \min\left(\frac{j}{m\,!} \mid j \in \mathbb{N} \cup \{0\},\ \frac{j}{m\,!} \geq w_i^*\right),$$

and $\mathbf{w}^{(m)} = (w_1^{(m)}, \ldots, w_n^{(m)})$.
Then $w_i^{(m)} \in \mathbb{Q}^+$ and $\sum_i w_i^{(m)} \geq 1$ for all $m \in \mathbb{N}$ (and if already all weights $w_i^* \in \mathbb{Q}^+$, then also $w_i^{(m)} = w_i^*$ for all $i$ and all sufficiently large $m$) and we define

$$A_{\mathbf{w}}(\mathbf{x}) = \liminf_{m \to \infty} A_{\mathbf{w}^{(m)}}(\mathbf{x}) \quad \text{for all } \mathrm{x} \in I^n. \tag{7}$$

The following result can be straightforwardly checked from Definition 6.

**Proposition 2.** *Let* $\triangle = (\mathbf{w}^{(n)})_{n=1}^{\infty}$ *be a weighting triangle, i.e., for each* $n \in \mathbb{N}$*, let* $\mathbf{w}^{(n)} = (w_{1,n}, \ldots, w_{n,n})$ *be a non–zero weighting vector. Under the notations and requirements in Definition 6, define the function* $A_\triangle : \bigcup_{n\in\mathbb{N}} I^n \to I$*,* $A_\triangle(\mathbf{x}) = A_{\mathbf{w}^{(n)}}(\mathbf{x})$*, whenever* $\mathbf{x} \in I^n$*. Then* $A_\triangle$ *is a well defined idempotent extended aggregation function.*

Note that the approach allowing to introduce integer (rational) weights as given in formulas (3) and (4) was already applied to decomposable idempotent symmetric extended aggregation functions, see [24]. However, our results cover a wider class of symmetric strongly idempotent extended aggregation functions. For example, let $g : [0, 1] \to [0, 1]$ be given by $g(x) = 2x - x^2$. Define the function $A : \bigcup_{n\in\mathbb{N}} I^n \to I$ by

$$A(x_1, \ldots, x_n) = \sum_{i=1}^{n} \left( g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right)\right) x'_i,$$

where $x'_i$ is the $i$–th order statistics from the sample $(x_1, \ldots, x_n)$. Then $A$, which is an extended OWA operator, is a symmetric strongly idempotent extended aggregation function which is not decomposable. Further observe that the limit in formula (7) need not exist, in general.

The idea of qualitative weights incorporation into aggregation is linked to the transformation of the inputs by means of the corresponding weights from [0, 1] (as parameters expressing the importance of the corresponding input coordinates/criteria),

$$A_{\mathbf{w}}(\mathrm{x}) = A(h(w_1, x_1), \ldots, h(w_n, x_n)), \tag{8}$$

where $h : [0, 1] \times I \to [0, 1]$ is an appropriate binary function. This idea was already applied, e.g., in expert systems, and for $I = [0, 1]$ it was introduced by Yager in [51], where $h$ is a function called a RET operator. More details about RET operators can also be found in [44].

To ensure (W1), the following property of $h$ is required:
(RET1) $h(1, x) = x$ for all $x \in I$ .
Similarly, to ensure (W2), $A$ is supposed to have a neutral element $e$ and then
(RET2) $h(0, x) = e$ for all $x \in I$.
Further, to ensure the monotonicity of $A_{\mathbf{w}}$, one requires
(RET3) $h(w, \cdot)$ is non-decreasing for all $w \in [0, 1]$.
Finally, to ensure the boundary conditions of aggregation functions, one requires
(RET4) $h(\cdot, b)$ is non-decreasing for all $b \geq e$;
(RET5) $h(\cdot, b)$ is non-increasing for all $b \leq e$.

**Proposition 3.** *Let* $A : \bigcup_{n\in\mathbb{N}} I^n \to I$ *be an extended aggregation function with neutral element e and let* $h : [0, 1] \times I \to I$ *fulfil properties (RET1)–(RET5). For any weighting vector* $\mathbf{w} \in [0, 1]^n$, $\max w_i = 1$, *define the function* $A_{\mathbf{w}}$ *by (1). Then* $A_{\mathbf{w}}$ *is an n-ary aggregation function satisfying axioms (W1), (W2) and (W3).*

We only recall a typical example of a RET operator given by

$$h:\ h(w,x) = (x-e)w + e.$$

If $e = 0$ and $I = [0,1]$, any binary semicopula fulfills (RET1)–(RET5), while for $e = 1$, any fuzzy implication satisfying the neutrality principle, which corresponds to (RET1), see, e.g. [28], can be applied.

In some special cases, $h$ can also be defined for weights exceeding 1, that is, $h$ maps $[0,\infty[\times I$ into $I$. For example, recall the introduction of weights for continuous t–norms and t–conorms. Take, e.g., a strict t-norm $T$ with an additive generator $f : [0,1] \to [0,\infty]$. Then $h(w,x) = f^{-1}(wf(x))$, and for an arbitrary weighting vector w (the only constraint is $\sum w_i > 0$) we can put $T_{\mathbf{w}}(\mathbf{x}) = f^{-1}\left(\sum w_i f(x_i)\right)$.

Recall that special classes of anonymous (i.e., symmetric) aggregation functions with neutral elements appropriate for qualitative weights incorporation are triangular norms, triangular conorms, uninorms.

Projections to a distinguished subspace of some metric space are often applied operators which are usually related to some (constraint) optimization problem. The crucial role is played here by the underlying metric, and in fact, we are always looking for the best approximation of a discussed point by some point from the considered subspace. A similar philosophy can be found in defuzzification methods [20, 50], where a fuzzy quantity is characterized by a unique real number. Based on the just mentioned ideas, we introduce a metric–like function on the space of all possible scores (finitely dimensional inputs from some real interval or ordinal scale). Next we transform our metric–like function into a fuzzy relation. This approach is already standard in the domain of $T$–equivalence relations where the transformation was done, see, e.g., [17, 39, 40]. For a fixed score $(x_1, x_2, \dots, x_n)$, we will look for an appropriate "projection" to the subspace of all unanymous scores $(r, r, \dots, r)$, $r \in I$, applying some defuzzification method. Thus, in fact, we will define a function with inputs and outputs from some real interval $I$. In the special case of the $MOM$ defuzzification method we will rediscover a generalization of the penalty method introduced by Yager and Rybalov [52], see also [9].

For a fixed real interval $I$ and $n \in \mathbb{N}$ we introduce a dissimilarity function $D : I^n \times I^n \to [0,\infty[$ by

$$D(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} D_i(x_i, y_i), \tag{9}$$

where all $D_i : I^2 \to [0,\infty[$ are particular one-dimensional dissimilarity functions, $D_i(x,y) = K_i(f_i(x) - f_i(y))$, with $K_i : ]-\infty,\infty[ \to ]-\infty,\infty[$ a convex function with the unique minimum $K_i(0) = 0$, and $f_i : I \to ]-\infty,\infty[$, a strictly monotone continuous real function. For more details see [34]. Note that if $K_i$ are even functions then $D$ is a metric on $I^n$.

**Definition 7.** *For a given dissimilarity D, the function $U : I^n \to [0,1]^I$ which assigns to a score* x *the fuzzy subset $U_{\mathrm{x}}$ of I with the membership function*

$$U_{\mathrm{x}}(r) = \frac{1}{1 + D(\mathrm{x}, \mathrm{r})}, \qquad (10)$$

*where* $\mathrm{r} = (r, \ldots, r)$*, will be called a D–fuzzy utility function.*

**Proposition 4.** *Each D–fuzzy utility function U assigns to each score* $\mathbf{x} \in I^n$ *a continuous quasi-convex fuzzy quantity* $U_{\mathrm{x}}$*, i.e., for all* $r, s \in I$*,* $\lambda \in [0, 1]$*,*

$$U_{\mathbf{x}}(\lambda \cdot r + (1 - \lambda)s) \geq \min(U_{\mathrm{x}}(r), U_{\mathbf{x}}(s)),$$

*and thus for any* $\alpha \in ]0, 1]$ *the* $\alpha$*–cut* $U_{\mathbf{x}}^{\alpha} = \{r \in I \mid U_{\mathrm{x}}(r) \geq \alpha\}$ *is a closed subinterval of I in the standard topology.*

For each defuzzification method $DEF$ acting on quasi–convex (continuous) fuzzy quantities, we can assign to each score x a characteristic $DEF(U_{\mathrm{x}})$. Supposing that for any fuzzy quantity $Q$, $DEF(Q) \in supp(Q)$, $DEF(U)$ is an $I^n \to I$ function. In general, this function must be neither idempotent nor non–decreasing. Note that in [34], the conditions on $DEF$ ensuring the idempotency and monotonicity of the aggregation function $DEF(U)$ are discussed. Observe that the $MOM$ defuzzification method (Mean of Maxima) satisfies these conditions and thus we will illustrate our approach on the $MOM$ defuzzification. Note that $MOM(U)(\mathrm{x}) = \frac{1}{2}\left(\inf U_{\mathrm{x}}^{\alpha^*} + \sup U_{\mathrm{x}}^{\alpha^*}\right)$, where $\alpha^* = \sup\{\alpha \in ]0, 1] \mid U_{\mathrm{x}}^{\alpha} \neq \emptyset\}$.

**Definition 8.** *For a given dissimilarity D, the* $MOM$*-based operator* $MOM(U)$ *will be denoted by* $A_D$*.*

As already mentioned above, for any dissimilarity $D$, $A_D$ is an idempotent aggregation function.

*Example 1.*

(i) For $D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} (f(x_i) - f(y_i))^2$, we have $A_D(\mathbf{x}) = f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right)$, i.e., $A_D$ is a quasi–arithmetic mean.

(ii) For $D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \mid x_i - y_i \mid$, we have $A_D(\mathbf{x}) = \mathrm{med}(x_1, \ldots, x_n)$, i.e., the median operator.

(iii) For $n = 2$, $D(\mathbf{x}, \mathbf{y}) = \mid x_1 - y_1 \mid + (x_2 - y_2)^2$, we have $A_D(\mathbf{x}) = \mathrm{med}(x_1, x_2 - 1/2, x_2 + 1/2)$.

(iv) For $D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} D_c(x_i, y_i)$, where $D_c(x, y) = \begin{cases} c(y - x), & \text{if } x \leq y \\ x - y, & \text{else} \end{cases}$, $A_D$ is the $\alpha$–quantil (order statistics) with $\alpha = \frac{1}{(1+c)}$.

(v) For $D(\mathbf{x}, \mathbf{y}) = \max_{i=1}^{n} \mid x_i - y_i \mid$ we have $A_D(\mathbf{x}) = \frac{\min_i x_i + \max_i x_i}{2}$, i.e., $A_D$ is a special OWA operator.

Dissimilarity based approach to aggregation functions allows a straightforward incorporation of weights. For a weighting vector $\mathbf{w} = (w_1, \ldots, w_n)$, the weighted dissimilarity $D_{\mathbf{W}}$ will be given by $D_{\mathbf{W}}(\mathrm{x}, \mathrm{y}) = \sum_{i=1}^{n} w_i \, D_i(x_i, y_i)$ and then we will apply Definition 8 to obtain the corresponding weighted aggregation function. In the case of standard aggregation functions we have obtained in Example 1 (i) and (ii), the standard weighted quasi–arithmetic mean and the weighted median are obtained, respectively. The weighted aggregation function corresponding to Example 1 (iii) is given by $A_{D_{\mathbf{W}}}(\mathrm{x}) = \mathrm{med}(x_1, x_2 - \frac{w_1}{2w_2}, x_2 + \frac{w_1}{2w_2})$.

Finally, following the ideas of Yager [49], we propose to introduce OWAF (ordered weighted aggregation functions) as follows.

**Definition 9.** *Let $A_{\mathrm{w}} : I^n \to I$ be a weighted aggregation function. Then the operator $A'_{\mathbf{W}} : I^n \to I$ given by $A'_{\mathbf{W}}(\mathrm{x}) = A_{\mathbf{W}}(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$, where $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a permutation for which $x_{\sigma(1)} \le \ldots \le x_{\sigma(n)}$, will be called an OWAF.*

Evidently, starting from a weighted arithmetic mean $M_{\mathbf{W}}$, Definition 9 yields the OWA operator $M'_{\mathbf{W}}$. Note that the ordered weighted $t$–norm $T'_{(0,1,1)}(x, y, z) = \beta \cdot \gamma$ and its dual ordered weighted $t$–conorm $S'_{(1,1,0)}(x, y, z) = \alpha + \beta - \alpha\beta$, $\alpha = \min(x, y, z)$, $\beta = \mathrm{med}(x, y, z)$, $\gamma = \max(x, y, z)$, were found to be important in the study of fuzzy preference structures [16].

## 6 Conclusion

We have discussed some aspects of the theory of aggregation functions, including the review of some properties and classes of aggregation functions, and some construction methods. Especially, we have splitted the properties of extended aggregation functions into local properties, i.e., the properties of relevant $n$–ary aggregation functions for each fixed $n$, and into global properties which are often called "strong". Global properties properties constraint different arities functions involved in each extended aggregation function and thus, in the next development of the theory of aggregation functions they should be investigated in more detail. We expect interesting generalizations based on modifications of these standard approaches in the near future. For example, copulas are due to their probabilistic nature strongly connected with the standard operations, especially with the sum. Switching to the possibilistic background which is related to the maximum, we end up with semicopulas. However, there are many appropriate pseudo–additions (t–conorms) varying between the sum and maximum.

**Acknowledgment** This work was supported by the grants VEGA 1/3012/06, 1/2032/05, Science and Technology Assistance Agency under the contract No. APVT–20–003204, and by the grant GAČR 402/04/1026.

## References


1. Alsina C, Frank MJ, Schweizer B (2003) Problems on associative functions. Aequationes Math 66:128–140.
2. Alsina C, Nelsen RB, Schweizer B (1993) On the characterization of a class of binary operations on distributions functions. Stat Probab Lett 17:85–89.
3. Bajraktarevič M (1963) Sur une généralisation des moyennes quasilinéaire. Publ Inst Math Beograd 17:69–76.
4. Bassan B, Spizzichino F (2005) Relations among univariate aging, bivariate aging and dependence for excheangable lifetimes. J Multivariate Anal 93:313–339.
5. Calvo T, Mesiar R (2001) Criteria importances in median-like aggregation. IEEE Trans on Fuzzy Systems 9:662–666.
6. Calvo T, Mesiar R (2001) Weighted means based on triangular conorms. Int J Uncertainty, Fuzziness and Knowledge–Based Systems 9:183–196.
7. Calvo T, Kolesárová A, Komorníková M, Mesiar R (2002) Aggregation Operators: Properties, Classes and Construction Methods. In: Calvo T, Mayor G, Mesiar R (eds) Aggregation Operators. Physica-Verlag, Heidelberg, pp. 3–107.
8. Calvo T, Mesiar R (2003) Fusion with quantitative weights. In: Proc. of Eusflat'03, Zittau, Germany, pp. 312–317.
9. Calvo T, Mesiar R, Yager RR (2004) Quantitative weights and aggregation. IEEE Transactions on Fuzzy Systems 12:62–69.
10. Calvo T, Mesiar R (2004) How to incorporate weights in the aggregation process? In: Atanassov KT, Hryniewicz O, Kacprzyk J (eds) Soft Computing Foundations and Theoretical Aspects. Akademicka Oficyna Wydawnicza, Warszawa.
11. Calvo T, Mesiarová A, Valášková Ľ (2003) Construction of aggregation operators: new composition method. Kybernetika 39:643–650.
12. Choquet G (1953–54) Theory of capacities. Ann Inst Fourier 5:131–295.
13. Darsow WF, Nguyen B, Olsen ET (1992) Copulas and Markov processes. Illinois J Math 36:600–642.
14. De Baets B (1995) Model implicators and their characterization. PhD Thesis, University Gent.
15. De Baets B (1999) Idempotent uninorms. Europ J Oper Research 180:631–642.
16. De Baets B, Fodor JC (eds) (2003) Principles of Fuzzy Preference Modelling and Decision Making. Academia Press, Ghent.
17. De Baets B, Mesiar R (1997) Pseudo-metrics and T-equivalences. J Fuzzy Mathematics 5:471–481.
18. Del Amo A, Montero J, Molina E (2001) Representation of recursive rules. Europ J Operat Research 130:29–53.
19. Denneberg D (1994) Non–additive Measure and Integral. Kluwer Academic Publishers, Dordrecht.
20. Driankov D, Hellendoorn H, Reinfrank M (1996) An Introduciton to Fuzzy Control. Second Edition, Springer Verlag, Berlin.
21. Dubois E, Prade H (1985) A review of fuzzy set aggregation connectives. Inform Sci 36:85–121.
22. Dubois D, Prade H (2004) On the use of aggregation operations in information fusion processes. Fuzzy Sets and Systems 142:143–161.
23. Fodor JC, Yager RR, Rybalov A (1997) Structure of uninorms. Int J Uncertainty, Fuzziness and Knowledge–Based Systems 5:411–427.
24. Fodor JC, Roubens M (1994) Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic Publishers, Dordrecht.
25. Frank MJ (1979) On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. Aequationes Math 19:194–226.
26. Grabisch M, Marichal J-L, Mesiar R, Pap E (200x) Aggregation Functions for Decision Making. Monograph in preparation.

27. Genest C, Quesada Molina JJ, Rodríguez Lallena JA, Sempi C (1999) A characterization of quasi–copulas. J Multivariate Anal 69:193–205.
28. Klement EP, Mesiar R, Pap E (2000) Triangular norms. Kluwer Academic Publishers, Dordrecht.
29. Klement EP, Mesiar R, Pap E (2004) Measure–based aggregation operators. Fuzzy Sets and Systems 142:3–14.
30. Klement EP, Mesiar R, Pap E (2002) Invariant copulas. Kybernetika 38:275–285.
31. Kolesárová A, Mesiar R, Sempi C (2006) Compatibility of copulas. In: Proccedings IPMU'06, Paris, 1:658–663.
32. Kolesárová A (2003) 1–Lipschitz aggregation operators and quasi–copulas. Kybernetika 39:615–629.
33. Marques Pereira RA, Ribeiro RA (2003) Aggregation with generalized mixture operators using weighting functions. Fuzzy Sets and Systems 137:43–58.
34. Mesiar R (200x) Fuzzy set approach to the utility, preference relations, and aggregation operators. Europ J Operational Research.
35. Mesiarová A (200x) A note on two open problems of Alsina, Frank and Schweizer. Aequationes Math.
36. Montero J (2004) Some new results on recursive aggregation rules. In: Proc. IPMU'2004, Perugia, pp. 1235–1242.
37. Nelsen RB (1999) An Introduction to Copulas. Lecture Notes in Statistics 139, Springer Verlag, New York.
38. Pap E (1995) Null–Additive Set Functions. Kluwer Academic Publishers, Dordrecht.
39. Pradera A, Trillas E (2002) A note on pseudo-metric aggregation. Int J General Systems 31:41–51.
40. Saminger S, Mesiar R, Bodenhofer U (2002) Domination of aggregation operators and preservation of transitivity. Int J Uncertainty, Fuzziness Knowledge-Based Systems 10:11–35.
41. Schweizer B, Sklar A (1983) Probabilistic Metric Spaces. North–Holland, New York.
42. Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. Publ Inst Statist Univ Paris 8:229–231.
43. Sugeno M (1974) Theory of Fuzzy Integrals and Applications. PhD Thesis, Tokyo Inst of Technology, Tokyo.
44. Šabo M, Kolesárová A, Varga Š (2001) Ret operators generated by triangular norms and copulas. Int J Uncertainty, Fuzziess and Knowlwdge–Based Systems 9:169–181.
45. Šabo M, Strežo P (2005) On reverses of some binary operations. Kybernetika 41:437–452.
46. Špirková J (2005) Weighting functions for aggregation operators. In: Proc. AGOP' 2005, Lugano, pp. 127–130.
47. Struk P (2006) Extremal fuzzy integrals. Soft Computing 10:502–505.
48. Narukawa Y, Torra V (2003) Twofold integral and multi–step Choquet integral. In: Proc. AGOP' 2003, Alcalá de Henares, pp. 135–140.
49. Yager RR (1988) On ordered weighted averaging aggregation operators in multi–criteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18:183–190.
50. Yager RR, Filev D (1994) Essentials of Fuzzy Modelling and Control. J.Wiley&Sons, New York.
51. Yager RR (2001) Uninorms in fuzzy system modeling. Fuzzy Sets and Systems 122:167–175.
52. Yager RR, Rybalov A (1997) Understanding the median as a fusion operator. Int J General Systems 26:239–263.

# Identification of Weights in Aggregation Operators

**Tomasa Calvo and Gleb Beliakov**

**Abstract** This chapter provides a review of various techniques for identification of weights in generalized mean and ordered weighted averaging aggregation operators, as well as identification of fuzzy measures in Choquet integral based operators. Our main focus is on using empirical data to compute the weights. We present a number of practical algorithms to identify the best aggregation operator that fits the data.

## 1 Introduction

Aggregation of several input values into a single output value is an indispensable tool in many disciplines. The arithmetic mean is probably the most commonly used aggregation operator, which simply averages the input values. Weighted arithmetic means are also commonly used when the inputs have associated weights, which usually reflect the relative importance of the inputs. There is also a wide range of other averaging aggregation operators, which provide great flexibility to the modeling process, among which Ordered Weighted Averaging (OWA), generalized means and Choquet integral based aggregation operators play an especially important role.

All weighted aggregation operators are defined with the help of an associated vector of weights (or, more generally, a weighting triangle which contains the weights of all $n$-ary aggregation operators, $n = 2, 3, \ldots$). Furthermore, Choquet integral based operators also require coefficients of an appropriate fuzzy measure, which number $2^n - 2$. The weights (coefficients) are problem specific, and typically have a direct interpretation in terms of the underlying task.

When it is possible to gather some data about the desired (or observed) performance of a system, which typically consists of pairs input-output, a key approach to identifying the most suitable aggregation operator is to calculate the weights from the data by using some kind of regression procedure. One has to be aware, however, that the weights are not fully independent (for instance they typically sum to one), and that the semantics of the problem may impose further quite complicated relationships. This implies that proper procedures for weight identification must accommodate a number of restrictions (constraints).

In this chapter we review a number of methods for weights identification based on empirical data, which are applicable to a wide range of aggregation operators. Our

H. Bustince et al. (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

main focus is on those methods that are based on classical constrained optimization problems, namely quadratic and linear programming problems. The main reason is that such methods guarantee the best solution and are by far more efficient than popular modern heuristics used in nonlinear programming.

## 2 Preliminaries

A recent comprehensive overview of aggregation operators is given in [11], from which we took some relevant definitions, see also [14, 15] and [28].

**Definition 1.** An *aggregation operator* is a function $F : \bigcup_{n\in\mathbb{N}} [0,1]^n \to [0,1]$ such that:

(i) $F(x_1,\ldots,x_n) \le F(y_1,\ldots,y_n)$ whenever $x_i \le y_i$ for all $i \in \{1,\ldots,n\}$.
(ii) $F(t) = t$ for all $t \in [0,1]$.
(iii) $F(\underbrace{0,\ldots,0}_{n-times}) = 0$ *and* $F(\underbrace{1,\ldots,1}_{n-times}) = 1$

Each aggregation operator $F$ can be represented by a family of $n$-ary operators $f_n : [0,1]^n \to [0,1]$ given by $f_n(x_1,\ldots,x_n) = F(x_1,\ldots,x_n)$.

An aggregation operator $F$ is called averaging if it is bounded by

$$\min(x) = \min_{i=1,\ldots,n} x_i \le F(x) \le \max_{i=1,\ldots,n} x_i = \max(x).$$

An aggregation operator is called idempotent if $F(t,t,\ldots,t) = t$ for any $t \in [0,1]$. Note that since aggregation operators are monotone, idempotency is equivalent to the averaging behaviour.

An aggregation operator is called symmetric (commutative) if $F(x) = F(x_P)$ for any $x \in [0,1]^n$ and any permutation $P$ of $\{1,\ldots,n\}$.

An aggregation operator $F$ has a weighting triangle $\triangle = (w_{in})$, if for each $f_n$ the vector of weights is $w_n = (w_{1n}, w_{2n}, \ldots, w_{nn}) \in [0,1]^n$ with $\sum_{i=1}^n w_{in} = 1$.

Throughout this paper the set of empirical data (dataset) will be denoted by $\mathcal{D} = \{(x^k, y^k)\}_{k=1}^K$. It consists of $K$ input-output tuples, $x^k \in [0,1]^n$, $y^k \in [0,1]$. The aim of the methods presented in this chapter is to identify the weighting vectors $w_n$ of aggregation operators of given classes by fitting the weights to the data, i.e., by minimizing the difference between the predicted and observed values

$$\left(\sum_{k=1}^K |f_n(x^k) - y^k|^p\right)^{1/p}, \; p \ge 1, \tag{1}$$

subject to $f_n$ being an aggregation operator from some class.

If we choose $p = 2$, we obtain the least squares problem, and if we choose $p = 1$ we obtain the least absolute deviation problem. The latter problem is frequently used in robust regression, since the optimal solution is less sensitive to outliers. Another possible choice is $p = \infty$, i.e., minimize the maximal difference

$$\max_{k=1,\ldots,K} |f_n(x^k) - y^k|.$$

The two cases $p = 1$ and $p = \infty$ sometimes present computational advantages, because the corresponding minimization problems can be solved by using linear programming (LP) methods. In the first case we use auxiliary variables $v_-^k, v_+^k$ : $v_+^k - v_-^k = f_n(x^k) - y^k$, in which case $v_+^k + v_-^k = |f_n(x^k) - y^k|$. In the second case we minimize $\varepsilon$, subject to $\varepsilon \geq f_n(x^k) - y^k$ and $\varepsilon \geq -(f_n(x^k) - y^k)$, $k = 1, \ldots, K$ along with other required constraints. In the case $p = 2$ we can use quadratic programming (QP) methods. A detailed overview of these and some other techniques for treating these cases can be found in [26, 39]. There are a number of specially adapted versions of the simplex method for numerical solution [2], and free program code is available at `http://www.netlib.org/`.

What distinguishes the methods described below is the type of constraints that appear in the minimization problem (1). These constraints depend on the class of aggregation operator, and also on some other properties that may be imposed to satisfy problem requirements.

In some cases it is possible to make a further step and identify not only weighting vectors $w_n$ for some $n$, but the whole weighting triangle $\triangle$, even if the data are of a fixed dimension. Of course, this is based on a given relationship between the weighting vectors. We will explore this relationship in Sect. 7.

## 3 Weighted Means

**Definition 2.** An aggregation operator is a *weighted arithmetic mean* if, for each $n \in \mathbb{N}$, it can be written as

$$f_n(x) = \sum_{i=1}^{n} w_{in} x_i$$

where $w_n = (w_{1n}, \ldots, w_{nn}) \in [0, 1]^n$ verifies $\sum_{i=1}^{n} w_{in} = 1$.

**Definition 3.** An aggregation operator is a *weighted quasi-arithmetic mean* if, for each $n \in \mathbb{N}$, it can be written as

$$f_n(x) = g^{-1}\left(\sum_{i=1}^{n} w_{in} g(x_i)\right)$$

where $g : [0, 1] \to [-\infty, +\infty]$ is a continuous strictly monotone function and $w_n = (w_{1n}, \ldots, w_{nn}) \in [0, 1]^n$ verifies $\sum_{i=1}^{n} w_{in} = 1$.

By taking $g(t) = t$ we obtain weighted arithmetic means, and by taking $g(t) = \log(t)$ we obtain weighted geometric mean

$$f_n(x) = \prod_{i=1}^{n} x_i^{w_{in}}.$$

If $g(t) = t^q, q \neq 0$, we obtain root-power operators, or weighted quasi-linear means. A distinguished member is the harmonic mean ($q = -1$). A detailed discussion is given in [1, 16, 11].

We note that the generating function $g$ is defined up to an arbitrary linear transformation, i.e., the function $\tilde{g}(t) = Ag(t) + B$ defines the same weighted quasi-arithmetic mean as $g$.

The simplest case is identification of weights of arithmetic means. Here we solve the following minimization problem (for a fixed $n$)

$$\text{minimize} \sum_{k=1}^{K} |\sum_{i=1}^{n} w_{in} x_i^k - y^k|^p, \tag{2}$$

subject to $w_{in} \geq 0$ and $\sum_{i=1}^{n} w_{in} = 1$. For $p = 2$ it is a quadratic programming problem, and for $p = 1$ and $p = \infty$ the problem is reduced to linear programming, as described earlier. The quadratic programming problem is efficiently solved by using *LSEI* formulation (Least Squares with Equality and Inequality constraints), see [26, 23, 24], and Algorithm 587 from `http://www.netlib.org`. This method is based on the active set approach, which is described in detail in [36, 37].

A particular attention is needed for the case when the quadratic (resp. linear) programming problems have singular matrices. Such cases appear when there are few data, or when the input values are linearly dependent. While modern quadratic and linear programming methods accommodate for such cases, the minimization problem will typically have multiple solutions. An additional criterion is then used to select one of these solutions, and typically this criterion relates to the dispersion of weights, or the entropy [36]. The measure of weights dispersion is defined as

$$Disp(w) = -\sum_{i=1}^{n} w_i \log w_i, \tag{3}$$

with the convention $0 \log 0 = 0$, and it measures the degree to which $F$ takes into account all information in the aggregation. For example, among the two optimal solutions $w^1 = (0, 1)$ and $w^2 = (0.5, 0.5)$ the second one is preferable, since the corresponding weighted mean operator uses information from two sources rather than a single source, and is consequently less sensitive to input inaccuracies. Torra [36] proposes to solve an auxiliary univariate optimization problem to maximize weights dispersion, subject to a given value of (3).

Consider now the case of quasi-arithmetic means, when a given generating function $g(t)$ is not affine. The optimization problem (1) becomes non-linear, however it can be reduced to the previously discussed constrained linear regression problem by a simple mathematical trick. Instead of (1) consider minimization of

$$\left(\sum_{k=1}^{K} |g(f_n(x^k)) - g(y^k)|^p\right)^{1/p}, p \geq 1, \tag{4}$$

subject to the same conditions. In the case of quasi-arithmetic means it converts to

$$\text{minimize} \sum_{k=1}^{K} |\sum_{i=1}^{n} w_{in} g(x_i^k) - g(y^k)|^p, \tag{5}$$

subject to $w_{in} \geq 0$ and $\sum_{i=1}^{n} w_{in} = 1$. This approach was discussed in detail in [36, 4, 9, 7].

Note that for the purposes of numerical solution, one can simply apply the solution to problem (2) to the data set $g\mathcal{D} = \{(g(x^k), g(y^k))\}_{k=1}^{K}$. Of course, in general solutions to problems (1) and (4) are different, but if the errors in the data are small, they will be close to each other. Further, if the solution to (1) is needed, the solution to (4) can be used as a good starting point in a local nonlinear optimization method.

## 4 Ordered Weighted Averaging

**Definition 4.** An aggregation operator is an *Ordered Weighted Averaging operator (OWA)* ([41, 42]) if, for each $n \in \mathbb{N}$, it can be written as

$$f_n(x) = \sum_{i=1}^{n} w_{in} x_{(i)}$$

where $w_n = (w_{1n}, \ldots, w_{nn}) \in [0, 1]^n$ verifies $\sum_{i=1}^{n} w_{in} = 1$ and $(x_{(1)}, \ldots, x_{(n)})$ is a vector obtained from $x$ by arranging its components in a non-increasing order.

**Definition 5.** An aggregation operator is a *generalized Ordered Weighted Averaging (GOWA) operator* ([11, 43]) if, for each $n \in \mathbb{N}$, it can be written as

$$f_n(x) = g^{-1}\left(\sum_{i=1}^{n} w_{in} g(x_{(i)})\right)$$

where $g : [0, 1] \to [-\infty, +\infty]$ is a continuous strictly monotone function, and $w_n, x_{()}$ are as in Definition 4. Such an operator is also called Ordered Weighted Quasi-Arithmetic mean (OWQA) in [11].

In [43] Yager proposed using $g(t) = t^q$ as a generalization of OWA operators,

$$f_n(x_1, \ldots x_n) = \left( \sum_{i=1}^{n} w_{in} x_{(i)}^q \right)^{1/q}, \tag{6}$$

which is similar to weighted quasi-linear means, but involves re-ordering of the components of $x$. The attitudinal character of the GOWA operator (or the measure of orness) is defined in [43] as

$$AC(w; p) = \left( \sum_{i=1}^{n} w_{in} \left( \frac{n-i}{n-1} \right)^q \right)^{1/q} = \frac{1}{n-1} \left( \sum_{i=1}^{n} w_{in} \, (n-i)^q \right)^{1/q}. \tag{7}$$

The values close to 0 correspond to min-like aggregation and values close to 1 correspond to max-like aggregation. For a general generating function $g$, this measure is defined as

$$AC(w; g) = g^{-1} \left( \sum_{i=1}^{n} w_{in} g \left( \frac{n-i}{n-1} \right) \right). \tag{8}$$

Torra [34] has generalized OWA operators in a different way, by using the second vector of weights that correspond to the individual inputs (WOWA operators). Consider two weighting vectors $w, p \in [0, 1]^n$, $\sum w_i = 1$, $\sum p_i = 1$. WOWA operator is a mapping

$$f_n(x) = \sum_{i=1}^{n} \omega_i x_{P(i)},$$

where $P$ is a permutation of $\{1, \ldots, n\}$ such that $x_{P(i-1)} \geq x_{P(i)}$, and the weights $\omega_i$ are defined as

$$\omega_i = W \left( \sum_{j \leq i} p_{P(j)} \right) - W \left( \sum_{j < i} p_{P(j)} \right).$$

$W$ is a monotone increasing function interpolating the points $(i/n, \sum_{j \leq i} w_j)$ and $(0, 0)$. $W$ is required to be linear if these points lie on a straight line.

The problem of identification of weights of OWA operators was studied by several authors [17, 18, 40, 45]. A common feature of all methods is to eliminate nonlinearity due to reordering of the components of $x$ by restricting the domain of the aggregation operator to the simplex $S \subset [0, 1]^n$ defined by the inequalities $x_1 \geq x_2 \geq \ldots \geq x_n$. On that domain OWA operator is a linear function. Once the coefficients of this function are found, OWA operator can be computed on the whole $[0, 1]^n$ by using its symmetry. Algorithmically, it amounts to using an auxiliary data set $\mathcal{D}_{()} = \{(z^k, y^k)\}$, where vectors $z^k = x^k_{()}$. Thus identification of weights of

OWA is a very similar problem to identification of weights of arithmetic means. Depending on the parameter $p$ being used, we solve it by using either quadratic or linear programming techniques.

Filev and Yager [17] proposed a nonlinear change in variables to obtain unrestricted minimization problem, which they propose to solve using nonlinear local optimization methods. Unfortunately the resulting nonlinear optimization problem is difficult due to a large number of local minimizers, and the traditional optimization methods are stuck in the local minima.

The approach relying on quadratic programming was used in [44, 36, 37, 4, 7], and it was shown to be numerically efficient and stable with respect to rank deficiency (e.g., when $K < n$, or the data are linearly dependent).

Often an additional requirement is needed: a desired value of the measure of orness $AC = \alpha$. This requirement can be easily incorporated into QP or LP problem as an additional linear equality constraint.

For generalized OWA operators with a given generating function $g$, the approach is similar to the one used for quasi-arithmetic means:

$$\text{minimize} \sum_{k=1}^{K} |\sum_{i=1}^{n} w_{in} g(x_{(i)}^k) - g(y^k)|^p, \tag{9}$$

subject to $w_{in} \geq 0$, $\sum_{i=1}^{n} w_{in} = 1$, and possibly $g(AC) = \sum_{i=1}^{n} w_{in} g\left(\frac{n-i}{n-1}\right) = g(\alpha)$. More details about this method are found in [36, 4, 9, 7].

The method proposed in [18, 31] identifies OWA operator weights with the maximal entropy, subject to a given measure of orness. This method does not use empirical data, and aims at generating the weights with maximal dispersion, as discussed in the previous section. Maximization of weights dispersion is also useful for data driven weights identification, if there are multiple optimal solutions. Then the solution maximizing $Disp(w)$ is chosen, as discussed in [36].

Identification of weights of WOWA operators is more complicated due to nonlinearities that appear in the definitions of weights $\omega_i$. This problem can be formulated as

$$\text{minimize} \sum_{k=1}^{K} \left( \sum_{i=1}^{n} \omega_i(p, w, P) x_{P(i)}^k - y^k \right)^2,$$

and the minimum with respect to $p, w$ can be found using nonlinear optimization methods. A cheaper alternative is to obtain the weights $\omega_i(P)$, after recognizing them as coefficients of a fuzzy measure, and WOWA operator as Choquet integral [35] (see Sect. 6).

## 5 Identification of the Generating Function

In this section we present the methods aiming at identifying not only the weights of quasi-arithmetic means and generalized OWA operators, but also the unknown generating function $g$. Previously we discussed the case of a fixed $g(t)$ in both classes of aggregation operators, and have shown that the optimal weights can be found by solving problem (5) or (9).

Now consider the function $g$ from some parametric family, but with unknown parameter $q$, such as $g(t) = t^q$. We denote such a family by $g(t; q)$. We convert the optimization problem with respect to both $q$ and $w$ into a bi-level problem

$$\min_q \min_w \sum_{k=1}^{K} | \sum_{i=1}^{n} w_{in} g(x_i^k; q) - g(y^k; q)|^p, \tag{10}$$

where at the inner level we solve problem (5) (or (9) respectively) with a fixed $q$. While the outer problem is nonlinear, it involves a single variable $q$, and thus its global optimum can be found iteratively by using some global optimization method, such as Pijavski algorithm [32]. However we need to ensure the efficiency of each iteration, i.e., use efficient algorithms for solving the inner problem.

A different situation arises when the parametric form of $g(t)$ is not given. The approach proposed in [4, 6, 9] is based on approximation of $g$ with a monotone spline (in particular linear spline), as

$$g(t) = \sum_{j=1}^{J} c_j B_j(t), \tag{11}$$

where $B_j$ are appropriately chosen basis functions, and $c_j$ are spline coefficients. The monotonicity of $g$ is ensured by imposing linear restrictions on spline coefficients, in particular non-negativity, as in [3]. Further, since the generating function is defined up to an arbitrary linear transformation, one has to fix a particular $g$ by specifying two interpolation conditions, like $g(a) = 0$, $g(b) = 1$, $a, b \in (0, 1)$, and if necessary, properly model asymptotic behaviour if $g(0)$ or $g(1)$ are infinite, see [6, 9].

After rearranging the terms of the sum, the problem of identification becomes [7]

$$\min_{c,w} \left( \sum_{k=1}^{K} \left| \sum_{j=1}^{J} c_j \left[ \sum_{i=1}^{n} w_{in} B_j(x_i^k) - B_j(y^k) \right] \right|^p \right)^{1/p}. \tag{12}$$

For a fixed $c$ (i.e., fixed $g(x)$) we have either a quadratic or linear programming problem to find $w$, and for a fixed $w$, we have a quadratic or linear programming problem to find $c$. However if we consider both $c, w$ as variables, we obtain a difficult global optimization problem. We convert it into a bi-level optimization problem

$$\min_c \min_w \left( \sum_{k=1}^{K} \left| \sum_{j=1}^{J} c_j \left[ \sum_{i=1}^{n} w_{in} B_j(x_i^k) - B_j(y^k) \right] \right|^p \right)^{1/p},$$

where at the inner level we have a QP or LP problem and at the outer level we have a nonlinear problem with multiple local minima. When the number of spline coefficients $J$ is not very large ($< 10$), this problem can be efficiently solved by using deterministic Cutting Angle method of global optimization [5]. If the number of variables is small and $J$ is large, then reversing the order of minimization (i.e., using $\min_w \min_c$) is more efficient.

## 6 Choquet Integral Based Aggregation

Choquet integral is frequently used as an aggregation tool [10, 11, 22]. The Choquet integral based aggregation operator is defined as

$$C_v(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_{(i)}[v(H_i) - v(H_{i+1})], \tag{13}$$

where $v : 2^N \to I$ is a fuzzy measure on the set $N = \{1, 2, \ldots, n\}$, which is a monotonic (i.e. $v(S) \leq v(T)$ whenever $S \subseteq T$) set function satisfying $v(\emptyset) = 0, v(N) = 1$; $x_{(1)} \leq \ldots \leq x_{(n)}$, $H_i = \{(i), \ldots, (n)\}$, and $H_{n+1} = \emptyset$ by convention. Equation (13) can also be written as [20, 21], p. 110,

$$C_v(x_1, \ldots, x_n) = \sum_{i=1}^{n} \left[ x_{(i)} - x_{(i-1)} \right] v(H_i). \tag{14}$$

where $x_{(0)} = 0$ by convention. In this notation, $C_v$ is a *linear* function of the coefficients of the fuzzy measure $v(H_i)$. Note that the order in which the components of $x_{()}$ are arranged is different to the order used in Sect. 4, this is done for consistency with the literature on OWA and Choquet integral based operators.

In multicriteria decision making, Choquet aggregation explicitly models the importance of not only individual criteria, but of their subsets, as well as various interactions between the criteria. In the context of learning aggregation operators from data, identification of Choquet aggregation operator is equivalent to identification of the fuzzy measure $v(T)$, described by $2^n$ coefficients. This problem was addressed in [8, 9, 13, 22].

Note that if fuzzy measure is symmetric, Choquet integral based aggregation becomes an OWA operator, and if the fuzzy measure is additive, we obtain a weighted arithmetic mean.

Identification of the $2^n - 2$ coefficients (two are given explicitly as $v(\emptyset) = 0$, $v(N) = 1$) involves a quadratic ($p = 2$) or linear ($p = 1, \infty$) programming problem

$$\text{minimize} \sum_{k=1}^{K} |C_v(x_1^k, \ldots, x_n^k) - y^k|^p,$$

subject to the conditions of monotonicity of the fuzzy measure (they translate into a number of linear constraints). In this problem the main difficulty is the large number of unknowns, and typically a much smaller number of data. As we mentioned, modern LP and QP methods handle well the resulting degenerate systems of equations, however for $n \geq 10$ one also has to take into account the sparse structure of such systems. Most LP solvers handle sparse systems, and are very efficient, for QP we recommend OOQP sparse solver [19] (`http://www.cs.wisc.edu/~swright/ooqp/`).

There are also other requirements that can be imposed on the fuzzy measure from the problem specifications. For example, if it is known that certain variables are sub-additive, super-additive or substitutive, then the following additional constraints are added [8]. For sub-additive criteria we add

$$v(T \cup \{i, j\}) \leq v(T \cup \{i\}) + v(T \cup \{j\}) - v(T), \quad \forall T \subseteq N \setminus \{i, j\}. \tag{15}$$

For substitutive criteria we add

$$\begin{aligned} \forall T \subseteq N \setminus \{i, j\} : v(T \cup \{i, j\}) &\leq v(T \cup \{i\}) + (1 - \nu_{ij}) v(T \cup \{j\}), \\ v(T \cup \{i, j\}) &\leq v(T \cup \{j\}) + (1 - \nu_{ij}) v(T \cup \{i\}), \end{aligned} \tag{16}$$

where $\nu_{ij} \in [0, 1]$ is the degree of substitutivity. When $\nu_{ij} = 1$, we obtain the equalities $v(T \cup \{i, j\}) = v(T \cup \{i\}) = v(T \cup \{j\})$, i.e., fully substitutive (identical) criteria.

As a way of reducing the complexity of the fuzzy measure (the number of coefficients), Grabisch [20] introduces the concept of k-additive fuzzy measures (see also [21, 30]). A fuzzy measure $v$ is $k$-additive whenever its Möbius transform vanishes on sets containing more than $k$ elements, that is, if $card(E) > k$, then

$$\mu(E) = \sum_{F \subset E} (-1)^{card(E \setminus F)} v(F) = 0.$$

This condition also translates into a system of linear inequalities, which can be added to the LP or QP for fuzzy measure identification. There are also other possible restrictions on the fuzzy measure $v$, like given values of Shapley index (or its generalizations, like interaction indices [20]), which also translate into linear inequalities involving components of $v$.

Another way of reducing the complexity of a fuzzy measure is to use $\lambda$-fuzzy measures, introduced in [33]. These fuzzy measures satisfy the following conditions $\forall F, E \subseteq N, F \cap E = \emptyset$:

$$v(F \cup E) = v(F) + v(E) + \lambda v(F) v(E) \tag{17}$$

where $\lambda \in (-1, \infty)$. Under these conditions, all the coefficients $v(F)$ are immediately computed from $n$ independent coefficients $v(\{i\}), i = 1, \dots, n$. We also have an explicit formula for a family of disjoint subsets of $N$

$$v_\lambda\left(\bigcup_{i=1}^{m}\{i\}\right) = \frac{1}{\lambda}\left(\prod_{i=1}^{m}(1 + \lambda v(\{i\})) - 1\right).$$

The coefficient $\lambda$ is determined from the boundary condition $v_\lambda(N) = 1$, which gives

$$\lambda + 1 = \prod_{i=1}^{n}(1 + \lambda v(\{i\})),$$

which can be solved on $(-1, 0)$ or $(0, \infty)$ numerically (note that $\lambda = 0$ is always a solution).

The price for the reduction of the number of variables in $\lambda$-fuzzy measures is that the resulting optimization problem is nonlinear, and we cannot apply QP or LP methods as earlier. Most research in this area concentrated on genetic algorithms, and we refer the interested reader to the articles [12, 25, 27, 38].

Yager proposes in [43] a generalized Choquet aggregation operator

$$C_{v,q}(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} x_{(i)}^q [v(H_i) - v(H_{i+1})]\right)^{1/q}. \tag{18}$$

It is not difficult to see that this implies

$$C_{v,q}(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} \left[x_{(i)}^q - x_{(i-1)}^q\right] v(H_i)\right)^{1/q}. \tag{19}$$

The sum in the brackets is again a linear function of the fuzzy measure coefficients. Thus we can apply the methods of fuzzy measure identification studied in [8, 9, 22], with one distinction that the data are linearized (i.e., taking $\{(x_{(i)}^k)^q, (y^k)^q\}$). The problem becomes a quadratic or linear programming problem for a fixed $q$, and it is solved as a bi-level optimization problem if $q$ also has to be identified from the data.

## 7 Generated Aggregation Operators

We now consider a situation where not only a particular vector of weights $w_n$ for a fixed $n$ needs to be identified, but the whole weighting triangle (i.e., all weight vectors $w_n, n = 2, 3, \dots$). We note that there will be no sufficiently many data to identify all weighting vectors independently, instead we want to exploit certain relations between aggregation operators of different dimensions but from a single family. For this purpose we consider identification of generating functions that can generate the whole weighting triangle. The following section summarizes the results presented in [9, 29].

Let us use the following notation: $N = \{1, \ldots, n\}$ and $\mathcal{A} = \{(n, E) | n \in \mathbb{N}, E \subseteq N\}$.

**Definition 6.** *A mapping* $\mathrm{M} : \mathcal{A} \to [0, 1]$ *is called a universal fuzzy measure whenever for each fixed* $n \in \mathbb{N}$, $v = \mathrm{M}(n, \cdot)$ *is a fuzzy measure, that is*

1. $\mathrm{M}(n, \emptyset) = 0, \mathrm{M}(n, N) = 1,$
2. $\mathrm{M}(n, E) \leq \mathrm{M}(n, F)$ *for all* $E \subseteq F \subseteq N$.

For a given universal fuzzy measure **M**, an aggregation operator can be built by means of any fuzzy integral [10]. We restrict our attention to the Choquet integral, and thus

$$f_n(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \mathrm{M}(n, E_{(i)}), \tag{20}$$

where $x_{(1)} \leq \ldots \leq x_{(n)}$, $E_{(i)} = \{(i), \ldots, (n)\}$ and $x_{(0)} = 0$ by convention. Let us also use notation $x_{()} = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$.

It is possible to construct *symmetric* and *additive* universal fuzzy measures with the help of a one-dimensional generator $g : [0, 1] \to [0, 1]$, a non-decreasing function with $g(0) = 0, g(1) = 1$ [29], namely

$$\mathrm{M}^g_{symm}(n, E) = g\left(\frac{card(E)}{n}\right), \tag{21}$$

and

$$\mathrm{M}^g_{add}(n, E) = \sum_{i \in E} \left(g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right)\right). \tag{22}$$

The corresponding Choquet integral based aggregation operators are given as follows. In the case of symmetric universal fuzzy measure it is

$$f_n(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_{(i)} \left(g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right)\right) = \sum_{i=1}^{n} w_{in} x_{(i)} \tag{23}$$

i.e., it is the OWA operator with the weights

$$w_{in} = g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right). \tag{24}$$

In the case of an additive fuzzy measure it is

$$f_n(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i \left( g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right)\right) = \sum_{i=1}^{n} w_{in} x_i \tag{25}$$

which is a weighted arithmetic mean with the weights as above. The only universal fuzzy measure which is both symmetric and additive is linked to the identity generator $g(x) = x$, in which case $f_n$ is the arithmetic mean.

The convexity of $g$ on [0, 1] is the necessary condition for a symmetric universal fuzzy measure $\mathrm{M}^g$ to be a belief measure (i.e., each $\mathrm{M}^g(n, \cdot)$ is a belief measure), while the concavity of $g$ on [0, 1] is the necessary condition for $\mathrm{M}^g$ to be a plausibility measure (i.e., each $\mathrm{M}^g(n, \cdot)$ is a plausibility measure).

Thus one can build Choquet integral based general aggregation operators with the help of a one-dimensional generator $g$, by defining universal fuzzy measures using (21) (symmetric measure) and (22) (additive measure). It is now our task to determine not only particular vectors of weights for fixed $n$, but the generating function $g$ itself, which will allow us to identify the weights for any given $n$.

Consider the data set $\mathcal{D}$ and a symmetric Choquet integral based aggregation operator (23),(24). No constraints on the generator $g$ besides monotonicity and boundary conditions $g(0) = 0$, $g(1) = 1$ are given. Let us represent $g$ with a linear spline (11). Now write (1), (23) as

$$\begin{aligned} f_n(x_1^k, \ldots, x_n^k) &= \sum_{i=1}^{n} x_{(i)}^k \left( g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right)\right) \\ &= \sum_{i=1}^{n} x_{(i)}^k \left( \sum_{j=1}^{J} c_j \left[ B_j\left(\frac{i}{n}\right) - B_j\left(\frac{i-1}{n}\right)\right]\right) \\ &= \sum_{j=1}^{J} c_j \left( \sum_{i=1}^{n} x_{(i)}^k \left[ B_j\left(\frac{i}{n}\right) - B_j\left(\frac{i-1}{n}\right)\right]\right) \\ &\approx y^k. \end{aligned}$$

Similarly, we obtain for an additive Choquet integral based aggregation operator (25)

$$f_n(x_1^k, \ldots, x_n^k) = \sum_{j=1}^{J} c_j \left( \sum_{i=1}^{n} x_i^k \left[ B_j\left(\frac{i}{n}\right) - B_j\left(\frac{i-1}{n}\right)\right]\right) \approx y^k,$$

Thus we obtain the following optimization problem

$$\text{minimize } ||Ac - y||_p^p$$

subject to non-negativity of the components of $c$. The components of the matrix $A$ are given as

$$A_{kj} = \sum_{i=1}^{n} x_{(i)}^{k} \left[ B_j\left(\frac{i}{n}\right) - B_j\left(\frac{i-1}{n}\right) \right],$$

or

$$A_{kj} = \sum_{i=1}^{n} x_{i}^{k} \left[ B_j\left(\frac{i}{n}\right) - B_j\left(\frac{i-1}{n}\right) \right],$$

respectively. Note that for $p = 2$ we obtain a QP, and for $p = 1, \infty$ we obtain an LP problem.

It is also possible to place other restrictions on the generated fuzzy measure $\mathrm{M}^g$ to make it either belief or plausibility measure. It amounts to forcing the spline $S$ to be concave or convex, which translates into further linear constraints on $c_j$, namely $c_j - c_{j-1} \leq 0$, or $c_j - c_{j-1} \geq 0,\ j = 2, \ldots, J$.

For symmetric $k$-additive universal fuzzy measures we have the following characterization of the generating function $g$.

**Proposition 1.** *[29] Let* $\mathrm{M}^g$ *be a symmetric universal fuzzy measure generated by a monotone generator* $g : [0,1] \to [0,1]$, $\mathrm{M}^g(n, E) = g(card(E)/n)$. *Then:*
*i)* $\mathrm{M}^g$ *is 2-additive if and only if* $g(x) = ax^2 + (1-a)x$ *for some* $a \in [-1, 1]$*;*
*ii)* $\mathrm{M}^g$ *is 3-additive if and only if* $g(x) = ax^3 + bx^2 + (1-a-b)x$ *for some* $a \in [-2, 4]$ *and b such that*
*- if* $a \in [-2, 1]$ *then* $b \in [-2a - 1, 1 - a]$*, and*
*- if* $a \in [1, 4]$ *then* $b \in [-3a/2 - \sqrt{3a(4-a)/4}, -3a/2 + \sqrt{3a(4-a)/4}]$ .

From the properties of $g$ we are able to characterize the subsets in the space of parameters $a, b$ on which the measure $\mathrm{M}^g$ is either belief or plausibility measure. Recall that the convexity of $g$ (respectively concavity of $g$), are the necessary conditions. It can be shown that for quadratic functions these are also sufficient conditions. In the case of a cubic generator $g$, for sufficiency we additionally need non-negativity of its third derivative. Hence, in the case of 2-additive symmetric fuzzy measure, it is a belief measure if and only if $a \in [0, 1]$, and it is a plausibility measure if and only if $a \in [-1, 0]$. For 3-additive case, $\mathrm{M}^g$ is a belief measure if and only if $a, b$ satisfy $a + b \leq 1, b \geq 0, a \geq 0, b \geq -3a$. It is a plausibility measure if and only if the parameters satisfy $2a + b \geq -1, b \leq 0, a \geq 0, b \leq -3a$.

This explicit characterization of the generators $g$ of 2- (3-) additive symmetric universal fuzzy measures simplifies the problem of fitting these generators to the data. The algebraic form of $g$ is now fixed, and one (two) linear parameters must be found from the data using linear regression. The additivity of the Choquet integral with respect to fuzzy measures allows one to simplify computations. If $g(x) = ax^2 + (1-a)x$ for some $a \in [-1, 1]$, then

$$f_n(x_1, x_2, \ldots, x_n) = aA_2(x_1, x_2, \ldots, x_n) + (1-a)A_1(x_1, x_2, \ldots, x_n),$$

where $A_1$ is the standard arithmetic mean and $A_2$ is an OWA operator with the weights $\left(\frac{1}{n^2}, \frac{3}{n^2}, \ldots, \frac{2n-1}{n^2}\right)$. Thus we have

$$
\begin{aligned}
f_n(x_1^k,\dots,x_n^k) &= \sum_{i=1}^{n}\left(a\frac{2i-1}{n^2}+\frac{1-a}{n}\right)x_{(i)}^k\\
&= a\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2i-1}{n}-1\right)x_{(i)}^k\right]+\left[\sum_{i=1}^{n}\frac{x_{(i)}^k}{n}\right]\\
&= S_1^k a + S_2^k \approx y^k,\ k=1,\dots,K,
\end{aligned}
$$

which is a trivial univariate regression problem, subject to $a \in [-1, 1]$. The values of $S_1^k$ and $S_2^k$ are the expressions in the square brackets. The explicit solution is

$$
a = \max\{-1, \min\{1, \frac{\sum_{k=1}^{K}\left(y^k - S_2^k\right)S_1^k}{\sum_{k=1}^{K}\left(S_1^k\right)^2}\}\}.
$$

For 3-additive symmetric universal fuzzy measure the domain for parameters $a, b$ is a convex compact set consisting of a triangle determined by the vertices (-2,3),(1,0),(1,-3) and an ellipse determined by the curve $3a^2 - 3a + 3ab + b^2 = 0$. Let us call this set $D$. To determine these parameters one solves

$$
\begin{aligned}
f_n(x_1^k,\dots,x_n^k) &= \sum_{i=1}^{n}\left(a\frac{3i^2-3i+1}{n^3}+b\frac{2i-1}{n^2}+(1-a-b)\frac{1}{n}\right)x_{(i)}^k\\
&= a\sum_{i=1}^{n}\frac{3i^2-3i+1-n^2}{n^3}x_{(i)}^k\\
&+ b\sum_{i=1}^{n}\frac{2i-1-n}{n_k^2}x_{(i)}^k+\sum_{i=1}^{n}\frac{x_{(i)}}{n}\\
&= P_1^k a + P_2^k b + P_3^k \approx y^k,\ k=1,\dots,K,
\end{aligned}
\tag{26}
$$

with the values of $P_1^k$, $P_2^k$ and $P_3^k$ determined by the corresponding sums.

The domain of $a, b$ is too complicated to write down an explicit solution, but since it is convex, one can find $a, b$ using the following method. First solve the linearly constrained least squares problem (26), subject to

$$
\begin{aligned}
a+b &\le 1,\\
2a+b &\ge -1.
\end{aligned}
\tag{27}
$$

Notice that $D$ is a subset of the set determined by (27). If the solution of this LSEI problem $(\hat{a}, \hat{b}) \in D$, then we stop, otherwise the solution must be on the boundary of the right part of the ellipse determined by $3a^2 - 3a + 3ab + b^2 = 0$, $a > 1$. The Lagrangian is therefore

$$L(a,b,\lambda) = \sum_{k=1}^{K} \left( P_1^k a + P_2^k b + P_3^k - y^k \right)^2 + \lambda(3a^2 - 3a + 3ab + b^2). \quad (28)$$

Minimization of the convex quadratic function (28) yields the optimal solution.

## 8 Conclusion

We have summarized a number of methods for identification of weights in aggregation operators from empirical data. We considered various classes of aggregation operators: weighted means, OWA, Choquet integral-based aggregation, and also their generalizations – quasi-arithmetic means, GOWA and WOWA operators. In all cases it was possible to formulate weights identification problems as classical quadratic or linear programming problems, and apply very efficient QP and LP algorithms. The methods presented in this chapter have been implemented in `AOTool` software package available from `http://www.deakin.edu.au/~gleb`.

**Acknowledgment** The work of the first author was supported by the projects MTM2004-3175, BFM2003-05308 and TIC2002-11942-E from Ministerio de Educación y Ciencia, Spain, PRIB-2004-9250, Govern de les Illes Balears, and TIN2004-20503-E, Red Temática Nacional sobre Sistemas de Acceso a la Información en la Web basados en Soft-Computing. She is a member of Fundamentos Matemáticos para el Soft-Computing research group.

## References

1. J. Aczel. On mean values. *Bulletin of the American Math. Society*, 54, 1948.
2. I. Barrodale and F.D.K. Roberts. Algorithm 552: Solution of the constrained $l_1$ linear approximation problem. *ACM Trans. Math. Software*, 6: 231–235, 1980.
3. G. Beliakov. Shape preserving approximation using least squares splines. *Approximation theory and applications*, 16: 80–98, 2000.
4. G. Beliakov. How to build aggregation operators from data? *Int. J. Intelligent Systems*, 18: 903–923, 2003.
5. G. Beliakov. The cutting angle method - a tool for constrained global optimization. *Optimization Methods and Software*, 19: 137–151, 2004.
6. G. Beliakov. Fitting triangular norms to empirical data. In P.E. Klement and R. Mesiar, editors, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*, pp. 255–265. Elsevier, New York, 2005.
7. G. Beliakov. Learning weights in the generalized OWA operators. *Fuzzy Optimization and Decision Making*, 4: 119–130, 2005.
8. G. Beliakov, T. Calvo, M.E. Garcia Barriocanal, and M.A. Sicilia. Choquet integral-based aggregation of interface usability parameters: Identification of fuzzy measure. In A. Rubinov, editor, *6th Intl. Conf. on Optimization Techniques and Applications*, Ballarat, Australia, 2004.
9. G. Beliakov, R. Mesiar, and L. Valaskova. Fitting generated aggregation operators to empirical data. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12: 219–236, 2004.

10. P. Benvenuti and R. Mesiar. Integrals with respect to a general fuzzy measure. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals. Theory and Applications.*, pp. 205–232. Physica-Verlag, Heidelberg, 2000.
11. T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Aggregation Operators. New Trends and Applications*, pp. 3–104. Physica-Verlag, Heidelberg, New York, 2002.
12. T.-Y. Chen and J.-C. Wang. Identification of [lambda]-fuzzy measures using sampling design and genetic algorithms. *Fuzzy Sets and Systems*, 123: 321–341, 2001.
13. T.-Y. Chen, J.-C. Wang, and G.H. Tzeng. Identification of general fuzzy measures by genetic algorithms based on partial information. *IEEE Trans. on Syst., Man and Cybernetics, B*, 30: 517 – 528, 2000.
14. D. Dubois and H. Prade. A review of fuzzy set aggregation connectives. *Information Sciences*, 36: 85–121, 1985.
15. D. Dubois and H. Prade. On the use of aggregation operations in information fusion processes. *Fuzzy Sets and Systems*, 142: 143–161, 2004.
16. H. Dyckhoff and W. Pedrycz. Generalized means as model of compensative connectives. *Fuzzy Sets and Systems*, 14: 143–154, 1984.
17. D. Filev and R. Yager. On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*, 94: 157–169, 1998.
18. R. Fuller and P. Majlender. An analytic approach for obtaining maximal entropy OWA operator weights. *Fuzzy Sets and Systems*, 124: 53–57, 2001.
19. E. M. Gertz and S. J. Wright. Object-oriented software for quadratic programming. *ACM Trans. Math. Software*, 29:58–81, 2003.
20. M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92: 167–189, 1997.
21. M. Grabisch. The interaction and Mobius representation of fuzzy measures on finite spaces, k-additive measures: a survey. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals. Theory and Applications.*, pp. 70–93. Physica-Verlag, Heidelberg, 2000.
22. M. Grabisch, H.T. Nguyen, and E.A. Walker. *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference*. Kluwer, Dordrecht, 1995.
23. R.J. Hanson and K.H. Haskell. Algorithm 587. Two algorithms for the linearly constrained least squares problem. *ACM Trans. Math. Software*, 8: 323–333, 1982.
24. K.H. Haskell and R. Hanson. An algorithm for linear least squares problems with equality and nonnegativity constraints. *Math. Progr.*, 21: 98–118, 1981.
25. K. Ishii and M. Sugeno. A model of human evaluation process using fuzzy measure. *International Journal of Man-Machine Studies*, 22:19–38, 1985.
26. C. Lawson and R. Hanson. *Solving Least Squares Problems*. SIAM, Philadelphia, 1995.
27. K.M. Lee and H. Leekwang. Identification of λ-fuzzy measure by genetic algorithms. *Fuzzy Sets and Systems*, 75: 301–309, 1995.
28. R. Mesiar, M. Komornikova, T. Calvo, and A. Kolesarova. A review of aggregation functions. In H. Bustince, F. Herrera, and J. Montero, editors, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, page in this volume. Springer, Heidelberg, 2007.
29. R. Mesiar and L. Valaskova. Universal fuzzy measures. In *Proc. IFSA Conf.*, pp. 139–142, Istanbul, 2003.
30. P. Miranda and M. Grabisch. Characterizing k-additive fuzzy measures. In B. Bouchon-Meunier, J. Gutierrez-Rios, L. Magdalena, and R. Yager, editors, *Technologies for Constructing Intelligent Systems*, Vol 2, pp. 209–222. Physica-Verlag, Heidelberg, 2002.
31. M.O. O'Hagan. Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. In *22nd Annual IEEE Asilomar Conf. Signals, Systems, Computers*, pp. 681–689. Pacific Grove, CA, 1988.
32. S.A. Pijavski. An algorithm for finding the absolute extremum of a function. *USSR Comput. Math. and Math. Phys.*, 2: 57–67, 1972.

33. M. Sugeno. *Theory of Fuzzy Integrals and Applications*. Phd thesis, Tokyo Inst. of Technology, 1974.
34. V. Torra. The weighted OWA operator. *Int. J. Intelligent Systems*, 12: 153–166, 1997.
35. V. Torra. On some relationships between WOWA operator and the Choquet integral. In *IPMU 1998*, pp. 818–824, Paris, 1998.
36. V. Torra. Learning weights for the quasi-weighted means. *IEEE Trans. on Fuzzy Syst.*, 10: 653–666, 2002.
37. V. Torra. OWA operators in data modeling and reidentification. *IEEE Trans. on Fuzzy Syst.*, 12: 652–660, 2004.
38. W. Wang, Z. Wang, and G.J. Klir. Genetic algorithms for determining fuzzy measures from data. *J. Intell. Fuzzy Systems*, 6: 171–183, 1998.
39. G.A. Watson. Approximation in normed linear spaces. *J. Comp. Appl. Math.*, 121: 1–36, 2000.
40. Z.S. Xu. An overview of methods for determining OWA weights. *Int. J. Intelligent Systems*, 20:843–865, 2005.
41. R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18: 183–190, 1988.
42. R. Yager. Families of OWA operators. *Fuzzy Sets and Systems*, 59:125–148, 1993.
43. R. Yager. Generalized OWA aggregation operators. *Fuzzy Optimization and Decision Making*, 3:93–107, 2004.
44. R. Yager and D. Filev. *Essentials of Fuzzy Modelling and Control*. J. Wiley & Sons, New York, 1994.
45. R. Yager and J. Kacprzyk, editors. *The Ordered Weighted Averaging Operators. Theory and Applications.* Kluwer, Boston, 1997.

# Linguistic Aggregation Operators: An Overview

**Zeshui Xu**

**Abstract** Linguistic aggregation operators are a powerful tool to aggregate linguistic information, which have been studied and applied in a wide variety of areas, including engineering, decision making, artificial intelligence, data mining, and soft computing. In this chapter, we provide a comprehensive survey of the existing main linguistic aggregation operators, and briefly discuss their characteristics and applications. Finally, we roughly classify all these linguistic aggregation operators and conclude with a discussion of some interesting further research directions.

## 1 Introduction

In the real-life world, there are many situations, such as selecting applications for different kinds of scholarships and selecting projects for different kinds of funding policies [31], and evaluating the "speed", "comfort" or "design" for different kinds of cars [3], in which the information cannot be assessed precisely in a quantitative form but may be in a qualitative one [13]. As was pointed out in [6], this may arise for different reasons: 1) the information may be unquantifiable due to its nature; and 2) the precise quantitative information may not be stated because either it is unavailable or the cost of its computation is too high and an "approximate value" may be tolerated. Thus, in such situations, a more realistic approach may be to use linguistic assessments instead of numerical values by means of linguistic variables, that is, variables whose values are not numbers but words or sentences in a natural or artificial language [62]. For example, when evaluating the "comfort" or "design" of a car, linguistic labels like "good", "fair", "poor" are usually be used, and evaluating a car's speed, linguistic labels like "very fast", "fast", "slow" can be used [3, 34].

When a problem is solved using linguistic information, it implies the need for computing with words [62]. Linguistic aggregation operators are a powerful tool to deal with this issue. Over the last decades, various linguistic aggregation operators have been proposed, including the linguistic max and min operators [53, 55, 56, 57, 58, 61], linguistic median operator [57, 58, 61], linguistic weighted median operator [57, 58, 61], linguistic max-min weighted averaging operator [56], extension principle-based linguistic aggregation operator [2, 4, 5, 7, 33], symbol-based linguistic aggregation operator [11], 2-tuple arithmetic mean

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*. © Springer 2008

operator [21], 2-tuple weighted averaging operator [21], 2-tuple OWA operator [21], linguistic weighted OWA operator [39], linguistic averaging operator [41, 44, 50], linguistic weighted disjunction operator [12], linguistic weighted conjunction operator [12], linguistic weighted averaging operator [43, 44, 50], ordinal OWA operator [3, 55, 56], ordinal linguistic aggregation operator [11], ordinal hybrid aggregation operator [43], linguistic OWA operator [9, 13, 16, 18, 19, 20, 40, 43, 47, 50], inverse-linguistic OWA operator [12, 20], linguistic hybrid aggregation operator [51], induced linguistic OWA operator [50], uncertain linguistic averaging operator [42, 50], uncertain linguistic weighted averaging operator [50], uncertain linguistic OWA operator [42, 50], induced uncertain linguistic OWA operator [49, 50], uncertain linguistic hybrid aggregation operator [42], etc. To date, these linguistic aggregation operators have been studied and applied in a wide variety of areas, such as engineering [35, 43], decision making [3, 6, 10, 11, 12, 13, 15, 16, 17, 18, 19, 22, 23, 24, 25, 29, 40, 41, 42, 43, 44, 45, 46, 48, 49, 51, 56], information retrieval [4, 8, 27, 28, 30], marketing [20, 60], scheduling [1], biotechnology [5], materials selection [7], software system [33], personnel management [20], educational grading system [32], supply chain management and maintenance service [43], etc.

In this chapter, we provide a comprehensive survey of linguistic aggregation operators. To do so, the remainder of this chapter is structured as follows. Section 2 reviews the existing main linguistic aggregation operators, and briefly discusses their characteristics and applications. In the concluding remarks, we roughly classify all these linguistic aggregation operators and finish with a discussion of some interesting further research directions.

## 2 A Review of Linguistic Aggregation Operators

In many real-world problems, the information about the satisfaction associated with an outcome and a state of nature may be at best expressed in terms of linguistic labels [4, 37, 38]. For convenience of description, let $S = \{s_0, s_1, \ldots, s_g\}$ be a finite and totally ordered discrete linguistic label set, whose cardinality value is odd, such as 7 and 9, where $g$ is a positive integer, $s_i$ represents a possible value for a linguistic variable, and it requires that [16]:

1) The set is ordered: $s_i \geq s_j$ iff $i \geq j$;
2) The negation operator is defined: neg $(s_i) = s_j$ such that $j = g - i$.

The cardinality of $S$ must be small enough so as not to impose useless precision on the experts and it must be rich enough in order to allow a discrimination of the performances of each object in a limited number of grades [3], the limit of cardinality is 11 or not more than 13. For example, a set of seven linguistic labels $S$ could be [13]:

$$S = \{s_0 = none, s_1 = very\ low, s_2 = low, s_3 = medium,\\ s_4 = high, s_5 = very\ high, s_6 = perfect\}$$

In many areas, such as engineering, decision making, data mining, artificial intelligence, medical diagnosis and fuzzy logic control, etc., linguistic information aggregation is an essential process, which has received great attention from researchers in the last decades. Various linguistic aggregation operators have also been developed. In the following, we give an overview of the existing main linguistic aggregation operators.

## *2.1 Linguistic Aggregation Operators Based on Linear Ordering*

Let $\{a_1, a_2, \ldots, a_n\}$ be a collection of linguistic arguments in $S$. In [55, 56, 57, 58, 61], Yager et al. introduced the linguistic max ($\text{LM}_1$) and linguistic min ($\text{LM}_2$) operators, respectively:

$$LM_1(a_1, a_2, \ldots, a_n) = \max_j\{a_j\}, \quad LM_2(a_1, a_2, \ldots, a_n) = \min_j\{a_j\} \tag{1}$$

Yager et al. [57, 58, 61] developed the linguistic median ($\text{LM}_3$) operator:

$$LM_3(a_1, a_2, \ldots, a_n) = \begin{cases} b_{(n+1)2} & if\ n\ is\ odd, \\ b_{n2} & if\ n\ is\ even. \end{cases} \tag{2}$$

where $b_j$ is the $j$th largest of the linguistic arguments in $\{a_1, a_2, \ldots, a_n\}$.

These three operators are the simplest linguistic aggregation operators, which are usually used to develop some other operators for aggregating linguistic information. Yager et al. [57, 58, 61] further considered the problem of weighted median aggregation as below:

Suppose that $\{(w_1, a_1), (w_2, a_2), \ldots, (w_n, a_n)\}$ is a collection of pairs, where $a_i$ is a linguistic label and $w_i$ is its associated weight, with the condition $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, and assume that the $a_i$ are reordered such that $b_j$ is the $j$th largest of the $a_i$. Then $\{(u_1, b_1), (u_2, b_2), \ldots, (u_n, b_n)\}$ is the ordered collection of $\{(w_1, a_1), (w_2, a_2), \ldots, (w_n, a_n)\}$, where $u_j$ is the weight that is associated with the $a_i$ that becomes $b_j$. For example, if $b_j = a_5$, then $u_j = w_5$. Let $T_i = \sum_{j=1}^{i} u_j$, if

$$LWM((u_1, b_1), (u_2, b_2), \ldots, (u_n, b_n)) = b_k \tag{3}$$

then *LWM* is called a linguistic weighted median (LWM) operator, where $k$ satisfies $T_{k-1} < 0.5$ and $T_k \geq 0.5$. The linguistic weighted median is the ordered value of the arguments for which the sum of the weights first crosses the value of 0.5, and $k$ is called a cross over value. Yager [58] showed that the LWM operator is of idempotency, commutativity and monotonicity.

Yager [54] introduced the ordered weighted averaging (OWA) operator to provide a method for aggregating several numerical inputs that lie between the max and min operators. The fundamental aspect of the OWA operator is the re-ordering step. Using an ordinal form of the OWA operator, Yager [55] defined the following:

**Definition 1.** [55] *A mapping* $OOWA : S^n \rightarrow S$ *is called an ordinal OWA (OOWA) operator of dimension n if it has an associated weighting vector* $w = (w_1, w_2, \ldots, w_n)$*, such that 1)* $w_j \in S$*; 2)*$w_j \geq w_i if j > i$*; and 3)* $\max_j\{w_j\} = s_g Furthermore,$.

$$OOWA(a_1, a_2, \ldots, a_n) = \max_j\{w_j \wedge b_j\} \tag{4}$$

*where* $b_j$ *is the* $j$ *th largest of the* $a_j$.

Especially, if $w = (s_g, s_g, \ldots, s_g)$, where $w_i = s_g$, for all $i$, then the OOWA operator is reduced to the $LM_1$ operator; if $w = (s_0, \ldots, s_0, s_g)$, where $w_i = s_0$, for all $i \neq n$, $w_n = s_g$, then the OOWA operator is reduced to the $LM_2$ operator; if $w = (s_\alpha, \ldots, s_\alpha, s_g)$, where $w_i = s_\alpha$, for all $i \neq n$, $w_n = s_g$, then the OOWA operator is reduced to the following:

$$OOWA(a_1, a_2, \ldots, a_n) = (s_\alpha \wedge \max_j\{a_j\}) \vee \min_j(a_j) \tag{5}$$

which is called a linguistic max-min weighted averaging (LMMA) operator. This operator is analogous to the Hurwicz type criteria used in the numeric case, $\beta \max_j(a_j) + (1 - \beta) \min_j(a_j)$, where $\beta \in [0,1]$.

The OOWA operators enjoys many useful properties. In particular, it is symmetric, idempotent and monotonic function [55, 56]. Ovchinnikov [36] investigated the invariance properties of the OOWA operator. Bordogna et al. [3] proposed a model based on the OOWA operator for group decision making in a linguistic context, and presented a human-consistent definition that can be useful in a consensus reaching phase and a procedure for its computation. Yager [56] developed a structure making use of the OOWA operator for the evaluation of decisions under ignorance in situations in which the payoff matrix contains linguistic information. In this structure, the decision maker's dispositional attitude is represented by the weighting vector. He also provided two measures for classifying the OOWA weighting vectors used in the aggregations. Yager et al. [60] developed an OOWA operator-based approach to aggregating market research data based on fuzzy reasoning. This approach uses operations from fuzzy logic to construct the models and the principal of minimal entropy to choose between competing models. They implemented this approach in a test case using economic attitudinal variables to predict the purchase of major consumer products.

Herrera and Herrera-Viedma [12] provided the linguistic weighted disjunction and conjunction operators.

**Definition 2.** [12] *Let* $\{a_1, a_2, \ldots, a_n\}$ *be a set of linguistic labels to be aggregated,* $c = (c_1, c_2, \ldots, c_n)$ *be the associated weighting vector,* $a_j, c_j \in S$*, for all j, then a linguistic weighted disjunction (LWD) operator is defined as:*

$$LWD((c_1, a_1), (c_2, a_2), \ldots, (c_n, a_n)) = \max_j \min\{c_j, a_j\} \tag{6}$$

**Definition 3.** [12] *Let* $\{a_1, a_2, \ldots, a_n\}$ *be a set of linguistic labels to be aggregated,* $c = (c_1, c_2, \ldots, c_n)$ *be the associated weighting vector,* $a_j, c_j \in S$, *for all j, then a linguistic weighted conjunction (LWC) operator is defined as:*

$$LWC((c_1, a_1), (c_2, a_2), \ldots, (c_n, a_n)) = \min_j \max\{neg(c_j), a_j\} \tag{7}$$

Herrera-Viedma [27] applied the LWD and LWC operators to develop a linguistic model for an information retrieval system (IRS), and presented an ordinal fuzzy linguistic approach and studied its use for modeling the imprecision and subjectivity that appear in the user-IRS interaction.

Xu [43] developed an ordinal hybrid aggregation operator that reflects the importance degrees of both the linguistic argument and its ordered position:

**Definition 4.** [43] *An ordinal hybrid aggregation (OHA) operator is mapping* $OHA : S^n \to S$, *which has associated with it a linguistic weighting vector* $w = (w_1, w_2, \ldots, w_n)$, *with* $w_j \in S$, *such that*

$$OHA(a_1, a_2, \ldots, a_n) = \max_j \min\{w_j, b_j\} \tag{8}$$

*where* $b_j$ *is the* $j$*th largest of the linguistic weighted arguments* $\bar{a}_i$ ($\bar{a}_i = \min\{\omega_i, a_i\}$, $i = 1, 2, \ldots, n$), $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ *is the linguistic weight vector of the* $a_i$, *with* $\omega_i \in S$.

Especially, if $w = (s_g, s_g, \ldots, s_g)$, where $w_i = s_g$, for all $i$, then

$$OHA(a_1, a_2, \ldots, a_n) = \max_j \min\{w_j, b_j\} = \max_i\{\bar{a}_i\} = \max_i \min\{\omega_i, a_i\} \tag{9}$$

thus, in this case, the OHA operator is reduced to the LWD operator.

If $\omega = (s_g, s_g, \ldots, s_g)$, where $\omega_i = s_g$, for all $i$, then

$$\bar{a}_i = \min\{\omega_i, a_i\} = a_i, i = 1, 2, \ldots, n \tag{10}$$

in this case, the OHA operator is reduced to the OOWA operator.

## 2.2 Linguistic Aggregation Operators Based on the Extension Principle and Symbols

In [2, 4, 5, 7, 10, 33], Bonissone et al. studied the linguistic aggregation operators based on the extension principle. Herrera and Martínez [21] described an extension principle-based linguistic aggregation operator as follows:

**Definition 5.** [21] *An extension principle-based linguistic aggregation operator acts according to*

$$S^n \overset{\tilde{F}}{\rightarrow} F(R) \overset{app_1(\cdot)}{\rightarrow} S \tag{11}$$

*where $S^n$ symbolizes the n Cartesian product of $S$, $\tilde{F}$ is an aggregation operator based on the extension principle, $F(R)$ is the set of fuzzy sets over the set of real numbers $R$, $app_1 : F(R) \rightarrow S$ is a linguistic approximation function that returns a label from $S$ whose meaning is the closest to the obtained unlabeled fuzzy number and $S$ is the initial linguistic label set.*

Chang and Chen [5] gave a decision algorithm based on the extension principle-based linguistic aggregation operator to solve the technology transfer strategy selection problem in the area of biotechnology management. The algorithm is based on the concepts of fuzzy set theory and the hierarchical structure analysis. The linguistic variables and fuzzy numbers were used to aggregate the decision makers' subjective assessments about criteria weightings and appropriateness of alternative transfer strategies versus selection criteria to obtain the final scores called fuzzy appropriateness indices. Chen [7] utilized the extension principle-based linguistic aggregation operator to solve the tool steel materials selection problem under fuzzy environment where the importance weights of different criteria and the ratings of various alternatives under different criteria are assessed in linguistic terms represented by fuzzy numbers. Lee [33] used the extension principle-based linguistic aggregation operator to build a structure model of risk in software development and proposed two algorithms to tackle the rate of aggregative risk in a fuzzy environment by fuzzy sets theory during any phase of the life cycle. Adamopoulos and Pappis [1] proposed a fuzzy approach by using the extension principle-based linguistic aggregation operator to solve a single machine scheduling problem. In this approach, the system's variables are defined using linguistic terms. Each of these variables may take values described via fuzzy triangular numbers. Law [32] built a structure model of a fuzzy educational grading system with the extension principle-based linguistic aggregation operator, and proposed an algorithm with it. They also discussed a method to build the membership functions of several linguistic values with different weights.

Delgado et al. [11] developed a symbol-based linguistic aggregation operator that makes computations on the indexes of the linguistic labels, which can be interpreted as [21, 24]:

$$S^n \overset{C}{\rightarrow} [0, g] \overset{app_2(\cdot)}{\rightarrow} \{0, \ldots, g\} \rightarrow S \tag{12}$$

where $C$ is a symbol-based linguistic aggregation operation, and $app_2(\cdot)$ is an approximation function used to obtain an index $\{0, \ldots, g\}$ associated to a linguistic label in $S = \{s_0, s_1, \ldots, s_g\}$ from a value in $[0, g]$. For example, $C$ is the convex combination of linguistic labels defined in [11], and $app_2(\cdot)$ is the usual round operation.

Herrera and Verdegay [26] presented a linguistic OWA operator, which is based on the OWA operator [54], and on the convex combination of linguistic labels [11].

**Definition 6.** [26] *Let $A = \{a_1, a_2, \ldots, a_n\}$ be a set of linguistic labels to be aggregated, then a linguistic OWA ($LOWA_1$) operator is defined as:*

$$\begin{aligned} LOWA_1(a_1, a_2, \ldots, a_n) &= w \cdot B^T = C^n\{w_k, b_k, k = 1, 2, \ldots, n\} \\ &= w_1 \otimes b_1 \oplus (1 - w_1) \otimes C^{n-1}\{\gamma_h, b_h, h = 2, \ldots, n\} \end{aligned} \tag{13}$$

*where* $w = (w_1, w_2, \ldots, w_n)$ *is a weighting vector, such that* $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, $\gamma_h = w_h \Big/ \sum_{k=2}^{n} w_k$, $h = 2, \ldots, n$, *and* $B = (b_1, b_2, \ldots, b_m)$ *is a vector associated to* $A$, *such that*

$$B = \sigma(A) = (a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}) \tag{14}$$

*where* $a_{\sigma(j)} \leq a_{\sigma(i)}$, *for all* $i \leq j$, *with* $\sigma$ *being a permutation over the set of linguistic labels* $A$.$C^n$ *is the convex combination operator of* $n$ *linguistic labels and if* $n = 2$, *then it is defined as:*

$$C^2\{w_i, b_i, i = 1, 2\} = w_1 \otimes s_j \oplus (1 - w_1) \otimes s_i = s_k, s_j, s_i \in S(j \geq i) \tag{15}$$

*such that* $k = \min\{g, i + round(w_1 \cdot (j - i))\}$, *where round is the usual round operation, and* $b_1 = s_j, b_2 = s_i$. *If* $w_j = 1$ *and* $w_i = 0$ *with* $i \neq j$, *for all* $i$, *then the convex combination is defined as:*

$$C^n\{w_i, b_i, i = 1, 2, \ldots, n\} = b_j \tag{16}$$

Many approaches have been developed for determining the weighting vector $w = (w_1, w_2, \ldots, w_n)$, see Xu [47] for a detailed overview on this field.

Herrera et al. [9, 13, 16, 17, 18, 19] utilized the $\text{LOWA}_1$ operator to develop various approaches to group decision making under linguistic assessments, and presented some consensus models for the consensus reaching process in heterogeneous linguistic group decision making problems. Herrera et al. [20] established a linguistic decision model based on the $\text{LOWA}_1$ operator for promotion mix management solved with genetic algorithms, and demonstrated the usefulness of the model by solving a real problem from the business world. Herrera-Viedma and Peis [30] presented a fuzzy linguistic evaluation method based on the $\text{LOWA}_1$ operator to characterize the information contained in SGML (standard generalized markup language)-based documents. The method generates linguistic recommendations from linguistic evaluation judgments provided by different recommenders on meaningful elements of document type definition (DTD).

Herrera and Herrera-Viedma [12] presented an extension of the $\text{LOWA}_1$ operator, called the inverse-$\text{LOWA}_1$ operator:

**Definition 7.** [12] *An inverse-*$LOWA_1$ *operator is a type of linguistic OWA operator, in which* $B = \sigma^I(A) = \{a_{\sigma(1)}, a_{\sigma(2)} \ldots, a_{\sigma(n)}\}$, *where* $a_{\sigma(i)} \leq a_{\sigma(j)}$, *for all* $i \leq j$. *If* $m = 2$, *then it is defined as:*

$$C^2\{w_i, b_i, i = 1, 2\} = w_1 \otimes s_j \oplus (1 - w_1) \otimes s_i = s_k, s_j, s_i \in S(j \leq i) \tag{17}$$

*such that* $k = \min\{g, i + round(w_1 \cdot (j - i))\}$.

Based on the LOWA$_1$ and inverse-LOWA$_1$ operators, Herrera and Herrera-Viedma [12] defined a linguistic weighted averaging operator using the concept of fuzzy majority represented by fuzzy linguistic quantifiers and two families of linguistic connectives. Let us first review the following two families of linguistic connectives [12]:

1) Linguistic conjunction operators $LC^{\rightarrow}$:

   i) The classical linguistic min operator

$$LC_1^{\rightarrow}(c, a) = \min(c, a) \tag{18}$$

   ii) The nipotent linguistic min operator:

$$LC_2^{\rightarrow}(c, a) = \begin{cases} \min(c, a) & if\ c > neg(a), \\ 0 & otherwise. \end{cases} \tag{19}$$

   iii) The weakest linguistic conjunction:

$$LC_3^{\rightarrow}(c, a) = \begin{cases} \min(c, a) & if \max(c, a) = s_g, \\ 0 & otherwise. \end{cases} \tag{20}$$

2) Linguistic implication operators $LI^{\rightarrow}$:

   i) Kleene-Dienes's linguistic implication operator:

$$LI_1^{\rightarrow}(c, a) = \max(neg(c), a) \tag{21}$$

   ii) Gödel's linguistic implication operator:

$$LI_2^{\rightarrow}(c, a) = \begin{cases} s_g & if\ c \leq a, \\ a & otherwise. \end{cases} \tag{22}$$

   iii) Fodor's linguistic implication operator:

$$LI_3^{\rightarrow}(c, a) = \begin{cases} s_g & if\ c \leq a, \\ \max(neg(c), a) & otherwise. \end{cases} \tag{23}$$

**Definition 8.** [12] *Let $\{a_1, a_2, \ldots, a_n\}$ be a set of linguistic labels to be aggregated, $c = (c_1, c_2, \ldots, c_n)$ be the associated weighting vector, $a_j, c_j \in S$, for all $j$, then a linguistic weighted averaging (LWA$_1$) operator is defined as:*

$$LWA_1((c_1, a_1), (c_2, a_2), \ldots, (c_n, a_n)) = f(z(c_1, a_1), g(c_2, a_2), \ldots, (c_n, a_n)) \tag{24}$$

*where $f$ is a $L\,OWA_1$ operator (or inverse-LOWA$_1$ operator) and h is an importance transaction function, such that $z \in LC^{\rightarrow}$, if $f$ is a LOWA$_1$ operator; and $z \in LI^{\rightarrow}$,* *if $LWA_1$ is an inverse-LOWA$_1$ operator, where $LC^{\rightarrow} = \{LC_1^{\rightarrow}, LC_2^{\rightarrow}, \quad LC_3^{\rightarrow}\}$ and $LI^{\rightarrow} = \{LI_1^{\rightarrow}, LI_2^{\rightarrow}, LI_3^{\rightarrow}\}$.*

Torra [38] defined a weighted OWA operator that combines the advantages of the OWA operator and the ones of the weighted mean. In order to aggregate linguistic information, he extended the operator to linguistic environments and defined a linguistic weighted OWA operator:

**Definition 9.** [38] *Let $\{a_1, a_2, \ldots, a_n\}$ be a set of linguistic labels to be aggregated, $p = (p_1, p_2, \ldots, p_n)$ and $w = (w_1, w_2, \ldots, w_n)$ be the weighting vectors of dimension n, such that $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$; $p_i \in [0, 1]$, $\sum_{i=1}^{n} p_i = 1$, in this case, a mapping LWOWA:$S^n \to S$ is called a linguistic weighted OWA (LWOWA) operator of dimension n, if*

$$
\begin{aligned}
LWOWA(a_1, a_2, \ldots, a_n) &= C^n\{\omega_k, b_k, k = 1, 2, \ldots, n\} \\
&= \omega_1 \otimes b_1 \oplus (1 - \omega_1) \otimes C^{n-1}\{\omega_h \Big/ \sum_{k=2}^{n} \omega_k, b_h, h = 2, \ldots, n\}
\end{aligned} \tag{25}
$$

*where $(b_1, b_2, \ldots, b_m) = (a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)})$ with $\sigma$ being a permutation over the set of linguistic labels $\{a_1, a_2, \ldots, a_n\}$ such that $a_{\sigma(j)} \le a_{\sigma(i)}$, for all $i \le j$. The convex combination C is defined according to Definition 6, and the weight $\omega_i$ is defined as $\omega_i = w^*(\sum_{j \le i} p_{\sigma(j)}) - w^*(\sum_{j<i} p_{\sigma(j)})$, where $w^*$ is a monotone increasing function that interpolates the points $(i/n, \sum_{j \le i} w_j)$ together with the point $(0, 0)$, $w^*$ is required to straight line when the points can be interpolated in this way.*

## *2.3 Linguistic Aggregation Operators Based on Linguistic 2-tuples*

An important limitation for the extension principle-based linguistic aggregation operator and symbol-based linguistic aggregation operator appears because the computational techniques used present a common drawback, the "loss of information", that implies a lack of precision in the final results [21]. To overcome this drawback, Herrera and Martínez [21] developed a fuzzy linguistic representation model, which represents the linguistic information with a pair of values called 2-tuple, $(s, \alpha)$, where $s$ is a linguistic label and $\alpha$ is a numerical value that represents the value of the symbolic translation.

**Definition 10.** [21] *Let $\beta$ be the result of an aggregation of the indexes of a set of labels assessed in the linguistic label set $S = \{s_0, s_1, \ldots, s_g\}$, i.e., the result of a symbolic aggregation operation. $\beta \in [0, g]$, being $g + 1$ the cardinality of S. Let $i = round(\beta)$ and $\alpha = \beta - i$ be two values, such that, $i \in [0, g]$ and $\alpha \in [-0.5, 0.5)$, then $\alpha$ is called a symbolic translation.*

Herrera and Martínez [22] pointed out that the symbolic translation of a linguistic label, $s_i$, is a numerical value assessed in $[-0.5, 0.5)$ that supports the "difference of information" between a counting of information $\beta \in [0, g]$ obtained after a symbolic aggregation operation and the closest value in $\{0, \ldots, g\}$ that indicates the index of the closest linguistic label in $S(i = round(\beta))$. The equivalent 2-tuple representation of linguistic label $s_i$ is obtained by means of the function $\theta$ as:

$$\theta : S \to (S \times [-0.5, 0.5)), \quad \theta(s_i) = (s_i, 0)\big/ s_i \in S \tag{26}$$

The 2-tuple $(s_i, \alpha)$ that expresses the equivalent information to $\beta$ is obtained with the following function:

$$\Delta : [0, g] \to S \times [-0.5, 0.5), \quad \Delta(\beta) = \begin{cases} s_i & i = round(\beta), \\ \alpha = \beta - i & \alpha \in [-0.5, 0.5). \end{cases} \tag{27}$$

with round is the usual round operation, $s_i$ has the closest index label to "$\beta$" and "$\alpha$" is the value of the symbolic translation. There is always a function $\Delta^{-1}$, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g]$, that is [22], $\Delta^{-1} : S \times [-0.5, 0.5) \to [0, g]$ and $\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$.

Herrera and Martínez [21] further introduced some 2-tupe aggregation operators as follows:

1) 2-tuple negation operator. The negation operator over a 2-tuple can be defined as: $neg(s_i, \alpha) = \Delta(g - (\Delta^{-1}(s_i, \alpha)))$.

2) 2-tuple weighted averaging operator. Let $\{(r_1, \alpha_1), (r_2, \alpha_2), \ldots, (r_n, \alpha_n)\}$ be a set of 2-tuples and $w = (w_1, w_2, \ldots, w_n)$ be their weight vector, $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, then a 2-tuple weighted averaging (TWA) operator is defined as:

$$TWA((r_1, \alpha_1), (r_2, \alpha_2), \ldots, (r_n, \alpha_n)) = \Delta(\sum_{i=1}^{n} w_i \cdot \Delta^{-1}(r_i, \alpha_i)) = \Delta(\sum_{i=1}^{n} w_i \beta_i) \tag{28}$$

Especially, if $w = (1/n, 1/n, \ldots, 1/n)$, then the TWA operator is reduced to a 2-tuple arithmetic mean (TAM) operator.

3) 2-tuple OWA operator. Let $\{(r_1, \alpha_1), (r_2, \alpha_2), \ldots, (r_n, \alpha_n)\}$ be a set of 2-tuples and $w = (w_1, w_2, \ldots, w_n)$ be an associated weighting vector that satisfies $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, then 2-tuple OWA (TOWA) operator is given as:

$$TOWA((r_1, \alpha_1), (r_2, \alpha_2), \ldots, (r_n, \alpha_n)) = \Delta(\sum_{j=1}^{n} w_j \beta_j^*) \tag{29}$$

where $\beta_j^*$ is the $j$th largest of the $\beta_i$ values.

Herrera and Martínez [24] utilized the 2-tuple arithmetic mean operator to develop a computational model to identify an appropriate transfer strategy in biotechnology, this application showed that the 2-tuple linguistic computational model is a more adequate model for dealing with linguistic information in processes of computing with words than classical ones. Martínez et al. [35] utilized the 2-tuple linguistic weighted average operator to propose a linguistic evaluation process based on multi-expert multi-criteria decision model that will be able to deal with multi-granular linguistic information without loss of information in order to evaluate different design options for an engineering system according to safety and cost criteria. Herrera and Martínez [23] developed a linguistic decision model using the 2-tuple arithmetic mean operator to deal with multigranular linguistic contexts and applied it to a multi-expert decision making problem. Delgado et al. [8] presented a distributed intelligent agent model where the communication of the evaluation of the retrieved information among the agents is carried out by using the 2-tuple arithmetic mean operator and 2-tuple weighted averaging operator as a way to endow the retrieval process with a high flexibility, uniformity and precision.

### 2.4 Linguistic Aggregation Operators Computing with Words Directly

In [43, 46], we defined the linguistic label set $S$ by another form $S' = \{s_\alpha | \alpha = -t, \ldots, t\}$, where $t$ is a positive integer, and $s_\alpha$ has the following characteristics: 1) $s_\alpha < s_\beta$ iff $\alpha < \beta$; and 2) There is the negation operator: $\mathrm{neg}(s_\alpha) = s_{-\alpha}$, especially, $\mathrm{neg}(s_0) = s_0$.

Obviously, the mid linguistic label $s_0$ represents an assessment of "indifference", and with the rest of the linguistic labels being placed symmetrically around it. Xu [46] extended the discrete linguistic label set $S$ to a continuous label set $\bar{S} = \{s_\alpha | \alpha \in [-q, q]\}$ in order to preserve all the given information, where $q (q > t)$ is a sufficiently large positive integer. If $s_\alpha \in S$, then $s_\alpha$ is termed an original linguistic label, otherwise, $s_\alpha$ is termed a virtual linguistic label. In general, the virtual linguistic labels can only appear in calculations.

**Definition 11.** [43] *Let* $s_\alpha, s_\beta \in \bar{S}, \lambda \in [0, 1]$*, then their operational laws can be defined as follows: 1)* $s_\alpha \oplus s_\beta = s_{\alpha+\beta}$*; and 2)* $\lambda s_\alpha = s_{\lambda\alpha}$*.*

Comparing to the commonly used linguistic label set $S = \{s_0, s_1, \ldots, s_g\}$ presented earlier in this paper, the representation of $S' = \{s_\alpha | \alpha = -t, \ldots, t\}$ has its own advantages. For example, we consider two linguistic labels $s_{-2} = poor$ and $s_2 = good$ in the linguistic label set:

$$\begin{aligned} S' = \{&s_{-4} = extremely\ poor, s_{-3} = very\ poor, s_{-2} = poor, \\ &s_{-1} = slightly\ poor, s_0 = fair, s_1 = slightly\ good, \\ &s_2 = good, s_3 = very\ good, s_4 = extremely\ good\} \end{aligned}$$

By the operational laws in Definition 11, we have $s_{-2} \oplus s_2 = s_0$, where $s_0 = fair$. However, if we consider two linguistic labels $s_2 = low$ and $s_4 = high$ in the linguistic label set:

$$\begin{aligned} S = \{&s_0 = none, s_1 = very\ low, s_2 = low, s_3 = medium, \\ &s_4 = high, s_5 = very\ high, s_6 = perfect\} \end{aligned}$$

then we have $s_2 \oplus s_4 = s_6$, where $s_6 = perfect$. Obviously, the representation of the linguistic label set $S'$ should be more in accord with actual situations than $S$ when using the operational laws given in Definition 11.

Based on Definition 11, Xu [41, 43, 44, 50, 51] developed various linguistic aggregation operators, which compute with words directly:

**Definition 12.** [50] *Let* $LWA_2 : \bar{S}^n \to \bar{S}$. *If*

$$LWA_2(s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_n}) = w_1 s_{\alpha_1} \oplus w_2 s_{\alpha_2} \oplus \cdots \oplus w_n s_{\alpha_n} = s_{\dot{\alpha}} \tag{30}$$

*where* $\dot{\alpha} = \sum_{j=1}^{n} w_j \alpha_j$, $w = (w_1, w_2, \ldots, w_n)$ *is the weighting vector of the linguistic label* $s_{\alpha_i}$, *and* $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, *then* $LWA_2$ *is called a linguistic weighted averaging (LWA*$_2$*) operator. The LWA*$_2$ *operator is an extension of the weighted averaging (WA) operator [53].*

Especially, if $w = (1/n, 1/n, \ldots, 1/n)$, then the LWA$_2$ operator is reduced to a linguistic averaging (LA) operator.

The fundamental aspect of the LWA$_2$ operator is that it computes the aggregated linguistic labels taking into account the importance of the sources of information.

**Definition 13.** [50] *Let* $LOWA_2 : (\bar{S}')^n \to \bar{S}'$, *which has associated with it a weighting vector* $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, *with* $\omega_i \in [0, 1]$, $\sum_{j=1}^{n} \omega_j = 1$, *such that*

$$LOWA_2(s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_n}) = \omega_1 s_{\beta_1} \oplus \omega_2 s_{\beta_2} \oplus \cdots \oplus \omega_n s_{\beta_n} = s_{\bar{\beta}} \tag{31}$$

*where* $\bar{\beta} = \sum_{j=1}^{n} w_j \beta_j$, $s_{\beta_j}$ *is the* $j$*th largest of the* $s_{\alpha_j}$, *then* $LOWA_2$ *is called a linguistic OWA (LOWA*$_2$*) operator.*

Especially, if $\omega = (1, 0, \ldots, 0)$, where $\omega_1 = 1, \omega_i = 0, i = 2, \ldots, n$, then the LOWA$_2$ operator is reduced to the LM$_1$ operator; if $w = (0, \ldots, 0, 1)$, where $\omega_i = 0, i = 2, \ldots, n-1, \omega_n = 1$, then the LOWA$_2$ operator is reduced to the LM$_2$ operator; if $\omega = (1/n, 1/n, \ldots, 1/n)$, then the LOWA$_2$ operator is reduced to the LA operator. The fundamental aspect of the LOWA$_2$ operator is the reordering of the linguistic arguments to be aggregated, based on their values.

The LOWA$_2$ operator is an extension of the OWA operator [54], and has some desirable properties similar to the OWA operator, such as monotonicity, idempotency, commutativity, and lies between the linguistic min and max operators. Xu [43] developed a LOWA$_2$ operator-based approach to multiple attribute decision making under linguistic environment and applied it to the firepower schedule selection in the defensive battle.

From Definitions 12 and 13, we know that the LWA$_2$ operator weights the linguistic argument, while the LOWA$_2$ operator weights the ordered position of the linguistic argument instead of weighting the argument itself. Therefore, the weights represent different aspects in both the LWA$_2$ and LOWA$_2$ operators. To overcome this drawback, Xu [51] proposed a linguistic hybrid aggregation operator.

**Definition 14.** [51] *A linguistic hybrid aggregation (LHA) operator is a mapping* $LHA : \bar{S}^n \to \bar{S}$, *which has an associated vector* $w = (w_1, w_2, \ldots, w_n)$ *with*

$w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, *such that*

$$LHA(s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_n}) = w_1 s_{\beta_1} \oplus w_2 s_{\beta_2} \oplus \cdots \oplus w_n s_{\beta_n} \tag{32}$$

*where* $s_{\beta_j}$ *is the* $j$*th largest of the linguistic weighted arguments* $\bar{s}_{\alpha_i}$ $(\bar{s}_{\alpha_i} = n\omega_i s_{\alpha_i}$, $i = 1, 2, \ldots, n)$, $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ *is the weight vector of the* $s_{\alpha_i}$ $(i = 1, 2, \ldots, n)$ *with* $\omega_i \in [0, 1]$, $\sum_{i=1}^{n} \omega_i = 1$, *and* $n$ *is the balancing coefficient, which plays a role of balance (in such a case, if the vector* $(\omega_1, \omega_2, \ldots, \omega_n)$ *approaches* $(1/n, 1/n, \ldots, 1/n)$, *then the vector* $(n\omega_1 s_{\alpha_1}, n\omega_2 s_{\alpha_2}, \ldots, n\omega_n s_{\alpha_n})$ *approaches* $(s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_n})$).

Especially, if $w = (1/n, 1/n, \ldots, 1/n)$, then the LHA operator is reduced to the $\text{LWA}_2$ operator; if $\omega = (1/n, 1/n, \ldots, 1/n)$, then the LHA operator is reduced to the $\text{LOWA}_2$ operator.

It is clear that the LHA operator generalizes both the $\text{LWA}_2$ and $\text{LOWA}_2$ operators, and reflects the importance degrees of both the given linguistic argument and its ordered position. Xu [51] proposed a practical approach, based on the linguistic hybrid aggregation operator, to multiple attribute group decision making with linguistic information, and applied the developed approach to the problem of evaluating university faculty for tenure and promotion.

In [59], Yager and Filev introduced the concept of induced OWA (IOWA) operator, which takes as its argument pairs, called OWA pairs, in which one component is used to induce an ordering over the second components which are then aggregated. The IOWA operators, however, can only be used in situations in which the aggregated arguments are the exact numerical values. Xu [50] introduced an induced linguistic OWA (ILOWA) operator, which can be used to aggregate linguistic arguments and is defined as follows:

$$ILOWA\left(\langle u_1, s_{\alpha_1}\rangle, \langle u_2, s_{\alpha_2}\rangle \ldots, \langle u_n, s_{\alpha_n}\rangle\right) = w_1 s_{\gamma_1} \oplus w_2 s_{\gamma_2} \oplus \cdots \oplus w_n s_{\gamma_n} = s_{\bar{\gamma}} \tag{33}$$

where $\bar{\gamma} = \sum_{i=1}^{n} w_i \gamma_i$, $w = (w_1, w_2, \ldots, w_n)$ is a weighting vector, such that $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, $s_{\gamma_j}$ is the $s_{\alpha_i}$ value of the OWA pair $\langle u_i, s_{\alpha_i}\rangle$ having the $jth$ largest $u_i$, and $u_i$ in $\langle u_i, s_{\alpha_i}\rangle$ is referred to as the order inducing variable and $s_i$ as the linguistic label. Especially, if $w = (1/n, 1/n, \ldots, 1/n)$, then the ILOWA operator is reduced to the LA operator; if $u_i = s_{\alpha_i}$, for all $i$, then the LOWA operator is reduced to the $\text{LOWA}_2$ operator; if $u_i = No.i$, for all $i$, where $No.i$ is the ordered position of the $s_i$, then the ILOWA operator is reduced to the $\text{LWA}_2$ operator.

However, if there is a tie between $\langle u_i, s_{\alpha_i}\rangle$ and $\langle u_j, s_{\alpha_j}\rangle$ with respect to order inducing variables. In this case, we can follow the policy presented by Yager and Filev [59], that is, we replace the arguments of the tied objects by the average of the arguments of the tied objects, i.e., we replace the argument component of each of $\langle u_i, s_{\alpha_i}\rangle$ and $\langle u_j, s_{\alpha_j}\rangle$ by their average $(s_{\alpha_i} \oplus s_{\alpha_j})/2$. If $k$ items are tied, we replace these by $k$ replica's of their average.

Sometimes, however, the input linguistic arguments may not match any of the original linguistic labels, and they may be located between two of them. For example, when evaluating the "design" of a car, an expert may provide his/her opinion with "between 'fair' and 'good' ". To solve this issue, Xu [42, 43] introduced the concept of uncertain linguistic variable and defined some operational laws of uncertain linguistic variables.

Let $\tilde{s} = [s_\alpha, s_\beta]$, where $s_\alpha, s_\beta \in \bar{S}$, $s_\alpha$ and $s_\beta$ are the lower and upper limits, respectively, then $\tilde{s}$ is called an uncertain linguistic variable. For convenience, we let $\tilde{S}$ be the set of all the uncertain linguistic variables.

For any three uncertain linguistic variables $\tilde{s} = [s_\alpha, s_\beta], \tilde{s}_1 = [s_{\alpha_1}, s_{\beta_1}]$, and $\tilde{s}_2 = [s_{\alpha_2}, s_{\beta_2}] \in \tilde{S}$, their operational laws can be as follows:

1) $\tilde{s}_1 \oplus \tilde{s}_2 = [s_{\alpha_1}, s_{\beta_1}] \oplus [s_{\alpha_2}, s_{\beta_2}] = [s_{\alpha_1} \oplus s_{\alpha_2}, s_{\beta_1} \oplus s_{\beta_2}]$;
2) $\lambda\tilde{s} = [\lambda s_\alpha, \lambda s_\beta], \lambda \in [0, 1]$.

In order to compare any two uncertain linguistic values $\tilde{s}_1 = [s_{\alpha_1}, s_{\beta_1}]$ and $\tilde{s}_2 = [s_{\alpha_2}, s_{\beta_2}]$, Xu [42] introduced a simple formula:

$$p(\tilde{s}_1 \geq \tilde{s}_2) = \min\{\max((\beta_1 - \alpha_2)\big/(l_{\tilde{s}_1} + l_{\tilde{s}_2}), 0), 1\} \tag{34}$$

where $l_{\tilde{s}_1} = \beta_1 - \alpha_1, l_{\tilde{s}_2} = \beta_2 - \alpha_2. p(\tilde{s}_1 \geq \tilde{s}_2)$ is called the degree of possibility of $\tilde{s}_1 \geq \tilde{s}_2. p(\tilde{s}_1 \geq \tilde{s}_2)$ has the following useful properties:

$$0 \leq p(\tilde{s}_1 \geq \tilde{s}_2) \leq 1, p(\tilde{s}_1 \geq \tilde{s}_2) + p(\tilde{s}_2 \geq \tilde{s}_1) = 1, p(\tilde{s}_1 \geq \tilde{s}_1) = 0.5 \tag{35}$$

Xu [42, 43, 50] further developed some operators for aggregating uncertain linguistic information:

**Definition 15.** [50] *Let $ULWA : \tilde{S}^n \to \tilde{S}$, if*

$$ULWA\,(\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n) = w_1\tilde{s}_1 \oplus w_2\tilde{s}_2 \oplus \cdots \oplus w_n\tilde{s}_n \tag{36}$$

*where $w = (w_1, w_2, \ldots, w_n)$ is the weighting vector of the $\tilde{s}_i$, and $w_i \in [0, 1]$, $\sum_{i=1}^n w_i = 1$, then $ULWA$ is called an uncertain linguistic weighted averaging (ULWA) operator. Especially, if $w = (1/n\ , 1/n\ , \ldots, 1/n\ )$, then the ULWA operator is reduced to an uncertain linguistic averaging (ULA) operator.*

**Definition 16.** [50] *An ULOWA operator of dimension* ***n*** *is a mapping $ULOWA : \tilde{S}^n \to \tilde{S}$ that has an associated weighting vector $w = (w_1, w_2, \ldots, w_n)$ with $w_i \in [0, 1]$, $\sum_{i=1}^n w_i = 1$, and is defined to aggregate a collection of uncertain linguistic variables $\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n$ according to the following expression:*

$$ULOWA\,(\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n) = w_1\tilde{s}_{\sigma_1} \oplus w_2\tilde{s}_{\sigma_2} \oplus \cdots \oplus w_n\tilde{s}_{\sigma_n} \tag{37}$$

*where $\tilde{s}_{\sigma_j}$ is the $j$th largest of the $\tilde{s}_i$. Especially, if $w = (1/n\ , 1/n\ , \ldots, 1/n\ )$, then the ULOWA operator is reduced to the ULA operator.*

To rank these uncertain linguistic arguments $\tilde{s}_i\,(i = 1, 2, \ldots, n)$ we first compare each argument $\tilde{s}_i$ with all arguments $\tilde{s}_i\,(i = 1, 2, \ldots, n)$ by using (27), and let $p_{ij} = p(\tilde{s}_i \geq \tilde{s}_j)$, then we can construct a complementary matrix [52]:$P = (p_{ij})_{n\times n}$, where $p_{ij} \geq 0$, $p_{ij} + p_{ji} = 1$, $p_{ii} = 1\big/2, i, j = 1, 2, \ldots, n$. Summing all elements in each line of matrix $P$, we have $p_i = \sum_{j=1}^n p_{ij}, i = 1, 2, \ldots, n$. Then we can rank the uncertain linguistic variables $\tilde{s}_i\,(i = 1, 2, \ldots, n)$ in descending order in accordance with the values of $p_i\,(i = 1, 2, \ldots, n)$.

**Definition 17.** [42] *An uncertain linguistic hybrid aggregation (ULHA) operator is a mapping* $ULHA : \tilde{S}^n \to \tilde{S}$*, which has an associated vector* $w = (w_1, w_2, \ldots, w_n)$ *with* $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$*, such that*

$$ULHA(\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_n) = w_1 \tilde{s}_{\beta_1} \oplus w_2 \tilde{s}_{\beta_2} \oplus \cdots \oplus w_n \tilde{s}_{\beta_n} \quad (38)$$

*where* $\tilde{s}_{\beta_j}$ *is the* $j$*th largest of the uncertain linguistic weighted arguments* $\tilde{s}_i^{'}(\tilde{s}_i^{'} = n\omega_i \tilde{s}_i, i = 1, 2, \ldots, n)$, $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ *is the weight vector of the* $\tilde{s}_i (i = 1, 2, \ldots, n\ )$ *with* $\omega_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$*, and n is the balancing coefficient, which plays a role of balance.*

Especially, if $w = (1/n\ , 1/n\ , \ldots, 1/n)$, then the ULHA operator is reduced to the ULWA operator; if $\omega = (1/, n1/n\ , \ldots, 1/n)$, then the ULHA operator is reduced to the ULOWA operator.

Xu [43] applied the ULOWA operator and the ULHA operator to the partner selection of an enterprise in the field of supply chain management and the maintenance service of product.

Similar to (33), Xu [50] defined an induced uncertain LOWA (IULOWA) operator as follows:

$$IULOWA\left(\langle u_1, \tilde{s}_1 \rangle, \langle u_2, \tilde{s}_2 \rangle \ldots, \langle u_n, \tilde{s}_n \rangle\right) = w_1 \tilde{s}_{\sigma_1} \oplus w_2 \tilde{s}_{\sigma_2} \oplus \cdots \oplus w_n \tilde{s}_{\sigma_n} \quad (39)$$

where $w = (w_1, w_2, \ldots, w_n)$ is a weighting vector, such that $w_i \in [0, 1]$, $\sum_{i=1}^{n} w_i = 1$, $\tilde{s}_{\sigma_j}$ is the $\tilde{s}_i$ value of the pair $\langle u_i, \tilde{s}_i \rangle$ having the $j$th largest $\mu_i$, and $\mu_i$ in $\langle u_i, \tilde{s}_i \rangle$ is referred to as the order inducing variable and $\tilde{s}_i$ as the uncertain linguistic argument variable.

Especially, if $u_i = \tilde{s}_i$, for all $i$, then the IULOWA operator is reduced to the ULOWA operator; if $u_i = No.i$, for all $i$, where $No.i$ is the ordered position of the $s_i$, then the IULOWA operator is reduced to the ULWA operator; if $w = (1/n\ , 1/n\ , \ldots, 1/n)$, then the IULOWA operator is reduced to the ULA operator.

However, if there is a tie between $\langle u_i, \tilde{s}_i \rangle$ and $\left\langle u_j, \tilde{s}_j \right\rangle$ with respect to order inducing variables. In this case, we can replace the argument component of each of $\langle u_i, \tilde{s}_i \rangle$ and $\left\langle u_j, \tilde{s}_j \right\rangle$ by their average $(\tilde{s}_i \oplus \tilde{s}_j) \big/ 2$. If $k$ items are tied, we replace these by $k$ replica's of their average.

## 3 Concluding Remarks

In this chapter, we have provided a detailed overview of the existing operators for aggregating linguistic information, and briefly discussed their characteristics and relationships. We have also made a comprehensive survey of their applications in a wide variety of areas. Based on their characteristics, all the linguistic aggregation operators can be roughly classified into the following five categories: 1) The linguistic aggregation operators based on linear ordering [3, 12, 43, 53, 55, 56, 57, 58, 61], which make computations by using max ($\vee$), min ($\wedge$), and negation;

2) The extension principle-based linguistic aggregation operators [2, 4, 5, 7, 10, 33], which make operations on the fuzzy numbers that support the semantics of the linguistic labels; 3) The linguistic aggregation operators based on symbols [11, 13, 16, 17, 18, 19, 20, 23, 26, 30, 38], which make computations on the indexes of the linguistic labels; 4) The linguistic aggregation operators based on linguistic 2-tuples [21, 22, 23, 24, 25], which represent the linguistic information with a pair of values called 2-tuple, composed by a linguistic term and a number; and 5) The linguistic aggregation operators, which compute with words directly [40, 41, 42, 43, 44, 45, 46, 48, 50, 51].

The operators in 1)$\sim$3) develop some approximation processes to express the results in the initial expression domains, which produce the consequent loss of information and hence the lack of precision, while those in 4) and 5) allow a continuous representation of the linguistic information on its domain, and thus, they can represent any counting of information obtained in a aggregation process without any loss of information [21].

Although many efforts and progresses have been made in the field of linguistic information aggregation, there are some promising research problems that needs to be answered in the future: 1) The existing linguistic labels in a linguistic label set are uniformly and symmetrically distributed. However, in some situations, the unbalanced linguistic information appears due to the nature of the linguistic variables used in the problems. Therefore, to develop unbalanced linguistic label sets and some operators for aggregating unbalanced linguistic labels is an important future research field [14]; and 2) The increasing complexity and uncertainty of real-world problems may lead to the situations where the input linguistic argument information is incomplete or dynamic (such as negotiation processes, the high technology project investment of venture capital firms, supply chain management, etc.). Thus, to develop incomplete or dynamic linguistic aggregation operators is an interesting and important research topic, which is worth paying attention to.

In addition, we need to further extend the applications of the developed linguistic aggregation operators to other fields, such as neural network, intelligent control, pattern recognition, computer vision, image/signal processing, personnel dynamic assessments, bioinformatics, etc.

**Acknowledgment** This work was supported by the National Natural Science Foundation of China under Grant 70472033.

## References

1. Adamopoulos GI, Pappis GP (1996) A fuzzy linguistic approach to a multicriteria-sequencing problem. European Journal of Operational Research 92: 628–636.
2. Bonissone PP, Decker KS (1986) Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity. In: Kanal LH, Lemmer JF (Eds) Uncertainty in Artificial Intelligence. Amsterdam, The Netherlands: North-Holland, pp. 217–247.
3. Bordogna G, Fedrizzi M, Passi G (1997) A linguistic modeling of consensus in group decision making based on OWA operator. IEEE Transactions on Systems, Man, and Cybernetics-Part A 27: 126–132.

4. Bordogna G, Pasi G (1993) A fuzzy linguistic approach generalizing boolean information retrieval: a model and its evaluation. Journal of the American Society for Information Science and Technology 44: 70–82.
5. Chang P, Chen Y (1994) A fuzzy multicriteria decision making method for technology transfer strategy selection in biotechnology. Fuzzy Sets and Systems 63: 131–139.
6. Chen SJ, Hwang CL (1992) Fuzzy multiple attribute decision making: methods and applications. Springer-Verlag, Berlin.
7. Chen SM (1997) A new method for tool steel materials selection under fuzzy environment. Fuzzy Sets and Systems 92: 265–274.
8. Delgado M, Herrera F, Herrera-Viedma E, Martin-Bautista MJ, Martinez L, Vila MA (2002) A communication model based on the 2-tuple fuzzy linguistic representation for a distributed intelligent agent system on internet. Soft Computing 6: 320–328.
9. Delgado M, Herrera F, Herrera-Viedma E, Martínez L (1998) Combining numerical and linguistic information in group decision making. Information Sciences 107: 177–194.
10. Delgado M, Verdegay JL, Vila MA (1993) Linguistic decision making models. International Journal of Intelligent Systems 7: 479–492.
11. Delgado M, Verdegay JL, Vila MA (1993) On aggregation operators of linguistic labels. International Journal of Intelligent Systems 8: 351–370.
12. Herrera F, Herrera-Viedma E (1997) Aggregation operators for linguistic weighted information. IEEE Transactions on Systems, Man, and Cybernetics-Part A 27: 646–656.
13. Herrera F, Herrera-Viedma E (2000) Linguistic decision analysis: steps for solving decision problems under linguistic information. Fuzzy Sets and Systems 115: 67–82.
14. Herrera F, Herrera-Viedma E (2003) Linguistic preference modeling: foundation models and new trends. Proceedings of 4th International Workshop on Preferences and Decisions, Trento (Italy) pp. 47–51.
15. Herrera F, Herrera-Viedma E, Martínez L (2000) A fusion approach for managing multi-granularity linguistic term sets in decision making. Fuzzy Sets and Systems 114: 43–58.
16. Herrera F, Herrera-Viedma E, Verdegay JL (1995) A sequential selection process in group decision making with a linguistic assessment approach. Information Sciences 85: 223–239.
17. Herrera F, Herrera-Viedma E, Verdgay JL (1996) A model of consensus in group decision making under linguistic assessments. Fuzzy Sets and Systems 78: 73–87.
18. Herrera F, Herrera-Viedma E, Verdegay JL (1996) Direct approach processes in group decision making using linguistic OWA operators. Fuzzy Sets and Systems 79: 175–190.
19. Herrera F, Herrera-Viedma E, Verdegay JL (1997) A rational consensus model in group decision making using linguistic assessments. Fuzzy Sets and Systems 88: 31–49.
20. Herrera F, López E, Rodríguez MA (2002) A linguistic decision model for promotion mix management solved with genetic algorithms. Fuzzy Sets and Systems 131: 47–61.
21. Herrera F, Martínez L (2000) A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Transactions on Fuzzy Systems 8: 746–752.
22. Herrera F, Martínez L (2000) An approach for combining linguistic and numerical information based on the 2-tuple fuzzy linguistic representation model in decision-making. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 8: 539–562.
23. Herrera F, Martínez L (2001) A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision making. IEEE Transactions on Systems, Man, and Cybernetics-Part B 31: 227–234.
24. Herrera F, Martínez L (2001) The 2-tuple linguistic computational model: advantages of its linguistic description, accuracy and consistency. Internal Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (Suppl.): 33–48.
25. Herrera F, Martínez L, Sánchez PJ (2005) Managing non-homogeneous information in group decision making. European Journal of Operational Research 166:115–132.
26. Herrera F, Verdegay JL (1993) Linguistic assessments in group decision. In: Proceedings of the First European Congress on Fuzzy and Intelligent Technologies, Aachen, pp. 941–948.
27. Herrera-Viedma E (2001) Modeling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach. Journal of the American Society for Information Science and Technology 52: 460–475.

28. Herrera-Viedma E, Cordon O, Luque M, Lopez AG, Muñoz AN (2003) A model of fuzzy linguistic IRS based on multi-granular linguistic information. International Journal of Approximate Reasoning 34: 221–239.
29. Herrera-Viedma E, Martínez L, Mata F, Chiclana F (2005) A consensus support system model for group decision making problems with multigranular linguistic preference relations. IEEE Transactions on Fuzzy Systems 13: 644–658.
30. Herrera-Viedma E, Peis E (2003) Evaluating the informative quality of documents in SGML format from judgments by means of fuzzy linguistic techniques based on computing with words. Information Processing and Management 39: 233–249.
31. Köksalan M, Ulu C (2003) An interactive approach for placing alternatives in preference classes. European Journal of Operational Research 144: 429–439.
32. Law CK (1996) Using fuzzy numbers in educational grading systems. Fuzzy Sets and Systems 83: 311–323.
33. Lee HM (1996) Applying fuzzy set theory to evaluate the rate of aggregative risk in software development. Fuzzy Sets and Systems 80: 323–336.
34. Levrat E, Voisin A, Bombardier S, Bremont J (1997) Subjective evaluation of car seat comfort with fuzzy set techniques. International Journal of Intelligent Systems 12: 891–913.
35. Martínez L, Liu J, Yang JB, Herrera F (2005) A multigranular hierarchical linguistic model for design evaluation based on safety and cost analysis. International Journal of Intelligent Systems 20: 1161–1194.
36. Ovchinnikov S (1999) Invariance properties of ordinal OWA operators. International Journal of Intelligent Systems 14: 413–418.
37. Torra V (1996) Negation functions based semantics for ordered linguistic labels. International Journal of Intelligent systems 11: 975–988.
38. Torra V (1997) The weighted OWA operator. International Journal of Intelligent Systems 12: 153–166.
39. Torra V (2001) Aggregation of linguistic labels when semantics is based on antonyms. International Journal of Intelligent Systems 16: 513–524.
40. Xu ZS (2004) EOWA and EOWG operators for aggregating linguistic labels based on linguistic preference relations. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12: 791–810.
41. Xu ZS (2004) A method based on linguistic aggregation operators for group decision making with linguistic preference relations. Information Sciences 166:19–30.
42. Xu ZS (2004) Uncertain linguistic aggregation operators based approach to multiple attribute group decision making under uncertain linguistic environment. Information Sciences 168: 171–184.
43. Xu ZS (2004) Uncertain multiple attribute decision making: methods and applications. Tsinghua University Press, Beijing.
44. Xu ZS (2005) An approach to group decision making based on incomplete linguistic preference relations. International Journal of Information Technology and Decision Making 4: 153–160.
45. Xu ZS (2005) An approach to pure linguistic multiple attribute decision making under uncertainty. International Journal of Information Technology and Decision Making 4: 197–206.
46. Xu ZS (2005) Deviation measures of linguistic preference relations in group decision making. Omega 33: 249–254.
47. Xu ZS (2005) An overview of methods for determining OWA weights. International Journal of Intelligent Systems 20: 843–865.
48. Xu ZS (2006) An approach based on the uncertain LOWG and the induced uncertain LOWG operators to group decision making with uncertain multiplicative linguistic preference relations. Decision Support Systems 41: 488–499.
49. Xu ZS (2006) Induced uncertain linguistic OWA operators applied to group decision making. Information Fusion 7: 231–238.
50. Xu ZS (2006) On generalized induced linguistic aggregation operators. International Journal of General Systems 35: 17–28.

51. Xu ZS (2006) A note on linguistic hybrid arithmetic averaging operator in group decision making with linguistic information. Group Decision and Negotiation, to appear.
52. Xu ZS, Da QL (2002) The uncertain OWA operator. International Journal of Intelligent Systems 17: 569–575.
53. Xu ZS, Da QL (2003) An overview of operators for aggregating information. International Journal of Intelligent Systems 18: 953–969.
54. Yager RR (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man, and Cybernetics 18: 183–190.
55. Yager RR (1992) Applications and extensions of OWA aggregations. International Journal of Man-Machine Studied 37: 103–132.
56. Yager RR (1995) An approach to ordinal decision making. International Journal of Approximate Reasoning 12: 237–261.
57. Yager RR (1996) Quantifier guided aggregation using OWA operators. International Journal of Intelligent Systems 11: 49–73.
58. Yager RR (1998) Fusion of ordinal information using weighted median aggregation. International Journal of Approximate Reasoning 18: 35–52.
59. Yager RR, Filev DP (1999) Induced ordered weighted averaging operators. IEEE Transactions on Systems, Man, and Cybernetics-Part B 29: 141–150.
60. Yager RR, Goldstein LS, Mendels E (1994) FUZMAR: An approach to aggregating market research data based on fuzzy reasoning. Fuzzy Sets and Systems 68: 1–11.
61. Yager RR, Rybalov A (1997) Understanding the median as a fusion operator. International Journal of General Systems 26: 239–263.
62. Zadeh LA, Kacprzyk J (1999) Computing with words in information/intelligent systems-part 1: foundations: part 2: applications. Physica-Verlag, Heidelberg, Vol. I.

# Aggregation Operators in Interval-valued Fuzzy and Atanassov's Intuitionistic Fuzzy Set Theory

**Glad Deschrijver and Etienne Kerre**

**Abstract** In this chapter we give an overview of some recent advances on aggregation operators on $\mathcal{L}^I$, where $\mathcal{L}^I$ is the underlying lattice of interval-valued fuzzy set theory (which is equivalent to Atanassov's intuitionistic fuzzy set theory). We discuss some special classes of t-norms on $\mathcal{L}^I$ and their properties. We show that the t-representable t-norms, which are constructed as a pair of t-norms on [0, 1], are not the t-norms with the most interesting properties. We study additive generators of t-norms on $\mathcal{L}^I$, uninorms on $\mathcal{L}^I$ and generators of uninorms on $\mathcal{L}^I$. We give the general definition and some special classes of aggregation operators on $\mathcal{L}^I$. Finally we discuss the generalization of Yager's OWA operators to interval-valued fuzzy set theory.

## 1 Introduction

Interval-valued fuzzy set theory [25, 39] is an extension of fuzzy set theory in which to each element of the universe a closed subinterval of the unit interval is assigned which approximates the unknown membership degree. Another extension of fuzzy set theory is intuitionistic fuzzy set theory introduced by Atanassov [1, 2, 3]. Atanassov's intuitionistic fuzzy sets assign to each element of the universe not only a membership degree, but also a non-membership degree which is less than or equal to 1 minus the membership degree (in fuzzy set theory the non-membership degree is always equal to 1 minus the membership degree). In [15] it is shown that Atanassov's intuitionistic fuzzy set theory is equivalent to interval-valued fuzzy set theory and that both are equivalent to $L$-fuzzy set theory in the sense of Goguen [24] w.r.t. a special lattice $\mathcal{L}^I$. Another equivalent notion is the notion of vague sets [5, 23].

Fuzzy sets are limited in their expressivity because they can only model vague information (gradations in the notion of membership). On the other hand, interval-valued fuzzy sets are also capable of dealing with uncertainty (lack of information): they approximate the exact, but incompletely known, membership degree by an interval. Furthermore, the computational complexity is not excessively higher than when ordinary fuzzy sets are used. In this paper we deal with the issue of

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

aggregating uncertain values modelled by interval-valued fuzzy sets; in particular we present several operators that can be used for this kind of aggregation.

**Definition 1.** *Define* $\mathcal{L}^I = (L^I, \leq_{L^I})$ *by*

$$L^I = \{[x_1, x_2] \mid (x_1, x_2) \in [0,1]^2 \text{ and } x_1 \leq x_2\},$$
$$[x_1, x_2] \leq_{L^I} [y_1, y_2] \iff (x_1 \leq y_1 \text{ and } x_2 \leq y_2),$$
$$\text{for all } [x_1, x_2], [y_1, y_2] \text{ in } L^I.$$

The structure $\mathcal{L}^I$ is a complete lattice (see Lemma 2.1 in [15]).

**Definition 2.** [25, 39] *An interval-valued fuzzy set on* $U$ *is a mapping* $A : U \to L^I$.

**Definition 3.** [1, 2, 3] *An intuitionistic fuzzy set (in the sense of Atanassov) on* $U$ *is a set*

$$A = \{(u, \mu_A(u), \nu_A(u)) \mid u \in U\},$$

*where* $\mu_A(u) \in [0,1]$ *denotes the membership degree and* $\nu_A(u) \in [0,1]$ *the non-membership degree of* $u$ *in* $A$ *and where for all* $u \in U$, $\mu_A(u) + \nu_A(u) \leq 1$.

An intuitionistic fuzzy set $A$ (in the sense of Atanassov) on $U$ can be represented by the $\mathcal{L}^I$-fuzzy set $A$ given by

$$A : U \to L^I :$$
$$u \mapsto [\mu_A(u), 1 - \nu_A(u)], \quad \forall u \in U$$

In Fig. 1 the set $L^I$ is shown. Note that to each element $x = [x_1, x_2]$ of $L^I$ corresponds a point $(x_1, x_2) \in \mathbb{R}^2$.

In the sequel, if $x \in L^I$, then we denote its bounds by $x_1$ and $x_2$, i.e. $x = [x_1, x_2]$. The smallest and the largest element of $\mathcal{L}^I$ are given by $0_{\mathcal{L}^I} = [0,0]$ and $1_{\mathcal{L}^I} = [1,1]$. Note that, for $x, y$ in $L^I$, $x <_{L^I} y$ is equivalent to $x \leq_{L^I} y$ and $x \neq y$, i.e. either $x_1 < y_1$ and $x_2 \leq y_2$, or $x_1 \leq y_1$ and $x_2 < y_2$. We denote by $x \ll_{L^I} y$: $x_1 < y_1$ and $x_2 < y_2$. We define for further usage the sets

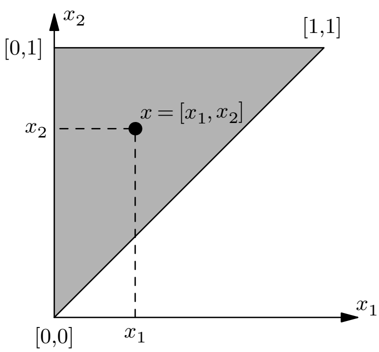



**Fig. 1** The grey area is a representation of $L^I$

$$
\begin{aligned}
D &= \{[x, x] \mid x \in [0, 1]\},\\
\bar{L}^I &= \{[x_1, x_2] \mid (x_1, x_2) \in [-\infty, +\infty]^2 \text{ and } x_1 \leq x_2\},\\
\bar{L}^I_+ &= \{[x_1, x_2] \mid (x_1, x_2) \in [0, +\infty]^2 \text{ and } x_1 \leq x_2\},\\
\bar{L}^I_{\infty,+} &= \{[x_1, x_2] \mid (x_1, x_2) \in [0, +\infty]^2 \text{ and } x_1 \leq x_2\},\\
\bar{D}_{\infty,+} &= \{[x, x] \mid x \in [0, +\infty]\}.
\end{aligned}
$$

Let $n \in \mathbb{N} \setminus \{0\}$, $F$ be an $n$-ary mapping on $L^I$ and $f$ an $n$-ary mapping on $[0, 1]$. We say that $F$ is a natural extension of $f$ to $\mathcal{L}^I$ if $F([x_1, x_1], \ldots, [x_n, x_n]) = [f(x_1, \ldots, x_n), f(x_1, \ldots, x_n)]$, for all $x_1, \ldots, x_n$ in $[0, 1]$.

## 2 Aggregation Operators on $\mathcal{L}^I$

We denote from now on by $\mathbb{N}^*$ the set $\mathbb{N} \setminus \{0\}$. In fuzzy set theory, aggregation operators are defined as follows (see e.g. [6, 27, 32]).

**Definition 4.** *An aggregation operator $A$ on* $([0, 1], \leq)$ *is a mapping* $A : \bigcup_{n\in\mathbb{N}^*}[0, 1]^n \to [0, 1]$ *with the following properties:*

*(a1)* $A(x) = x$, *for all* $x \in [0, 1]$*;*
*(a2)* *if* $x_i \leq y_i$ *for all* $i \in \{1, 2, \ldots, n\}$, *then* $A(x_1, \ldots, x_n) \leq A(y_1, \ldots, y_n)$, *for all* $n \in \mathbb{N}^*$ *and for all* $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ *in* $[0, 1]^n$*;*
*(a3)* $A\Big(\underbrace{0, \ldots, 0}_{n\ \text{times}}\Big) = 0$, *for all* $n \in \mathbb{N}^*$*;*
*(a4)* $A\Big(\underbrace{1, \ldots, 1}_{n\ \text{times}}\Big) = 1$, *for all* $n \in \mathbb{N}^*$.

We extend this definition to $\mathcal{L}^I$ as follows.

**Definition 5.** [17] *An aggregation operator $\mathcal{A}$ on $\mathcal{L}^I$ is a mapping* $\mathcal{A} : \bigcup_{n\in\mathbb{N}^*}(L^I)^n \to L^I$ *with the following properties:*

*(A1)* $\mathcal{A}(x) = x$, *for all* $x \in L^I$*;*
*(A2)* *if* $x_i \leq_{L^I} y_i$ *for all* $i \in \{1, 2, \ldots, n\}$, *then* $\mathcal{A}(x_1, \ldots, x_n) \leq_{L^I} \mathcal{A}(y_1, \ldots, y_n)$, *for all* $n \in \mathbb{N}^*$ *and for all* $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ *in* $(L^I)^n$*;*
*(A3)* $\mathcal{A}\Big(\underbrace{0_{\mathcal{L}^I}, \ldots, 0_{\mathcal{L}^I}}_{n\ \text{times}}\Big) = 0_{\mathcal{L}^I}$, *for all* $n \in \mathbb{N}^*$*;*
*(A4)* $\mathcal{A}\Big(\underbrace{1_{\mathcal{L}^I}, \ldots, 1_{\mathcal{L}^I}}_{n\ \text{times}}\Big) = 1_{\mathcal{L}^I}$, *for all* $n \in \mathbb{N}^*$.

**Definition 6.** *A negation on $\mathcal{L}^I$ is a decreasing mapping $\mathcal{N} : L^I \to L^I$ for which $\mathcal{N}(0_{\mathcal{L}^I}) = 1_{\mathcal{L}^I}$ and $\mathcal{N}(1_{\mathcal{L}^I}) = 0_{\mathcal{L}^I}$. If $\mathcal{N}(\mathcal{N}(x)) = x$, for all $x \in L^I$, then $\mathcal{N}$ is called involutive.*

The mapping $\mathcal{N}_s$ defined by $\mathcal{N}_s(x) = [1-x_2, 1-x_1]$, for all $x \in L^I$, is an involutive negation, called the standard negation on $\mathcal{L}^I$, which is a natural extension of the standard negation $N_s$ on $([0, 1], \leq)$ defined by $N_s(x) = 1 - x$, for all $x \in [0, 1]$.

Let $\mathcal{A}$ be an aggregation operator on $\mathcal{L}^I$ and $\mathcal{N}$ an involutive negation on $\mathcal{L}^I$. The mapping $\mathcal{A}^* : \bigcup_{n\in\mathbb{N}^*}(L^I)^n \to L^I$ defined by

$$\mathcal{A}^*(x_1, \dots, x_n) = \mathcal{N}(\mathcal{A}(\mathcal{N}(x_1), \dots, \mathcal{N}(x_n))),$$

for all $n \in \mathbb{N}^*$ and $x_1, \dots, x_n \in L^I$, is an aggregation operator on $\mathcal{L}^I$, called the dual aggregation operator of $\mathcal{A}$ w.r.t. $\mathcal{N}$.

Aggregation operators on $\mathcal{L}^I$ can be constructed in the following way. Let $A_1$ and $A_2$ be aggregation operators on $([0, 1], \leq)$. Define the mapping $\mathcal{A} : \bigcup_{n\in\mathbb{N}^*}(L^I)^n \to L^I$ by

$$\mathcal{A}(x_1, \dots, x_n) = [A_1((x_1)_1, \dots, (x_n)_1), A_2((x_1)_2, \dots, (x_n)_2)],$$

for all $n \in \mathbb{N}^*$ and $x_1 = [(x_1)_1, (x_1)_2], \dots, x_n = [(x_n)_1, (x_n)_2]$ in $L^I$. Then $\mathcal{A}(x_1, \dots, x_n) \in L^I$ if and only if $A_1 \leq A_2$. Clearly, $\mathcal{A}$ is an aggregation operator on $\mathcal{L}^I$ if and only if $A_1$ and $A_2$ are aggregation operators on $([0, 1], \leq)$. We call $\mathcal{A}$ a representable aggregation operator [17].

Not all aggregation operators on $\mathcal{L}^I$ can be constructed in this way [17]. Consider for instance the mapping $\mathcal{A} : \bigcup_{n\in\mathbb{N}^*}(L^I)^n \to L^I$ defined by, for any aggregation operator $A$ on $([0, 1], \leq)$, for all $n \in \mathbb{N}^*$ and $x_1, \dots, x_n$ in $L^I$,

$$\mathcal{A}(x_1, \dots, x_n) = \Big[A((x_1)_1, \dots, (x_n)_1), \max\Big(A((x_1)_1, (x_2)_2, \dots, (x_n)_2),$$
$$A((x_1)_2, (x_2)_1, (x_3)_2, \dots, (x_n)_2), \dots, A((x_1)_2, \dots, (x_{n-1})_2, (x_n)_1)\Big)\Big]$$

if $n > 1$, and $\mathcal{A}(x_1) = [A((x_1)_1), A((x_1)_2)]$ if $n = 1$. Then $\mathcal{A}$ is an aggregation operator on $\mathcal{L}^I$. Let now $A$ be an aggregation operator on $([0, 1], \leq)$ for which $A(0, 1) = A(1, 0) = 0$, $x_1 = [1, 1]$, $x_1' = [0, 1]$ and $x_2 = [0, 1]$, then $(\mathcal{A}(x_1, x_2))_2 = 1 \neq (\mathcal{A}(x_1', x_2))_2 = 0$. Hence no aggregation operators $A_1$ and $A_2$ on $([0, 1], \leq)$ can be found such that $\mathcal{A}(x_1, x_2) = [A_1((x_1)_1, (x_2)_1), A_2((x_1)_2, (x_2)_2)]$, since otherwise $(\mathcal{A}(x_1, x_2))_2$ would be independent of $(x_1)_1$.

**Definition 7.** [17] *Let $n \in \mathbb{N} \setminus \{0, 1\}$. An n-ary aggregation operator $\mathcal{A}$ on $\mathcal{L}^I$ is a mapping $\mathcal{A} : (L^I)^n \to L^I$ with the following properties:*

*(A1') if $x_i \leq_{L^I} y_i$ for all $i \in \{1, 2, \dots, n\}$, then $\mathcal{A}(x_1, \dots, x_n) \leq_{L^I} \mathcal{A}(y_1, \dots, y_n)$, for all $(x_1, \dots, x_n), (y_1, \dots, y_n)$ in $(L^I)^n$;*

*(A2') $\mathcal{A}(\underbrace{0_{\mathcal{L}^I}, \dots, 0_{\mathcal{L}^I}}_{n \text{ times}}) = 0_{\mathcal{L}^I}$;*

*(A3')* $\mathcal{A}\Big(\underbrace{1_{\mathcal{L}^I}, \ldots, 1_{\mathcal{L}^I}}_{n\ times}\Big) = 1_{\mathcal{L}^I}$.

If $\mathcal{A}$ is an aggregation operator on $\mathcal{L}^I$ and $n \in \mathbb{N} \setminus \{0, 1\}$, then the mapping $\mathcal{A}_n : (L^I)^n \to L^I$ defined by $\mathcal{A}_n(x_1, \ldots, x_n) = \mathcal{A}(x_1, \ldots, x_n)$, for all $(x_1, \ldots, x_n) \in (L^I)^n$ is an $n$-ary aggregation operator on $\mathcal{L}^I$. Conversely, if for all $n \in \mathbb{N}\setminus\{0, 1\}$, $\mathcal{A}_n$ is an $n$-ary aggregation operator on $\mathcal{L}^I$, then the mapping $\mathcal{A} : \bigcup_{n\in\mathbb{N}^*}(L^I)^n \to L^I$ defined by $\mathcal{A}(x_1, \ldots, x_n) = \mathcal{A}_n(x_1, \ldots, x_n)$, for all $(x_1, \ldots, x_n) \in (L^I)^n$, for all $n \in \mathbb{N}^*$, and $\mathcal{A}(x) = x$, for all $x \in L^I$, is an aggregation operator on $\mathcal{L}^I$.

*Example 1.* [17] Some examples of binary aggregation operators on $\mathcal{L}^I$ are, for $x, y$ in $L^I$:

(i) $\mathcal{A}_{A_1,A_2}(x, y) = [A_1(x_1, y_1), \max(A_2(x_1, y_2), A_2(y_1, x_2))]$,
(ii) $\mathcal{A}^*_{A_1,A_2}(x, y) = [\min(A_1(x_1, y_2), A_1(y_1, x_2)), A_2(x_2, y_2)]$,
(iii) $\mathcal{A}^{**}_{A_1,A_2}(x, y) = [\min(A_1(x_1, y_2), A_1(y_1, x_2)), \max(A_2(x_1, y_2), A_2(y_1, x_2))]$,
(iv) $\mathcal{A}_{A_1,A_2,A_3,A_4}(x, y) = [A_3(A_1(x_1,y_2), A_1(y_1,x_2)), A_4(A_2(x_1,y_2), A_2(y_1,x_2))]$,

where $A_1$, $A_2$, $A_3$ and $A_4$ are aggregation operators on $([0, 1], \leq)$ satisfying $A_1 \leq A_2$ and $A_3 \leq A_4$.

## 3 Triangular Norms and Conorms on $\mathcal{L}^I$

**Definition 8.**

- *A t-norm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{T} : (L^I)^2 \to L^I$ which satisfies $\mathcal{T}(1_{\mathcal{L}^I}, x) = x$, for all $x \in L^I$.*
- *A t-conorm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{S} : (L^I)^2 \to L^I$ which satisfies $\mathcal{S}(0_{\mathcal{L}^I}, x) = x$, for all $x \in L^I$.*

Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$, then we denote $x^{(n)_\mathcal{T}} = \mathcal{T}(x, x^{(n-1)_\mathcal{T}})$, for $n \in \mathbb{N} \setminus \{0, 1\}$ and $x^{(1)_\mathcal{T}} = x$. We say that a t-norm $\mathcal{T}$ on $\mathcal{L}^I$ satisfies the residuation principle if and only if, for all $x, y, z$ in $L^I$,

$$\mathcal{T}(x, y) \leq_{L^I} z \iff y \leq_{L^I} \mathcal{I}_\mathcal{T}(x, z) = \sup\{\gamma \mid \gamma \in L^I \text{ and } \mathcal{T}(x, \gamma) \leq_{L^I} z\}.$$

Clearly a t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is a binary aggregation operator on $\mathcal{L}^I$. Note that a t-norm on $\mathcal{L}^I$ displays downward reinforcement, i.e. $\mathcal{T}(x, y) \leq_{L^I} \inf(x, y)$, for all $x, y$ in $L^I$. On the other hand, a t-conorm on $\mathcal{L}^I$ displays upward reinforcement since $\mathcal{S}(x, y) \geq_{L^I} \sup(x, y)$, for all $x, y$ in $L^I$ (see [45] for more information on reinforcement operators on the unit interval).

For t-norms on $\mathcal{L}^I$, we consider the following special classes [13, 14]. Let $\alpha \in [0, 1]$ and let $T$ and $T'$ be t-norms on $([0, 1], \leq)$ such that $T(x, y) \leq T'(x, y)$, for all $x, y$ in [0, 1]. Then the t-norms $\mathcal{T}_{T,T'}$, $\mathcal{T}_T$, $\mathcal{T}_{T,\alpha}$ and $\mathcal{T}'_T$ are defined by, for all $x, y$ in $L^I$,

- $\mathcal{T}_{T,T'}(x, y) = [T(x_1, y_1), T'(x_2, y_2)]$ (t-representable t-norms);
- $\mathcal{T}_T(x, y) = [T(x_1, y_1), \max(T(x_1, y_2), T(x_2, y_1))]$ (pseudo-t-representable t-norms);
- $\mathcal{T}_{T,\alpha}(x, y) = [T(x_1, y_1), \max(T(\alpha, T(x_2, y_2)), T(x_1, y_2), T(x_2, y_1))]$ (generalized pseudo-t-representable t-norms);
- $\mathcal{T}'_T(x, y) = [\min(T(x_1, y_2), T(x_2, y_1)), T(x_2, y_2)]$.

Clearly, for any $\alpha \in [0, 1]$ and any t-norm $T$ on $([0, 1], \leq)$, $\mathcal{T}_{T,T}$, $\mathcal{T}_T$, $\mathcal{T}_{T,\alpha}$ and $\mathcal{T}'_T$ are natural extensions of $T$ to $\mathcal{L}^I$.

*Example 2.* [13] Some examples of these operators are, for $x, y$ in $L^I$,

- $\mathcal{T}_W(x, y) = [\max(0, x_1 + y_1 - 1), \max(0, x_1 + y_2 - 1, x_2 + y_1 - 1)]$, which is a pseudo-t-representable t-norm on $\mathcal{L}^I$, called the Łukasiewicz t-norm on $\mathcal{L}^I$, with representant the Łukasiewicz t-norm $T_W$ on $([0, 1], \leq)$ given by $T_W(x, y) = \max(0, x + y - 1)$, for all $x, y$ in [0, 1];
- $\mathcal{S}_W(x, y) = [\min(1, x_1 + y_2, x_2 + y_1), \min(1, x_2 + y_2)]$ is a natural extension of the Łukasiewicz t-conorm $S_W$ on $([0, 1], \leq)$ given by, $S_W(x, y) = \min(1, x + y)$, for all $x, y$ in [0, 1]. Furthermore, $\mathcal{S}_W(x, y) = \mathcal{N}_s(\mathcal{T}_W(\mathcal{N}_s(x), \mathcal{N}_s(y)))$, for all $x, y$ in $L^I$.

## 4 Archimedean t-norms on $\mathcal{L}^I$

In fuzzy set theory the class of Archimedean t-norms play an important role: continuous t-norms can be fully characterized by means of Archimedean t-norms, the Archimedean property is closely related to additive and multiplicative generators, etc. In particular a t-norm $T$ on $([0, 1], \leq)$ is a continuous Archimedean t-norm if and only if there exists a continuous, strictly decreasing mapping $f : [0, 1] \to [0, +\infty]$ with $f(1) = 0$ such that $T(x, y) = f^{-1}(\min(f(0), f(x) + f(y)))$, for all $x, y$ in [0, 1] [29, 31, 34]. The mapping $f$ is called an additive generator of $T$. Generators play an important role in aggregation (see e.g. [7, 21, 26, 32]). Therefore, in this section we will investigate Archimedean t-norms on $\mathcal{L}^I$ and in the next section we will discuss additive generators of t-norms on $\mathcal{L}^I$ (see [9] for more information on multiplicative generators).

In order to define Archimedean t-norms, we need a metric on $L^I$. Well-known metrics include the Euclidean distance and the Hamming distance. In the two-dimensional space $\mathbb{R}^2$ they are defined as follows:

- the Euclidean distance between two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in $\mathbb{R}^2$ is given by $d^E(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$,

- the Hamming distance between two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in $\mathbb{R}^2$ is given by $d^H(x, y) = |x_1 - y_1| + |x_2 - y_2|$.

If we restrict these distances to $L^I$ then we obtain the metric spaces $(L^I, d^E)$ and $(L^I, d^H)$.

Archimedean t-norms on the unit interval are defined as follows. A t-norm $T$ on $([0, 1], \leq)$ is Archimedean if for all $x, y$ in $]0, 1[$, there exists an $n \in \mathbb{N}^*$ such that $x^{(n)_T} < y$. We say that $T$ has the limit property if for all $x \in [0, 1]$, $\lim_{n\to+\infty} x^{(n)_T} = 0$. An element $x \in [0, 1]$ is called a zero-divisor of $T$ if there exists an $y \in [0, 1]$ such that $T(x, y) = 0$.

For t-norms on the unit interval we have the following characterizations of the Archimedean property.

**Theorem 1.** [29, 30] *A t-norm $T$ on $([0, 1], \leq)$ is Archimedean if and only if it satisfies the limit property.*

**Theorem 2.** [29, 30] *Let $T$ be a continuous t-norm on $([0, 1], \leq)$. Then $T$ is Archimedean if and only if $T(x, x) < x$, for all $x \in [0, 1]$.*

We extend the previous definitions to $\mathcal{L}^I$. There are several possible extensions of the Archimedean property, which we call Archimedean, weakly Archimedean and strong Archimedean property. Throughout this section we will use the sets $L_1^I = \{x \mid x \in L^I \text{ and } x_1 \in [0, 1]\}$ and $L_{12}^I = \{x \mid x \in L^I \text{ and } x_1 > 0 \text{ and } x_2 < 1\}$.

**Definition 9.** [11] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. We say that*

- *$\mathcal{T}$ is Archimedean if*

$$(\forall (x, y) \in (L_1^I)^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} <_{L^I} y);$$

- *$\mathcal{T}$ is strong Archimedean if*

$$(\forall (x, y) \in (L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\})^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} <_{L^I} y);$$

- *$\mathcal{T}$ is weakly Archimedean if*

$$(\forall (x, y) \in (L_{12}^I)^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} <_{L^I} y).$$

Obviously, if a t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is Archimedean, then it is weakly Archimedean, and if $\mathcal{T}$ is strong Archimedean, then it is Archimedean. The converse implications do not hold (see [11] for counterexamples).

In [4] the Archimedean property is introduced for t-norms on a general bounded poset. If we apply their definition to $\mathcal{L}^I$, then we obtain the following condition for a t-norm $\mathcal{T}$ on $\mathcal{L}^I$:

$$(\forall (x, y) \in (L^I)^2)((\forall n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} \geq_{L^I} y) \Rightarrow (x = 1_{\mathcal{L}^I} \text{ or } y = 0_{\mathcal{L}^I})). \quad (1)$$

On the unit interval the Archimedean property is equivalent to the counterpart of (1) on the unit interval. Although the definition of the strong Archimedean property on $\mathcal{L}^I$ is similar to the definition of the Archimedean property on the unit interval, on $\mathcal{L}^I$ property (1) is *not* equivalent to the strong Archimedean property: the following theorem shows that the Archimedean property defined using (1) corresponds to our definition of the Archimedean property given in Definition 9.

**Theorem 3.** [11] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. Then $\mathcal{T}$ is Archimedean (in the sense of Definition 9) if and only if $\mathcal{T}$ satisfies* (1).

**Definition 10.** [11] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. We say that*

- *$\mathcal{T}$ has the limit property if for all $x \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, $\lim_{n\to+\infty} x^{(n)_\mathcal{T}} = 0_{\mathcal{L}^I}$;*
- *$\mathcal{T}$ has the weak limit property if for all $x \in L^I_{12}$, $\lim_{n\to+\infty} x^{(n)_\mathcal{T}} = 0_{\mathcal{L}^I}$.*

Now we extend Theorem 1 to $\mathcal{L}^I$.

**Theorem 4.** [11] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. Then the following are equivalent:*

*(A1)* $\mathcal{T}$ *is Archimedean;*
*(A2)* $(\forall (x, y) \in (L^I_1)^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} \ll_{L^I} y)$;
*(A3)* $\mathcal{T}$ *satisfies the limit property;*
*(A4)* $(\forall x \in L^I_1)(\lim_{n\to+\infty} x^{(n)_\mathcal{T}} = 0_{\mathcal{L}^I})$.

**Theorem 5.** [11] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. Then the following are equivalent:*

*(wA1)* $\mathcal{T}$ *is weakly Archimedean;*
*(wA2)* $(\forall (x, y) \in (L^I_{12})^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} \ll_{L^I} y)$;
*(wA3)* $\mathcal{T}$ *satisfies the weak limit property.*

**Theorem 6.** [11] *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$. Then $\mathcal{T}$ satisfies the Archimedean property if and only if $\mathcal{T}(x, x) <_{L^I} x$, for all $x \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$.*

In Table 1 it is shown under which conditions the different classes of t-norms that we considered before satisfy the different types of the Archimedean property. Clearly, the only class which satisfies the Archimedean property is the class of pseudo-t-representable t-norms from which the t-representable t-norms are excluded.

**Table 1** The Archimedean property for the different classes of t-norms

| | Weakly Archimedean | Archimedean | Strong Archimedean |
|---|---|---|---|
| $\mathcal{T}_{T,T'}$ | iff $T$ and $T'$ are Archimedean | no | no |
| $\mathcal{T}_{T,t}$ $(t < 1)$ | iff $T$ is Archimedean | iff $T$ is Archimedean | iff $T$ is Archimedean and has zero-divisors |
| $\mathcal{T}'_T$ | iff $T$ is Archimedean | no | no |

For t-norms on the unit interval we have the following characterization (see [22]): a t-norm $T$ is continuous, Archimedean and nilpotent if and only if there exists a permutation $\varphi$ of $[0, 1]$ such that $T$ is the $\varphi$-transform of the Łukasiewicz t-norm, i.e. $T = \varphi^{-1} \circ T_W \circ (\varphi \times \varphi)$, where $\times$ denotes the product operation [18].

For t-norms on $\mathcal{L}^I$ we have a similar representation. Therefore we introduce the following property:

$$(\exists(x, y) \in (L^I)^2)(x_1 > 0 \text{ and } y_1 > 0 \text{ and } \mathcal{T}(x, y) = 0_{\mathcal{L}^I}). \tag{2}$$

Note that for continuous t-norms $T$ on $([0, 1], \leq)$ nilpotency is equivalent to the property that there exist $x, y$ in $[0, 1]$ such that $T(x, y) = 0$. So the following theorem can be seen as a natural extension of the previously mentioned representation theorem for t-norms on $([0, 1], \leq)$.

**Theorem 7.** [13] *Consider a mapping $\mathcal{T} : (L^I)^2 \to L^I$. Then the following statements are equivalent:*

*(i) $\mathcal{T}$ is a continuous Archimedean t-norm on $\mathcal{L}^I$ satisfying the residuation principle, formula* (2)*, $\mathcal{I}_{\mathcal{T}}(D, D) \subseteq D$ and $\mathcal{T}([0, 1], [0, 1]) = 0_{\mathcal{L}^I}$;*
*(ii) there exists a continuous increasing permutation $\Phi$ of $L^I$ with increasing inverse such that*

$$\mathcal{T} = \Phi^{-1} \circ \mathcal{T}_W \circ (\Phi \times \Phi).$$

Theorem 7 and Table 1 show that the pseudo-t-representable t-norms on $\mathcal{L}^I$ have more interesting properties than the t-representable t-norms.

## 5 Generators of t-norms on $\mathcal{L}^I$

In [10] the following arithmetic operations on $\bar{L}^I$ are introduced.

**Definition 11.** *The addition and the substraction of two elements $x$ and $y$ of $\bar{L}^I$ are given by,*

$$x \oplus_{\mathcal{L}^I} y = [\min(x_1 + y_2, x_2 + y_1), x_2 + y_2];$$
$$x \ominus_{\mathcal{L}^I} y = [x_1 - y_2, \max(x_1 - y_1, x_2 - y_2)].$$

Note that it does not necessarily hold that $x \oplus_{\mathcal{L}^I} y \in L^I$ or $x \ominus_{\mathcal{L}^I} y \in L^I$, for any $x, y$ in $L^I$. This is not a big issue, since it also does not hold that $x + y \in [0, 1]$ or $x - y \in [0, 1]$, for all $x, y$ in $[0, 1]$. On the other hand, $x \oplus_{\mathcal{L}^I} y \in \bar{L}^I$ and $x \ominus_{\mathcal{L}^I} y \in \bar{L}^I$, for all $x, y$ in $\bar{L}^I$. Moreover $\oplus_{\mathcal{L}^I}$ and $\ominus_{\mathcal{L}^I}$ are natural extensions of the classical addition and substraction on $\mathbb{R}$.

The reason for defining these operations in this way is that they allow us to write the Łukasiewicz t-norm and some related operations using a similar algebraic formula as their counterparts on the unit interval. Indeed, we find, for all $x, y$ in $L^I$,

$$\mathcal{N}_s(x) = 1_{\mathcal{L}^I} \ominus_{\mathcal{L}^I} x,$$
$$\mathcal{T}_W(x, y) = \sup(0_{\mathcal{L}^I}, x \ominus_{\mathcal{L}^I} (1_{\mathcal{L}^I} \ominus_{\mathcal{L}^I} y)),$$
$$\mathcal{S}_W(x, y) = \inf(1_{\mathcal{L}^I}, x \oplus_{\mathcal{L}^I} y),$$
$$\mathcal{I}_{\mathcal{T}_W}(x, y) = \inf(1_{\mathcal{L}^I}, y \oplus_{\mathcal{L}^I} (1_{\mathcal{L}^I} \ominus_{\mathcal{L}^I} x)).$$

On the other hand, for all $x, y$ in $L^I$,

$$\mathcal{T}'_{\mathcal{T}_W}(x, y) = \sup(0_{\mathcal{L}^I}, (x \oplus_{\mathcal{L}^I} y) \ominus_{\mathcal{L}^I} 1_{\mathcal{L}^I}) = \sup(0_{\mathcal{L}^I}, x \oplus_{\mathcal{L}^I} (y \ominus_{\mathcal{L}^I} 1_{\mathcal{L}^I})).$$

Some other natural extensions of t-norms, t-conorms and implications on $([0, 1], \leq)$ can be written using a similar algebraic formula as their counterparts on the unit interval (see [10, 12] for more details).

We extend $\oplus_{\mathcal{L}^I}$ to $\bar{L}^I \cup \bar{L}^I_{\infty,+}$ as follows: if $x, y \in \bar{L}^I \cup \bar{L}^I_{\infty,+}$, then $x \oplus_{\mathcal{L}^I} y = [\min(x_1 + y_2, x_2 + y_1), x_2 + y_2]$, where any of $x_1, x_2, y_1$ and $y_2$ can be $+\infty$. For instance, if $x \in \bar{L}^I$ and $y = [y_1, +\infty]$, where $y_1 \in [0, +\infty[$, then $x \oplus_{\mathcal{L}^I} y = [x_2 + y_1, +\infty]$. In a similar way, $\ominus_{\mathcal{L}^I}$ is extended to $\bar{L}^I \cup \bar{L}^I_{\infty,+}$.

**Definition 12.** [22, 28, 29] *A mapping $f : [0, 1] \to [0, +\infty]$ satisfying the following conditions:*

*(ag.1)* *$f$ is strictly decreasing;*
*(ag.2)* *$f(1) = 0$;*
*(ag.3)* *$f$ is right-continuous in $0$;*
*(ag.4)* *$f(x) + f(y) \in \text{rng}(f) \cup [f(0), +\infty]$, for all $x, y$ in $[0, 1]$;*

*is called an additive generator on* $[0, 1]$.

**Definition 13.** [28, 29] *Let $f : [0, 1] \to [0, +\infty]$ be a strictly decreasing function. The pseudo-inverse $f^{(-1)} : [0, +\infty] \to [0, 1]$ of $f$ is defined by, for all $y \in [0, +\infty]$,*

$$f^{(-1)}(y) = \sup(\{0\} \cup \{x \mid x \in [0, 1] \textit{ and } f(x) > y\}.$$

We extend these definitions to $L^I$ as follows.

**Definition 14.** [9] *Let $f : L^I \to \bar{L}^I_{\infty,+}$ be a strictly decreasing function. The pseudo-inverse $f^{(-1)} : \bar{L}^I_{\infty,+} \to L^I$ of $\mathfrak{f}$ is defined by, for all $y \in \bar{L}^I_{\infty,+}$,*

$$f^{(-1)}(y) = \begin{cases} \sup\{x \mid x \in L^I \textit{ and } f(x) \gg_{L^I} y\}, \textit{ if } y \ll_{L^I} f(0_{\mathcal{L}^I}); \\ \sup(\{0_{\mathcal{L}^I}\} \cup \{x \mid x \in L^I \textit{ and } (f(x))_1 > y_1 \\ \quad \textit{and } (f(x))_2 \geq (f(0_{\mathcal{L}^I}))_2\}), \textit{ if } y_2 \geq (f(0_{\mathcal{L}^I}))_2; \\ \sup(\{0_{\mathcal{L}^I}\} \cup \{x \mid x \in L^I \textit{ and } (f(x))_2 > y_2 \\ \quad \textit{and } (f(x))_1 \geq (f(0_{\mathcal{L}^I}))_1\}), \textit{ if } y_1 \geq (f(0_{\mathcal{L}^I}))_1. \end{cases}$$

**Definition 15.** [9] *A mapping* $f : L^I \rightarrow \bar{L}^I_{\infty,+}$ *satisfying the following conditions:*

*(AG.1)* $\mathfrak{f}$ *is strictly decreasing;*
*(AG.2)* $\mathfrak{f}(1_{\mathcal{L}^I}) = 0_{\mathcal{L}^I}$*;*
*(AG.3)* $\mathfrak{f}$ *is right-continuous in* $0_{\mathcal{L}^I}$*;*
*(AG.4)* $\mathfrak{f}(x) \oplus_{\mathcal{L}^I} \mathfrak{f}(y) \in \mathcal{R}(\mathfrak{f})$*, for all* $x, y$ *in* $L^I$*, where* $\mathcal{R}(\mathfrak{f}) = \text{rng}(\mathfrak{f})$
$\cup \{x \mid x \in \bar{L}^I_{\infty,+}$ *and* $[x_1, (\mathfrak{f}(0_{\mathcal{L}^I}))_2] \in \text{rng}(\mathfrak{f})$ *and* $x_2 \geq (\mathfrak{f}(0_{\mathcal{L}^I}))_2\}$
$\cup \{x \mid x \in \bar{L}^I_{\infty,+}$ *and* $[(\mathfrak{f}(0_{\mathcal{L}^I}))_1, x_2] \in \text{rng}(\mathfrak{f})$ *and* $x_1 \geq (\mathfrak{f}(0_{\mathcal{L}^I}))_1\}$
$\cup \{x \mid x \in \bar{L}^I_{\infty,+}$ *and* $x \geq_{L^I} \mathfrak{f}(0_{\mathcal{L}^I})\}$*;*
*(AG.5)* $\mathfrak{f}^{(-1)}(\mathfrak{f}(x)) = x$*, for all* $x \in L^I$*;*

*is called an additive generator on* $\mathcal{L}^I$*.*

In [9] it is shown that condition (AG.5) is necessary in the above definition in order to be able to use additive generators on $\mathcal{L}^I$ for the representation of t-norms on $\mathcal{L}^I$.

**Theorem 8.** [9] *Let* $f$ *be an additive generator on* $([0, 1], \leq)$ *and let* $\mathfrak{f} : L^I \rightarrow \bar{L}^I_{\infty,+}$ *the mapping defined by, for all* $x \in L^I$*,*

$$\mathfrak{f}(x) = [f(x_2), f(x_1)]. \tag{3}$$

*Then, for all* $y \in \bar{L}^I_{\infty,+}$*,* $\mathfrak{f}^{(-1)}(y) = [f^{(-1)}(y_2), f^{(-1)}(y_1)]$.

**Definition 16.** [9] *Let* $\mathfrak{f}$ *be an additive generator on* $\mathcal{L}^I$*. If there exists an additive generator* $f$ *on* $[0, 1]$ *such that* (3) *holds, for all* $x \in L^I$*, then* $\mathfrak{f}$ *is called representable and* $f$ *is called the representant of* $\mathfrak{f}$.

**Theorem 9.** [9] *A mapping* $f : L^I \rightarrow \bar{L}^I_{\infty,+}$ *is a continuous additive generator on* $\mathcal{L}^I$ *for which* $\mathfrak{f}(D) \subseteq \bar{D}_{\infty,+}$ *if and only if there exists a continuous additive generator* $\mathfrak{f}$ *on* $([0, 1], \leq)$ *such that* (3) *holds for all* $x \in L^I$*.*

**Definition 17.** [9] *An additive generator* $f$ *on* $\mathcal{L}^I$ *is called an additive generator of a t-norm* $\mathcal{T}$ *on* $\mathcal{L}^I$ *if and only if, for all* $x, y \in L^I$*,*

$$\mathcal{T}(x, y) = \mathfrak{f}^{(-1)}(\mathfrak{f}(x) \oplus_{\mathcal{L}^I} \mathfrak{f}(y)). \tag{4}$$

**Theorem 10.** [9] *Let* $\mathfrak{f}$ *be an additive generator on* $\mathcal{L}^I$*. Define the mapping* $\mathcal{T} : (L^I)^2 \rightarrow L^I$ *by* (4) *for all* $x, y \in L^I$*. Then* $\mathcal{T}$ *is a t-norm on* $\mathcal{L}^I$*.*

**Theorem 11.** [9] *Let* $\mathfrak{f}$ *be a representable additive generator on* $\mathcal{L}^I$*. Then the mapping* $\mathcal{T} : (L^I)^2 \rightarrow L^I$ *defined by* (4) *for all* $x, y \in L^I$*, is a pseudo-t-representable t-norm on* $L^I$*.*

**Theorem 12.** [9] *Let* $\mathcal{T}$ *be a pseudo-t-representable t-norm on* $\mathcal{L}^I$ *with representant* $T$*. Then* $\mathcal{T}$ *has an additive generator* $\mathfrak{f}$ *for which* $\mathfrak{f}(D) \subseteq \bar{D}_{\infty,+}$*, if and only if* $T$ *has an additive generator. Furthermore, if* $f$ *is an additive generator of* $T$*, then the mapping* $\mathfrak{f} : L^I \rightarrow \bar{L}^I_{+}$ *defined by* (4) *for all* $x \in L^I$*, is an additive generator of* $\mathcal{T}$*.*

*Example 3.* The mapping $f_W : [0,1] \to [0,+\infty]$ defined by $f_W(x) = 1 - x$, for all $x \in [0,1]$ is an additive generator of $T_W$ [28, 29]. Let $\mathfrak{f}_W$ be the additive generator on $\mathcal{L}^I$ defined by $\mathfrak{f}_W(x) = 1_{\mathcal{L}^I} \ominus_{\mathcal{L}^I} x = [1 - x_2, 1 - x_1]$, for all $x \in L^I$, then $f_W^{(-1)}(y) = [\max(0, 1 - y_2), \max(0, 1 - y_1)]$, for all $y \in \bar{L}^I_+$. We obtain, for all $x, y$ in $L^I$, that $\mathfrak{f}_W^{(-1)}(\mathfrak{f}_W(x) \oplus_{\mathcal{L}^I} \mathfrak{f}_W(y)) = \mathcal{T}_W(x, y)$. So $\mathfrak{f}_W$ is an additive generator of $\mathcal{T}_W$. Hence the continuous additive generator $f_W$ of the Łukasiewicz t-norm on $[0,1]$ has a natural extension $\mathfrak{f}_W$ to $\mathcal{L}^I$ which is a continuous additive generator of the Łukasiewicz t-norm on $\mathcal{L}^I$.

From Theorem 9 and Theorem 11 it follows that only pseudo-t-representable t-norms can have a continuous additive generator on $\mathcal{L}^I$ which is a natural extension of a generator on $[0,1]$.

## 6 Uninorms on $\mathcal{L}^I$

Compensation behavior (this means that the aggregated value lies between the highest and the lowest of the given values) seems to be a crucial property in many applications in which data is aggregated (see e.g. [46]). However, t-norms and t-conorms do not satisfy this property. For this reason, Yager et al. [20, 44] introduced uninorms on $([0,1], \leq)$ as a generalization of both t-norms and t-conorms. This notion is extended to $\mathcal{L}^I$ as follows.

**Definition 18.** [16] *A uninorm on $\mathcal{L}^I$ is a commutative, associative, increasing mapping $\mathcal{U} : (L^I)^2 \to L^I$ for which there exists an $e \in L^I$ such that $\mathcal{U}(e, x) = x$, for all $x \in L^I$. The element $e$ is called the neutral element of $\mathcal{U}$.*

If $e = 0_{\mathcal{L}^I}$, then $\mathcal{U}$ is a t-conorm on $\mathcal{L}^I$, if $e = 1_{\mathcal{L}^I}$, then we have a t-norm on $\mathcal{L}^I$. The neutral element $e$ of a uninorm $\mathcal{U}$ can be considered as the score that we would give to an argument which should not have any influence in the aggregation. It can be seen as a null vote.

Yager and Rybalov [45] show that uninorms on $([0,1], \leq)$ display both downward and upward reinforcement if $e \in [0,1]$. For uninorms on $\mathcal{L}^I$ a similar property holds. Let $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ and define $E = \{x \mid x \in L^I \text{ and } x \leq_{L^I} e\}$ and $E' = \{x \mid x \in L^I \text{ and } x \geq_{L^I} e\}$. Then the mappings $\Phi_e : L^I \to E$ and $\Psi_e : L^I \to E'$ defined by, for all $x \in L^I$,

$$\Phi_e(x) = [e_1 x_1, e_1 x_2],$$
$$\Psi_e(x) = [e_1 + x_1 - e_1 x_1, e_1 + x_2 - e_1 x_2],$$

are increasing bijections from $L^I$ to respectively $E$ and $E'$ which have an increasing inverse. If $e \in L^I \setminus D$, then there does not exist an increasing bijection from $L^I$ to $E$ or $E'$ with an increasing inverse [16].

**Theorem 13.** [16] *Let $\mathcal{U}$ be a uninorm on $\mathcal{L}^I$ with neutral element $e \in D \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$. Then there exist a t-norm $\mathcal{T}_\mathcal{U}$ and a t-conorm $\mathcal{S}_\mathcal{U}$ on $\mathcal{L}^I$ such that*

*(i)* $(\forall(x, y) \in E^2)(\mathcal{U}(x, y) = \Phi_e(\mathcal{T}_{\mathcal{U}}(\Phi_e^{-1}(x), \Phi_e^{-1}(y))))$,
*(ii)* $(\forall(x, y) \in E'^2)(\mathcal{U}(x, y) = \Psi_e(\mathcal{S}_{\mathcal{U}}(\Psi_e^{-1}(x), \Psi_e^{-1}(y))))$.

From Theorem 13 it follows that $\mathcal{U}$ displays downward reinforcement for $x$, $y$ in $E$ and upward reinforcement for $x$, $y$ in $E'$.

A uninorm $\mathcal{U}$ on $\mathcal{L}^I$ displays compensation behavior for $x \in E$ and $y \in E'$ or for $x \in E'$ and $y \in E$. Indeed, if $x \in E$ and $y \in E'$, then $\inf(x, y) = x = \mathcal{U}(x, e) \leq_{L^I} \mathcal{U}(x, y) \leq_{L^I} \mathcal{U}(e, y) = y = \sup(x, y)$.

In [20, 8, 19] a special class of uninorms on the unit interval is introduced: a uninorm on $([0, 1], \leq)$ is called representable if there exists a strictly increasing continuous mapping $h : [0, 1] \to \bar{\mathbb{R}}$ with $h(0) = -\infty$, $h(e) = 0$ and $h(1) = +\infty$ such that

$$U(x_1, y_1) = h^{-1}(h(x_1) + h(y_1)), \quad \forall(x_1, y_1) \in [0, 1]^2 \setminus \{(0, 1), (1, 0)\} .$$

The mapping $h$ is called a generator of $U$. They proved that a uninorm $U$ is representable if and only if $U$ is strictly decreasing and continuous in the open unit square and it is self-dual w.r.t. an involutive negation that has the neutral element of $U$ as fixed point. In [16] this necessary and sufficient condition is used to define representable uninorms on $\mathcal{L}^I$.

**Definition 19.** [16] *A uninorm $\mathcal{U}$ on $\mathcal{L}^I$ with neutral element $e$ is called representable if*

*(i)* *$\mathcal{U}$ is strictly decreasing and continuous in $(L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\})^2$;*
*(ii)* *there exists an involutive negation $\mathcal{N}_{\mathcal{U}}$ on $\mathcal{L}^I$ with fixed point $e$, i.e. $\mathcal{N}_{\mathcal{U}}(e) = e$, such that*

$$\Big(\forall(x, y) \in (L^I)^2 \setminus \{(0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}), (1_{\mathcal{L}^I}, 0_{\mathcal{L}^I})\}\Big)$$
$$\Big(\mathcal{U}(x, y) = \mathcal{N}_{\mathcal{U}}(\mathcal{U}(\mathcal{N}_{\mathcal{U}}(x), \mathcal{N}_{\mathcal{U}}(y)))\Big).$$

We consider an additional condition for a uninorm $\mathcal{U}$ on $\mathcal{L}^I$:

$$(\forall x \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\})(\forall Z \mid \varnothing \subset Z \subseteq L^I)$$
$$\left(\sup Z <_{L^I} 1_{\mathcal{L}^I} \Rightarrow \sup_{z \in Z} \mathcal{U}(x, z) = \mathcal{U}(x, \sup Z)\right) \tag{5}$$

It is shown in [16] that a similar condition for uninorms $U$ on the unit interval is equivalent to the left-continuity of $U$ in $[0, 1]^2$, and therefore holds for all representable uninorms on $([0, 1], \leq)$.

From Theorem 4.2, Lemma 6.5 and Theorem 6.7 in [16] it easily follows that the following holds.

**Theorem 14.** *Let $\mathcal{U}$ be a representable uninorm on $\mathcal{L}^I$ with neutral element $e \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$, satisfying* (5) *and $\mathcal{U}(D, D) \subseteq D$. Then there exists a representable uninorm $U$ on* $([0, 1], \leq)$ *such that, for all* $(x, y) \in (L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\})^2$,

$$\mathcal{U}(x, y) = [U(x_1, y_1), U(x_2, y_2)].$$

From Theorem 14 it follows that a representable uninorm $\mathcal{U}$ on $\mathcal{L}^I$ which satisfies (5) and $\mathcal{U}(D, D) \subseteq D$ can be represented by means of a strictly increasing continuous mapping $h : [0, 1] \to \bar{\mathbb{R}}$ for which $h(0) = -\infty$, $h(e) = 0$ and $h(1) = +\infty$: for all $x, y$ in $L^I$ such that neither $(x, y)$ or $(y, x)$ is an element of $(\{0_{\mathcal{L}^I}\} \times \{[a, 1] \mid a \in [0, 1]\}) \cup (\{[0, b] \mid b \in [0, 1]\} \times \{1_{\mathcal{L}^I}\})$,

$$\mathcal{U}(x, y) = [h^{-1}(h(x_1) + h(y_1)), h^{-1}(h(x_2) + h(y_2))].$$

## 7 OWA Operators on Interval-valued Fuzzy Sets

The OWA (ordered weighted average) operator was introduced by Yager [42] as a method for aggregating a set of $M$ crisp input values $a_i \in \mathbb{R}$, $i \in \{1, 2, \ldots, M\}$. Let

$$a_{(i)} = a_j, \text{ if } a_j \text{ is the } i\text{-th largest value.} \tag{6}$$

Let $(w_1, \ldots, w_M) \in [0, 1]^M$ be a vector of weights, for which $\sum_{i=1}^M w_i = 1$. Then the mapping $f : \mathbb{R}^M \to \mathbb{R}$ defined as

$$f(a_1, \ldots, a_M) = \sum_{i=1}^{M} w_i a_{(i)},$$

is called an OWA operator.

This operator can be extended to a fuzzy OWA operator, which can be used to aggregate $M$ continuous fuzzy sets $A_i$, $i \in \{1, \ldots, M\}$, on the real line [36, 37, 38, 40]. The corresponding "ordered" fuzzy sets $A_{(i)}$ are defined by applying Zadeh's extension principle to (6): for all $b \in \mathbb{R}$,

$$A_{(i)}(b) = \max\big\{\min\big(A_1(a_1), A_2(a_2), \ldots, A_M(a_M)\big) \mid (a_1, \ldots, a_M) \in \mathbb{R}^M \text{ and } b = a_{(i)}\big\}. \tag{7}$$

Define for each $\alpha \in ]0, 1]$, the $\alpha$-cut of a fuzzy set $A$ on $\mathbb{R}$ as $A_\alpha = \{x \mid x \in \mathbb{R} \text{ and } A(x) \geq \alpha\}$, and define $A_0 = \{x \mid x \in \mathbb{R} \text{ and } A(x) > 0\}$. Then, since $A_j$ is continuous for all $j \in \{1, \ldots, M\}$, $b \in (A_{(i)})_\alpha$ if and only if there exist $(a_1, \ldots, a_M) \in \mathbb{R}^M$ such that $b = a_{(i)}$ and $a_j \in (A_j)_\alpha$ for all $j \in \{1, \ldots, M\}$. If for all $j \in \{1, \ldots, M\}$, $(A_j)_\alpha = [\underline{A}_j, \overline{A}_j] \subseteq \mathbb{R}$, then $(A_{(i)})_\alpha = [\underline{A}_{(i)}, \overline{A}_{(i)}]$, i.e. the lower bound of $(A_{(i)})_\alpha$ is equal to the $i$-th largest lower bound of the $(A_j)_\alpha$, $j \in \{1, \ldots, M\}$, and similarly for the upper bound.

A fuzzy OWA operator $F$ is then defined by $F(A_1, \ldots, A_M) = \sum_{i=1}^M w_i A_{(i)}$, where the addition of fuzzy sets and the multiplication with a real number is calculated pointwise. Note that for any $M$-tuple $(A_1, \ldots, A_M)$ of continuous fuzzy sets on $\mathbb{R}$, $F(A_1, \ldots, A_M)$ is a fuzzy set on $\mathbb{R}$.

In [35] this operator is extended to an interval-valued fuzzy OWA operator. Given $M$ continuous interval-valued fuzzy sets $A_i : \mathbb{R} \to L^I$, $i \in \{1, 2, \ldots, M\}$, we have to define the corresponding "ordered" interval-valued fuzzy sets $A_{(i)}$, $i \in \{1, 2, \ldots, M\}$. The simplest approach would be to separately calculate for all $b \in \mathbb{R}$,

$$\begin{aligned} f^w_{(i)}(b) = \max\big\{&\min\big((A_1(a_1))_1, (A_2(a_2))_1, \ldots, (A_M(a_M))_1\big) \\ &\mid (a_1, \ldots, a_M) \in \mathbb{R}^M \text{ and } b = a_{(i)}\big\}, \end{aligned} \tag{8}$$

$$\begin{aligned} g^w_{(i)}(b) = \max\big\{&\min\big((A_1(a_1))_2, (A_2(a_2))_2, \ldots, (A_M(a_M))_2\big) \\ &\mid (a_1, \ldots, a_M) \in \mathbb{R}^M \text{ and } b = a_{(i)}\big\}, \end{aligned} \tag{9}$$

i.e. we apply (7) separately to the lower and the upper bounds of $A_i$, $i \in \{1, \ldots, M\}$. In [35] it is shown that for any $M$-tuple $(A_1, \ldots, A_M)$ of continuous interval-valued fuzzy sets, the mappings $f^w_{(i)}$ and $g^w_{(i)}$ obtained by using (8) and (9) are weakly coupled, i.e. they satisfy the inequalities $0 \le f^w_{(i)}(b) \le g^w_{(i)}(b) \le 1$, for all $b \in \mathbb{R}$. Hence the mapping $A_{(i)} : \mathbb{R} \to L^I$ defined by, for all $b \in \mathbb{R}$,

$$A_{(i)}(b) = [f^w_{(i)}(b), g^w_{(i)}(b)],$$

is an interval-valued fuzzy set.

In some circumstances we may require that for each $b \in \mathbb{R}$, $f_{(i)}(b)$ and $g_{(i)}(b)$ are strongly coupled, i.e. they depend on the lower and upper bounds of the same input interval-valued fuzzy set $A_j$ with $j \in \{1, \ldots, M\}$. Consider the set $\big\{\{a^{(1)}_1, \ldots, a^{(1)}_M\}, \{x^{(2)}_1, \ldots, x^{(2)}_M\}, \ldots, \{x^{(k)}_1, \ldots, x^{(k)}_M\}, \ldots\big\}$ of different $M$-tuples for which $b = x^{(k)}_{(i)}$, $k \in \{1, 2, \ldots\}$. Define for any $k \in \{1, 2, \ldots\}$,

$$\begin{aligned} f^{(k)}_{(i)}(b) &= \min((A_1(a^{(k)}_1))_1, \ldots, (A_M(a^{(k)}_M))_1), \\ g^{(k)}_{(i)}(b) &= \min((A_1(a^{(k)}_1))_2, \ldots, (A_M(a^{(k)}_M))_2). \end{aligned}$$

Then $f^w_{(i)}(b) = \max_k(f^{(k)}_{(i)}(b))$ and $g^w_{(i)}(b) = \max_k(g^{(k)}_{(i)}(b))$. In order to obtain strongly coupled results, we have to find an index $k_s$ such that

$$f^s_{(i)}(b) = f^{(k_s)}_{(i)}(b), \qquad g^s_{(i)}(b) = g^{(k_s)}_{(i)}(b).$$

In [35] the following criteria for choosing $k_s$ are discussed.

- Maximize $f^{(k)}_{(i)}(b)$. Then $k_s$ is the index for which $f^{(k_s)}_{(i)}(b) = \max_k\big(f^{(k)}_{(i)}(b)\big)$. This criterion may be regarded as "pessimistic", since it only takes into account the lower bounds of the intervals. Note that in this case $f^s_{(i)}(b) = f^w_{(i)}(b)$ and the upper bound function $g^s_{(i)}$ is completely determined by the lower bound function $f^s_{(i)}$: for any $b \in \mathbb{R}$, $f^s_{(i)}(b)$ is obtained by applying (7) to the lower bounds of

$A_i$, $i \in \{1, \ldots, M\}$; and if $f^s_{(i)}(b) = f^{(k_s)}_{(i)}(b)$ for some index $k_s$, then $g^s_{(i)}(b) = g^{(k_s)}_{(i)}(b)$.

- Maximize $g^{(k)}_{(i)}(b)$. Then $k_s$ is the index for which $g^{(k_s)}_{(i)}(b) = \max_k\big(g^{(k)}_{(i)}(b)\big)$. This criterion may be regarded as "optimistic", since it only takes into account the upper bounds of the intervals. Note that in this case $g^s_{(i)}(b) = g^w_{(i)}(b)$ and the lower bound function $f^s_{(i)}$ is completely determined by the upper bound function $g^s_{(i)}$.
- Maximize the sum $f^{(k)}_{(i)}(b) + g^{(k)}_{(i)}(b)$. Then $k_s$ is the index for which $f^{(k_s)}_{(i)}(b) + g^{(k_s)}_{(i)}(b) = \max_k\big(f^{(k)}_{(i)}(b) + g^{(k)}_{(i)}(b)\big)$. This criterion may be regarded as "neutral", since it gives equal weight to both the lower and the upper bounds of the intervals.

Now we can define the interval-valued fuzzy OWA operator, which is an extension of the fuzzy OWA operator, as follows: for any $M$-tuple $(A_1, \ldots, A_M)$ of continuous interval-valued fuzzy sets,

$$F(A_1, \ldots, A_M) = \sum_{i=1}^{M} w_i A_{(i)},$$

where the addition of interval-valued fuzzy sets and the multiplication with a real number is performed componentwise, i.e. for all $x \in \mathbb{R}$,

$$(F(A_1, \ldots, A_M)(x))_1 = \sum_{i=1}^{M} w_i (A_{(i)}(x))_1,$$

and similarly for the second component. Note that for any $M$-tuple $\{A_1, \ldots, A_M\}$ of continuous interval-valued fuzzy sets on $\mathbb{R}$, $F(A_1, \ldots, A_M)$ is an interval-valued fuzzy set on $\mathbb{R}$.

For examples and applications of interval-valued fuzzy OWA operators, we refer to [35].

## 8 Aggregation of Interval-valued Fuzzy Values Based on S-OWA Operators

Yager [42] introduced a measure on the OWA weights characterizing an OWA operator on $([0, 1], \leq)$, namely the orness, as follows:

$$\text{orness}(w_1, \ldots, w_M) = \frac{1}{M-1} \sum_{i=1}^{M} (M-i) w_i.$$

The dual measure, the andness, is defined as $\text{andness}(w_1, \ldots, w_M) = 1 - \text{orness}(w_1, \ldots, w_M)$. The orness and andness may be seen as the degree to which the operator obtained by the OWA weights represents, respectively, max ("or") and min ("and").

E.g. orness$(1, 0, \ldots, 0) = 1$, and using these weights we obtain the OWA operator $F(a_1, \ldots, a_M) = a_{(1)} = \max(a_1, \ldots, a_M)$.

Yager and Filev [43] introduced two families of OWA-operators: the 'orlike' and the 'andlike' S-OWA operators. The 'orlike' S-OWA operators, denoted $F_{SO}$, are defined by a family of OWA weights such that

$$w_i = \begin{cases} \frac{1}{M}(1-\alpha)+\alpha, & \text{if } i = 1, \\ \frac{1}{M}(1-\alpha), & \text{else}, \end{cases}$$

where $\alpha \in [0, 1]$. Then, for all $(a_1, \ldots, a_M) \in [0, 1]^M$,

$$F_{SO}(a_1, \ldots, a_M) = \alpha \max(a_1, \ldots, a_M) + \frac{1}{M}(1-\alpha)\sum_{i=1}^{M} a_i.$$

The orness measure of this aggregation operator is given by orness$(F_{SO}) = \frac{1}{2}(\alpha + 1)$. Note that for all $\alpha \in [0, 1]$, orness$(F_{SO}) \in [0.5, 1]$.

The 'andlike' S-OWA operators, denoted $F_{SA}$, are defined by a family of OWA weights such that

$$w_i = \begin{cases} \frac{1}{M}(1-\beta)+\beta, & \text{if } i = M; \\ \frac{1}{M}(1-\beta), & \text{else}, \end{cases}$$

where $\beta \in [0, 1]$. Then, for all $(a_1, \ldots, a_M) \in [0, 1]^M$,

$$F_{SA}(a_1, \ldots, a_M) = \beta \min(a_1, \ldots, a_M) + \frac{1}{M}(1-\beta)\sum_{i=1}^{M} a_i.$$

The orness measure of this aggregation operator is given by orness$(F_{SA}) = \frac{1}{2}(1 - \beta)$. Note that for all $\beta \in [0, 1]$, orness$(F_{SA}) \in [0, 0.5]$.

Wang et al. [41] defined the andlike S-OWA aggregation $\mathcal{F}_{SA}(a_1, \ldots, a_M)$ of $M$ elements $a_i$ $(i \in \{1, \ldots, M\})$ of $L^I$ as follows:

$$\mathcal{F}_{SA}(a_1, \ldots, a_M) = [F_{SA}((a_1)_1, \ldots, (a_M)_1), F_{SA}((a_1)_2, \ldots, (a_M)_2)],$$

where $F_{SA}$ is an andlike S-OWA aggregation operator on $([0, 1], \leq)$. They also define the andlike S-OWA aggregation of $M$ interval-valued fuzzy sets $A_i$ $(i \in \{1, \ldots, M\})$ on $\mathbb{R}$ as follows: for any $x \in \mathbb{R}$,

$$\begin{aligned} \mathcal{F}_{SA}(A_1, \ldots, A_M)(x) = [&F_{SA}((A_1(x))_1, \ldots, (A_M(x))_1), \\ &F_{SA}((A_1(x))_2, \ldots, (A_M(x))_2)]. \end{aligned}$$

The orlike S-OWA aggregation $\mathcal{F}_{SO}$ of $M$ elements of $L^I$ and the orlike S-OWA aggregation of $M$ interval-valued fuzzy sets are defined in a similar way.

Clearly, the andlike and orlike S-OWA aggregation of elements of $L^I$ are representable aggregation operators on $\mathcal{L}^I$.

### 8.1 Importance Weighted S-OWA Aggregation

In many instances of multi-criteria aggregation, the criteria differ in importance to the decision maker. We assume that the importance measure is modelled by a vector $V = (v_1, \ldots, v_M)$ such that $v_i \in [0, 1]$, for all $i \in \{1, \ldots, M\}$, and $\max(v_1, \ldots, v_M) = 1$.

Larsen [33] presented importance weighted OWA aggregation as follows: let $F$ be an OWA operator with weights $(w_1, \ldots, w_M)$, then the importance weighted OWA aggregation of $M$ elements $a_i$ of $[0, 1]$ is given by

$$F^V(a_1, \ldots, a_M) = F(b_1, \ldots, b_M) = \sum_{i=1}^{M} w_i b_{(i)},$$

where $b_i = \gamma + v_i(a_i - \gamma)$, where $\gamma$ denotes the andness of $F$.

Wang et al. [41] extended the notion of importance weighted OWA aggregation to $\mathcal{L}^I$ as follows.

**Definition 20.** *Let $a_1, \ldots, a_M$ be elements of $L^I$. Given a value of orness $\lambda$, with $\lambda \in [0, 0.5]$, and let $F_{SA}$ be the andlike S-OWA operator on $([0, 1], \leq)$ with orness $\lambda$. Then the importance weighted andlike S-OWA operator is defined as,*

$$\mathcal{F}_{SA}^V(a_1, \ldots, a_M) = F_{SA}(b_1, \ldots, b_M) = \sum_{i=1}^{M} w_i b_{(i)},$$

*where $b_i = 1 - \lambda + v_i(S(a_i) - 1 + \lambda)$, where $S : L^I \to [0, 1]$ is the score function defined by $S(x) = x_1 + x_2$, for all $x \in L^I$.*

The importance weighted orlike S-OWA operator is defined in a similar way. An application of these operators to fuzzy multi-criteria decision making can be found in [41].

## 9 Conclusion and Future Work

In this paper we have discussed aggregation operators in interval-valued fuzzy and Atanassov's intuitionistic fuzzy set theory. We have given an overview of the already obtained results on these operators. First we have discussed some properties of t-norms on the underlying lattice $\mathcal{L}^I$ of interval-valued fuzzy set theory and we investigated their additive and multiplicative generators.

Another class of aggregation operators we discussed are uninorms, in particular the class of representable uninorms is studied. We have shown that representable uninorms can be represented by means of a generator of a uninorm on the unit interval. In the future we will introduce generators of uninorms on $\mathcal{L}^I$ which are mappings on $L^I$ instead of on [0, 1].

The OWA operators introduced by Yager can be extended to interval-valued fuzzy sets. Mitchell [35] introduced two kinds of interval-valued fuzzy OWA operators: the first kind is simply a t-representable extension of the fuzzy OWA operator introduced by the same author, in the second kind the two components of the interval-valued fuzzy OWA are related to each other. Wang [41] extended a special class of OWA operators introduced by Yager and Filev, the so-called S-OWA operators to interval-valued fuzzy sets. Their approach yields also t-representable aggregation operators.

In the future we will extend other aggregation operators to interval-valued fuzzy set theory. We will also try to find non-t-representable extensions and investigate whether these extensions have better properties than the t-representable ones. We will also study in general how the type of representability (t-representable, pseudo-t-representable, ...) influences which properties satisfied by an aggregation operator on the unit interval is inherited by its extension to $\mathcal{L}^I$.

## References


1. K. T. Atanassov. Intuitionistic fuzzy sets, 1983. VII ITKR's Session, Sofia (deposed in Central Sci.-Technical Library of Bulg. Acad. of Sci., 1697/84) (in Bulgarian).
2. K. T. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20(1):87–96, 1986.
3. K. T. Atanassov. *Intuitionistic fuzzy sets*. Physica-Verlag, Heidelberg/New York, 1999.
4. G. Birkhoff. *Lattice Theory*, Vol. 25. AMS Colloquium Publications, Providence, Rhode Island, 1973.
5. H. Bustince and P. Burillo. Vague sets are intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 79(3):403–405, 1996.
6. T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Aggregation Operators: New Trends and Applications*, pp. 3–104. Physica-Verlag, Heidelberg/New York, 2002.
7. T. Calvo and R. Mesiar. Continuous generated associative aggregation operators. *Fuzzy Sets and Systems*, 126(2):191–197, 2002.
8. B. De Baets and J. Fodor. On the structure of uninorms and their residual implicators. In S. Gottwald and E. P. Klement, editors, *Proceedings of the 18th Linz Seminar on Fuzzy Set Theory: Enriched Lattice Structures for Many-Valued and Fuzzy Logics*, pp. 81–87, 1997.
9. G. Deschrijver. Additive and multiplicative generators in interval-valued fuzzy set theory. *IEEE Transactions on Fuzzy Systems*. in press.
10. G. Deschrijver. Algebraic operators in interval-valued fuzzy set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. submitted.
11. G. Deschrijver. The Archimedean property for t-norms in interval-valued fuzzy set theory. *Fuzzy Sets and Systems*. in press.
12. G. Deschrijver. Generators of t-norms in interval-valued fuzzy set theory. In *Proceedings of the 4th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2005)*, pp. 253–258, 2005.

13. G. Deschrijver, C. Cornelis, and E. E. Kerre. On the representation of intuitionistic fuzzy t-norms and t-conorms. *IEEE Transactions on Fuzzy Systems*, 12(1):45–61, 2004.
14. G. Deschrijver and E. E. Kerre. Classes of intuitionistic fuzzy t-norms satisfying the residuation principle. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(6):691–709, 2003.
15. G. Deschrijver and E. E. Kerre. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets and Systems*, 133(2):227–235, 2003.
16. G. Deschrijver and E. E. Kerre. Uninorms in $L^*$-fuzzy set theory. *Fuzzy Sets and Systems*, 148(2):243–262, 2004.
17. G. Deschrijver and E. E. Kerre. Implicators based on binary aggregation operators in interval-valued fuzzy set theory. *Fuzzy Sets and Systems*, 153(2):229–248, 2005.
18. M. Eisenberg. *Axiomatic theory of sets and classes*. Holt, Rinehart and Winston, Inc., New York, 1971.
19. J. Fodor and B. De Baets. On uninorms and their residual implicators. In P. Sinčák, J. Vaščák, V. Kvasnička, and J. Pospíchal, editors, *Intelligent Technologies - Theory and Applications*, pp. 91–97. IOS Press, 2002.
20. J. Fodor, R. Yager, and A. Rybalov. Structure of uninorms. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems*, 5:411–427, 1997.
21. J. C. Fodor and S. Ovchinnikov. On aggregation of T-transitive fuzzy binary relations. *Fuzzy Sets and Systems*, 72(2):135–145, 1995.
22. J. C. Fodor and M. Roubens. *Fuzzy preference modelling and multicriteria decision support*. Kluwer Academic Publishers, Dordrecht, 1994.
23. W. L. Gau and D. J. Buehrer. Vague sets. *IEEE Trans. Systems Man. Cybernet.*, 23(2): 610–614, 1993.
24. J. Goguen. $L$-fuzzy sets. *Journal of Mathematical Analysis and Applications*, 18(1):145–174, 1967.
25. M. B. Gorzałczany. A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, 21(1):1–17, 1987.
26. M. Grabisch. Fuzzy integral in multicriteria decision making. *Fuzzy Sets and Systems*, 69(3):279–298, 1995.
27. E. P. Klement, R. Mesiar, and E. Pap. On the relationship of associative compensatory operators to triangular norms and conorms. *Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4:129–144, 1996.
28. E. P. Klement, R. Mesiar, and E. Pap. Quasi- and pseudo-inverses of monotone functions, and the construction of t-norms. *Fuzzy Sets and Systems*, 104(1):3–13, 1999.
29. E. P. Klement, R. Mesiar, and E. Pap. *Triangular norms*. Kluwer Academic Publishers, Dordrecht, 2002.
30. E. P. Klement, R. Mesiar, and E. Pap. Triangular norms. Position paper I: basic analytical and algebraic properties. *Fuzzy Sets and Systems*, 143(1):5–26, 2004.
31. E. P. Klement, R. Mesiar, and E. Pap. Triangular norms. Position paper II: general constructions and parametrized families. *Fuzzy Sets and Systems*, 145(3):411–438, 2004.
32. A. Kolesárová and M. Komorníková. Triangular norm-based iterative compensatory operators. *Fuzzy Sets and Systems*, 104(1):109–120, 1999.
33. H. L. Larsen. Importance weighted OWA aggregation of multicriteria queries. In *Proceedings of the North American Fuzzy Information Processing Society Conference (NAFIPS'99)*, pp. 740–744, New York, 1999.
34. C.-H. Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212, 1965.
35. H. B. Mitchell. An intuitionistic OWA operator. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(6):843–860, 2004.
36. H. B. Mitchell and D. D. Estrakh. An OWA operator with fuzzy ranks. *International Journal of Intelligent Systems*, 13:69–82, 1998.
37. H. B. Mitchell and P. A. Schaefer. Multiple priorities in an induced ordered weighted averaging operator. *International Journal of Intelligent Systems*, 15:317–328, 2000.

38. H. B. Mitchell and P. A. Schaefer. On ordering fuzzy numbers. *International Journal of Intelligent Systems*, 15:981–994, 2000.
39. R. Sambuc. *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*. Ph. D. Thesis, Université de Marseille, France, 1975.
40. P. A. Schaefer and H. B. Mitchell. A generalized OWA operator. *International Journal of Intelligent Systems*, 14:123–144, 1999.
41. J. Wang, S. Y. Liu, J. Zhang, and S. Y. Wang. On the parametrized OWA operators for fuzzy MCDM based on vague set theory. *Fuzzy Optimization and Decision Making*, 5:5–20, 2006.
42. R. Yager. On ordered weighted averaging operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.
43. R. Yager and D. Filev. Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 29:141–150, 1999.
44. R. Yager and A. Rybalov. Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80(1): 111–120, 1996.
45. R. Yager and A. Rybalov. Full reinforcement operators in aggregation techniques. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 28(6):757–769, 1998.
46. H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4(1):37–51, 1980.

# Part II
# From Decision Making to Data Mining, Web Intelligence and Computer Vision

# Fuzzy Preference Modelling: Fundamentals and Recent Advances

**János Fodor*** **and Bernard de Baets**

**Abstract** The construction of fuzzy strict preference, indifference and incomparability relations from a fuzzy large preference relation is usually cast into an axiomatic framework based on t-norms. In this contribution, we show that this construction is essentially characterized by the choice of an indifference generator, a symmetrical mapping located between the Łukasiewicz t-norm and the minimum operator. Interesting constructions are obtained by choosing as indifference generator a commutative quasi-copula, an ordinal sum of Frank t-norms or a particular Frank t-norm.

## 1 Introduction

The study of fuzzy preference structures has a longstanding tradition and is characterized by a peculiar historical development [3]. Roughly speaking, three phases can be distinguished. In the first phase, different proposals for constructing strict preference and indifference relations emerge. In the second one, this construction is approached in an axiomatic way, by imposing conditions on the fuzzy strict preference, indifference and incomparability relations, leading to functional equations identifying suitable strict preference, indifference and incomparability generators. Surprisingly, the definition of a fuzzy preference structure was only given in the third phase, leading to the study of additive fuzzy preference structures. We show that a given additive fuzzy preference structure is not necessarily the result of applying monotone generators to a large preference relation. In order to cover all additive fuzzy preference structures, we therefore start all over again, looking for the most general strict preference, indifference and incomparability generators. We pinpoint the central role of the indifference generator and clarify that the monotonicity of a generator triplet is totally determined by using a commutative quasi-copula as indifference generator.

* J. Fodor has been supported in part by OTKA T046762.

## 2 Boolean Preference Structures

Preference modelling is a fundamental step of (multi criteria) decision making, operations research, social choice and voting procedures, and has been studied extensively for several years. Its most basic concept is that of a **preference structure**. Consider a set of alternatives $A$ and suppose that a decision maker wants to judge these alternatives by pairwise comparison. Given two alternatives, the decision maker can act in one of the following three ways:

(i) he/she clearly prefers one to the other;
(ii) the two alternatives are indifferent to him/her;
(iii) he/she is unable to compare the two alternatives.

According to these cases, three binary relations can be defined in $A$: the **strict preference relation** $P$, the **indifference relation** $I$ and the **incomparability relation** $J$. For any $(a, b) \in A^2$, we classify:

$$\begin{aligned} (a,b) \in P \quad &\Leftrightarrow \quad \text{he/she prefers } a \text{ to } b;\\ (a,b) \in I \quad &\Leftrightarrow \quad a \text{ and } b \text{ are indifferent to him/her;}\\ (a,b) \in J \quad &\Leftrightarrow \quad \text{he/she is unable to compare } a \text{ and } b. \end{aligned}$$

One easily verifies that the triplet $(P, I, J)$ defined above satisfies the conditions formulated in the following definition of a preference structure. For a binary relation $R$ in $A$, we denote its **converse** by $R^t$ and its **complement** by $\operatorname{co} R$.

**Definition 1.** [11] *A preference structure on $A$ is a triplet $(P, I, J)$ of binary relations in $A$ that satisfy:*

(B1) *$P$ is irreflexive, $I$ is reflexive and $J$ is irreflexive;*
(B2) *$P$ is asymmetrical, $I$ is symmetrical and $J$ is symmetrical;*
(B3) *$P \cap I = \emptyset$, $P \cap J = \emptyset$ and $I \cap J = \emptyset$;*
(B4) *$P \cup P^t \cup I \cup J = A^2$.*

This definition is exhaustive: it lists all properties of the components $P$, $I$ and $J$ of a preference structure. The asymmetry of $P$ can also be written as $P \cap P^t = \emptyset$. Condition (B4) is called the **completeness condition** and can be expressed equivalently (up to symmetry) in the following alternative ways: $\operatorname{co}(P \cup I) = P^t \cup J$, $\operatorname{co}(P \cup P^t) = I \cup J$, $\operatorname{co}(P \cup P^t \cup I) = J$, $\operatorname{co}(P \cup P^t \cup J) = I$ and $\operatorname{co}(P^t \cup I \cup J) = P$.

It is possible to associate a single reflexive relation to any preference structure so that it completely characterizes this structure. A preference structure $(P, I, J)$ on $A$ is characterized by the reflexive binary relation $R = P \cup I$, its large preference relation, in the following way:

$$(P, I, J) = (R \cap \operatorname{co} R^t, R \cap R^t, \operatorname{co} R \cap \operatorname{co} R^t).$$

Conversely, a triplet $(P, I, J)$ constructed in this way from a reflexive binary relation $R$ in $A$ is a preference structure on $A$. The interpretation of the large preference relation is

$$(a, b) \in R \quad \Leftrightarrow \quad b \text{ is considered at most as good as } a.$$

The above definition of a preference structure can be written in the following minimal way, identifying a relation with its characteristic mapping [4]: $I$ is reflexive and symmetrical, and for any $(a, b) \in A^2$:

$$P(a, b) + P^t(a, b) + I(a, b) + J(a, b) = 1.$$

Classical preference structures can therefore also be considered as **Boolean preference structures**, employing 1 and 0 for describing presence or absence of strict preferences, indifferences and incomparabilities. Complement, intersection and union then correspond to Boolean negation, conjunction (i.e. minimum) and disjunction (i.e. maximum) on characteristic mappings.

## 3 Beyond Booleanity

As preference structures are based on classical set theory and are therefore restricted to two-valued relations, they do not allow to express degrees of strict preference, indifference or incomparability. This is seen as an important drawback to their practical use, leading researchers already at an early stage to the theory of fuzzy sets, and in particular to the calculus of fuzzy relations. In that case, preference degrees are expressed on the continuous scale [0, 1] and operations from fuzzy logic are used for manipulating these degrees. The notion of a binary **fuzzy relation** $R$ in $A$ appears first in the seminal paper of Zadeh [37] and is simply defined as an $A^2 \to [0, 1]$ mapping.

In that context, it is essential to have access to suitable operators for combining the degrees of preference. In this paper, we are mainly interested in two classes of operators: the class of t-norms [16] and the class of (quasi-)copulas [7, 10].

**Definition 2.** *A binary operation* $T : [0, 1]^2 \to [0, 1]$ *is called a t-norm if it satisfies:*

*(i) Neutral element* 1*:* $(\forall x \in [0, 1])(T(x, 1) = T(1, x) = x)$.
*(ii) Monotonicity: $T$ is increasing in each variable.*
*(iii) Commutativity:* $(\forall (x, y) \in [0, 1]^2)(T(x, y) = T(y, x))$.
*(iv) Associativity:* $(\forall (x, y, z) \in [0, 1]^3)(T(x, T(y, z)) = T(T(x, y), z))$.

T-conorms are the dual operations of t-norms, in the sense that for a given t-norm $T$, the operation $S : [0, 1]^2 \to [0, 1]$ defined by

$$S(x, y) = 1 - T(1 - x, 1 - y)\,,$$

is a t-conorm. Formally, the only difference between t-conorms and t-norms is that the former have neutral element 0, while the latter have neutral element 1.

**Definition 3.** *A binary operation* $C : [0,1]^2 \to [0,1]$ *is called a quasi-copula if it satisfies:*

- *(i)* *Neutral element* 1*:* $(\forall x \in [0,1])(C(x,1) = C(1,x) = x)$.
- *(i')* *Absorbing element* 0*:* $(\forall x \in [0,1])(C(x,0) = C(0,x) = 0)$.
- *(ii)* *Monotonicity: C is increasing in each variable.*
- *(iii)* *1-Lipschitz property:* $(\forall (x_1, x_2, y_1, y_2) \in [0,1]^4)$

$$(|C(x_1, y_1) - C(x_2, y_2)| \leq |x_1 - x_2| + |y_1 - y_2|) .$$

*If instead of (iii), C satisfies*

- *(iv)* *Moderate growth:* $(\forall (x_1, x_2, y_1, y_2) \in [0,1]^4)$

$$((x_1 \leq x_2 \wedge y_1 \leq y_2) \Rightarrow C(x_1, y_1) + C(x_2, y_2) \geq C(x_1, y_2) + C(x_2, y_1)) ,$$

*then it is called a copula.*

Note that in case of a quasi-copula condition (i') is superfluous. For a copula, condition (ii) can be omitted (as it follows from (iv) and (i')). As implied by the terminology used, any copula is a quasi-copula, and therefore has the 1-Lipschitz property; the opposite is, of course, not true. It is well known that a copula is a t-norm if and only if it is associative; conversely, a t-norm is a copula if and only if it is 1-Lipschitz. Finally, note that for any quasi-copula $C$ it holds that $T_{\mathrm{L}} \leq C \leq T_{\mathrm{M}}$, where $T_{\mathrm{L}}(x, y) = \max(x + y - 1, 0)$ is the Łukasiewicz t-norm and $T_{\mathrm{M}}(x, y) = \min(x, y)$ is the minimum operator.

We consider a **continuous De Morgan triplet** $(T, S, N)$ on $[0, 1]$, consisting of a continuous t-norm $T$, a strong negation $N$ (i.e. a decreasing permutation of $[0, 1]$) and the $N$-dual t-conorm $S$ defined by

$$S(x, y) = N(T(N(x), N(y))) .$$

Note that a strong negation is uniquely determined by the corresponding automorphism $\phi$ of the unit interval, $N_\phi(x) := \phi^{-1}(1 - \phi(x))$.

Further on, the complement of a fuzzy set $A$ is denoted by $\mathrm{co}_N(A)$; the intersection (resp. union) of two fuzzy sets $A$ and $B$ is denoted by $A \cap_T B$ (resp. $A \cup_S B$).

A **fuzzy preference structure (FPS)** on $A$ is a triplet $(P, I, J)$ of binary fuzzy relations in $A$ satisfying:

- (F1) $P$ is irreflexive, $I$ is reflexive and $J$ is irreflexive;
- (F2) $P$ is $T$-asymmetrical (i.e. $P \cap_T P^t = \emptyset$), $I$ and $J$ are symmetrical;
- (F3) $P \cap_T I = \emptyset$, $P \cap_T J = \emptyset$ and $I \cap_T J = \emptyset$;
- (F4) a completeness conditions, such as $\mathrm{co}_N\,(P \cup_S I) = P^t \cup_S J$, $\mathrm{co}_N\,(P \cup_S P^t \cup_S I) = J$ or $P \cup_S P^t \cup_S I \cup_S J = A^2$.

Invoking the **assignment principle**: *for any pair of alternatives* $(a, b)$ *the decision maker is allowed to assign at least one of the degrees* $P(a, b)$, $P(b, a)$, $I(a, b)$ *and* $J(a, b)$ *freely in the unit interval*, shows that only a **nilpotent t-norm** $T$ is acceptable, i.e. a $\phi'$-transform of the Łukasiewicz t-norm: $T(x, y) := \phi'^{-1}(\max(\phi'(x)+\phi'(y)-1, 0))$ [12]. For the sake of simplicity, we consider $\phi = \phi'$. Consequently, we will be working with a Łukasiewicz triplet $(T_\phi^\infty, S_\phi^\infty, N_\phi)$. The latter notation is used to indicate that the Łukasiewicz t-norm belongs to the Frank t-norm family $(T^s)_{s\in[0,\infty]}$ (which is also a family of copulas) and corresponds to the parameter value $s = \infty$ (note that the minimum operator and the algebraic product correspond to the parameter values $s = 0$ and $s = 1$, respectively). Moreover, in that case, the completeness conditions $\text{co}_\phi\,(P \cup_\phi^\infty I) = P^t \cup_\phi^\infty J$ and $\text{co}_\phi\,(P \cup_\phi^\infty P^t) = I \cup_\phi^\infty J$ become equivalent and turn out to be stronger than the other completeness conditions, with $P \cup_\phi^\infty P^t \cup_\phi^\infty I \cup_\phi^\infty J = A^2$ as weakest condition [12]. Restricting to the **strongest completeness condition(s)**, we then obtain the following definition. Given a [0, 1]-automorphism $\phi$, a $\phi$-FPS $(P, I, J)$ on $A$ is a triplet of binary fuzzy relations in $A$ satisfying:

(F1) $P$ is irreflexive, $I$ is reflexive and $J$ is irreflexive;
(F2) $P$ is $T_\phi^\infty$-asymmetrical (i.e. $P \cap_\phi^\infty P^t = \emptyset$), $I$ and $J$ are symmetrical;
(F3) $P \cap_\phi^\infty I = \emptyset$, $P \cap_\phi^\infty J = \emptyset$ and $I \cap_\phi^\infty J = \emptyset$;
(F4) $\text{co}_\phi\,(P \cup_\phi^\infty I) = P^t \cup_\phi^\infty J$.

Moreover, a minimal formulation of this definition, similar to the classical one, exists: a triplet $(P, I, J)$ of binary fuzzy relations in $A$ is a $\phi$-FPS on $A$ if and only if $I$ is reflexive and symmetrical, and for any $(a, b) \in A^2$:

$$\phi(P(a, b)) + \phi(P^t(a, b)) + \phi(I(a, b)) + \phi(J(a, b)) = 1.$$

In view of the above equality, fuzzy preference structures are also called **additive fuzzy preference structures**.

## 4 Axiomatic Constructions

Again choosing a continuous de Morgan triplet $(T, S, N)$, we could transport the classical construction formalism to the fuzzy case and define, given a reflexive binary fuzzy relation $R$ in $A$:

$$(P, I, J) = (R \cap_T \text{co}_N R^t, R \cap_T R^t, \text{co}_N R \cap_T \text{co}_N R^t).$$

At the same time, we want to keep $R$ as the fuzzy large preference relation of the triplet $(P, I, J)$, i.e. $R = P \cup_S I$ and $\text{co}_N R = P^t \cup_S J$. Fodor and Roubens observed that the latter is not possible in general, and proposed four axioms for defining fuzzy strict preference, indifference and incomparability relations [5, 6].

According to the first axiom, the **independence of irrelevant alternatives**, there exist three $[0, 1]^2 \to [0, 1]$ mappings $p, i, j$ such that $P(a, b) = p(R(a, b), R(b, a))$, $I(a, b) = i(R(a, b), R(b, a))$ and $J(a, b) = j(R(a, b), R(b, a))$. The second and third axiom state that the mappings $p(x, N(y))$, $i(x, y)$ and $j(N(x), N(y))$ are increasing in both $x$ and $y$, and that $i$ and $j$ are symmetrical. The fourth and main axiom requires that $P \cup_S I = R$ and $P^t \cup_S J = \text{co}_N R$, or explicitly, for any $(x, y) \in [0, 1]^2$:

$$S(p(x, y), i(x, y)) = x$$
$$S(p(x, y), j(x, y)) = N(y).$$

The latter axiom implies that $\text{co}_N (P \cup_S I) = P^t \cup_S J$, i.e. the first completeness condition.

**Theorem 1.** [5, 6] *If $(T, S, N)$ and $(p, i, j)$ satisfy the above axioms, then there exists a $[0, 1]$-automorphism $\phi$ such that*

$$(T, S, N) = (T_\phi^\infty, S_\phi^\infty, N_\phi)$$

*and, for any $(x, y) \in [0, 1]^2$:*

$$\begin{aligned} T_\phi^\infty(x, N_\phi(y)) \le \quad & p(x, y) \quad \le T^0(x, N_\phi(y)) \\ T_\phi^\infty(x, y) \le \quad & i(x, y) \quad \le T^0(x, y) \\ T_\phi^\infty(N_\phi(x), N_\phi(y)) \le \quad & j(x, y) \quad \le T^0(N_\phi(x), N_\phi(y)). \end{aligned}$$

*Moreover, for any reflexive binary fuzzy relation $R$ in $A$, the triplet $(P, I, J)$ of binary fuzzy relations in $A$ defined by*

$$P(a, b) = p(R(a, b), R(b, a))$$
$$I(a, b) = i(R(a, b), R(b, a))$$
$$J(a, b) = j(R(a, b), R(b, a))$$

*is a $\phi$-FPS on $A$ such that $R = P \cup_\phi^\infty I$ and $\text{co}_\phi R = P^t \cup_\phi^\infty J$.*

Although in general the function $i$ is of two variables, and there is no need to extend it for more than two arguments, it might be a t-norm. The following theorem states that the only construction methods of the above type based on continuous t-norms are the ones using two Frank t-norms with reciprocal parameters.

**Theorem 2.** [5, 6] *Consider a $[0, 1]$-automorphism $\phi$ and two continuous t-norms $T_1$ and $T_2$. Define $p$ and $i$ by $p(x, y) = T_1(x, N_\phi(y))$ and $i(x, y) = T_2(x, y)$. Then $(p, i, j)$ satisfies the above axioms if and only if there exists a parameter $s \in [0, \infty]$ such that $T_1 = T_\phi^{1/s}$ and $T_2 = T_\phi^s$. In this case, we have that $j(x, y) = i(N_\phi(x), N_\phi(y))$.*

Summarizing, we have that for any reflexive binary fuzzy relation $R$ in $A$ the triplets

$$(P_s, I_s, J_s) := (R \cap_\phi^{\frac{1}{s}} \mathrm{co}_\phi R^t, R \cap_\phi^s R^t, \mathrm{co}_\phi R \cap_\phi^s \mathrm{co}_\phi R^t),$$

with $s \in [0, \infty]$, are the only t-norm-based constructions of fuzzy preference structures that satisfy $R = P \cup_\phi^\infty I$ and $\mathrm{co}_\phi R = P^t \cup_\phi^\infty J$. Consequently, $R$ is again called the **large preference relation**. Note that

$$\phi(R(a, b)) = \phi(P(a, b)) + \phi(I(a, b)).$$

In fact, in [6] it was only shown that ordinal sums of Frank t-norms should be used. For the sake of simplicity, only the ordinally irreducible ones were considered. However, we can prove that this is the only option.

Finally, we deal with the **reconstruction** of a $\phi$-FPS from its large preference relation. As expected, an additional condition is required. A $\phi$-FPS $(P, I, J)$ on $A$ is called:

(i) an $(s, \phi)$-FPS, with $s \in \{0, 1, \infty\}$, if $P \cap_\phi^s P^t = I \cap_\phi^{\frac{1}{s}} J$;
(ii) an $(s, \phi)$-FPS, with $s \in ]0, 1[ \cup ]1, \infty[$, if

$$s^{\phi(P \cap_\phi^s P^t)} + s^{-\phi(I \cap_\phi^{1/s} J)} = 2.$$

One can verify that the triplet $(P_s, I_s, J_s)$ constructed above is an $(s, \phi)$-FPS. Moreover, any $(s, \phi)$-FPS can be reconstructed from its large preference relation by means of the corresponding construction. The characterizing condition of a $(0, \phi)$-FPS, resp. $(\infty, \phi)$-FPS, can also be written as $P \cap^0 P^t = \emptyset$, i.e. $\min(P(a, b), P(b, a)) = 0$ for any $(a, b)$, resp. $I \cap^0 J = \emptyset$, i.e. $\min(I(a, b), J(a, b)) = 0$ for any $(a, b)$.

## 5 Local Characteristic Behaviour and Non-monotonicity

An important question now arises: to what extent is the class of $\phi$-FPS covered by the classes of $(s, \phi)$-FPS, $s \in [0, \infty]$? We will show next that this covering is, at least locally, total. To that end, we need some additional propositions.

**Proposition 1.** [2] *Consider* $(a, b, c, d) \in [0, 1]^4$ *such that* $a + b + c + d = 1$ *and* $s \in ]0, 1[ \cup ]1, \infty[$, *then it holds that*

$$s^{T^s(a,b)} + s^{-T^{1/s}(c,d)} = 2$$

*if and only if*

$$s^a + s^b + s = s^{1-c} + s^{1-d} + 1.$$

**Proposition 2.** [2] *Consider* $(a, b, c, d) \in [0, 1]^4$ *such that* $a + b + c + d = 1$. *Let* $u = a + c$ *and* $v = b + c$*, then it holds that*

$$(a, b, c, d) = (T^{1/s}(u, 1 - v), T^{1/s}(1 - u, v), T^s(u, v), T^s(1 - u, 1 - v)),$$

*where*

(i) $s = 0$ *if* $\min(a, b) = 0$*;*
(ii) $s = \infty$ *if* $\min(c, d) = 0$*;*
(iii) $s = 1$ *if* $ab = cd$*;*
(iv) $s$ *is the unique solution in* $]0, 1[\cup]1, \infty[$ *of the equation*

$$s^a + s^b + s = s^{1-c} + s^{1-d} + 1$$

*if none of the above holds.*

The fact that under the given conditions the latter equation has a unique solution is shown in [2]. Note that the first three cases in the foregoing theorem are not mutually exclusive. We now obtain the following conclusive characterization of additive fuzzy preference structures.

**Theorem 3.** [1] *Consider a* $[0, 1]$*-automorphism* $\phi$ *and a* $\phi$*-FPS* $(P, I, J)$ *on* $A$*. For any two alternatives* $a$ *and* $b$ *in* $A$ *there exists* $s_{a,b} \in [0, \infty]$ *such that the restriction of* $(P, I, J)$ *to* $\{a, b\}^2$ *is an* $(s_{a,b}, \phi)$*-FPS on* $\{a, b\}$.

Hence, to any $\phi$-FPS on $A$ corresponds a (not necessarily unique) symmetrical matrix of Frank parameters. This theorem also has strong implications on the system of functional equations corresponding to the axioms of Fodor and Roubens. Indeed, the most general solution for $i$, for instance, is given by $i(x, y) = T_\phi^{s(x,y)}(x, y)$, where $s$ is a symmetrical $[0, 1]^2 \to [0, \infty]$ mapping assuring that $i$ is increasing. This is, for instance, the case when $s$ is decreasing.

There is, however, no reason why the mapping $i$ should be increasing. Consider the set of alternatives $A = \{a, b, c, d\}$ such that

(i) $P(a, b) = P(b, a) = 0.4$, $I(a, b) = 0.2$ and $J(a, b) = 0$;
(ii) $P(c, d) = P(d, c) = 0.2$, $I(c, d) = 0.3$ and $J(c, d) = 0.3$.

The first case implies that $i(0.6, 0.6) = 0.2$, while the second case implies that $i(0.5, 0.5) = 0.3$. Clearly, the mapping $i$ generating the fuzzy indifference relation cannot be increasing.

## 6 Generator Triplets

The above observations have motivated us to reconsider the construction of additive fuzzy preference structures, not by rephrasing the conclusions resulting from an axiomatic study, but by starting from the minimal definition of an additive fuzzy preference structure. For the sake of brevity, we consider the case $\phi(x) = x$.

**Definition 4.** *A triplet* $(p, i, j)$ *of* $[0, 1]^2 \to [0, 1]$ *mappings is called a* **generator triplet** *compatible with a continuous t-conorm S and a strong negator N if and only if for any reflexive binary fuzzy relation R on a set of alternatives A it holds that the triplet* $(P, I, J)$ *of binary fuzzy relations on A defined by:*

$$P(a, b) = p(R(a, b), R(b, a))$$
$$I(a, b) = i(R(a, b), R(b, a))$$
$$J(a, b) = j(R(a, b), R(b, a))$$

*is a FPS on A such that* $P \cup_S I = R$ *and* $P^t \cup_S J = \mathrm{co}_N R$.

The above conditions $P \cup_S I = R$ and $P^t \cup_S J = \mathrm{co}_N R$ require the reconstructability of the fuzzy large preference relation $R$ from the fuzzy preference structure it generates. The following theorem expresses that for that purpose only nilpotent t-conorms can be used.

**Theorem 4.** *If* $(p, i, j)$ *is a generator triplet compatible with a continuous t-conorm S and a strong negator* $N = N_\phi$*, then* $S = S_\psi^\infty$*, i.e. S is nilpotent.*

Let us again consider the case $\psi(x) = x$. The above theorem implies that we can omit the specification 'compatible with a continuous t-conorm $S$ and strong negation $N$' and simply talk about generator triplets. The minimal definition of a fuzzy preference structure then immediately leads to the following proposition.

**Proposition 3.** *A triplet* $(p, i, j)$ *is a generator triplet if and only if, for any* $(x, y) \in [0, 1]^2$*:*

(i) $i(1, 1) = 1$;
(ii) $i(x, y) = i(y, x)$;
(iii) $p(x, y) + p(y, x) + i(x, y) + j(x, y) = 1$;
(iv) $p(x, y) + i(x, y) = x$.

From this proposition it follows that a generator triplet is uniquely determined by, for instance, the generator $i$. Indeed, for any generator triplet $(p, i, j)$ it holds that

$$p(x, y) = x - i(x, y)$$
$$j(x, y) = i(x, y) - (x + y - 1).$$

The fact that $p$ and $j$ take values in $[0, 1]$ implies that $T^\infty \leq i \leq T^0$. Moreover, from any symmetrical $i$ such that $T^\infty \leq i \leq T^0$ a generator triplet can be built. It is therefore not surprising that additional properties of generator triplets $(p, i, j)$ are completely determined by additional properties of $i$. In fact, in practice it would be sufficient to talk about a single generator $i$. We could simply talk about **the generator** of the FPS. Note that the symmetry of $i$ implies the symmetry of $j$.

Firstly, we try to characterize generator triplets fitting into the axiomatic framework of Fodor and Roubens.

**Definition 5.** *A generator triplet* $(p, i, j)$ *is called* **monotone** *if:*

(i) *$p$ is increasing in the first and decreasing in the second argument;*
(ii) *$i$ is increasing in both arguments;*
(iii) *$j$ is decreasing in both arguments.*

Inspired by a recent paper [9], we can show that monotone generator triplets are characterized by a 1-Lipschitz indifference generator, i.e. by a commutative quasi-copula.

**Theorem 5.** *A generator triplet* $(p, i, j)$ *is monotone if and only if* $i$ *is a commutative quasi-copula.*

The following theorem shows that when $i$ is a symmetrical ordinal sum of Frank t-norms, $j(1-x, 1-y)$ is also a t-norm, and $p(x, 1-y)$ is symmetrical. Note that by symmetrical ordinal sum we mean the following: if $(a, b, T)$ is a summand, then also $(1-b, 1-a, T)$ is a summand.

The associativity of $p(x, 1-y)$, however, can only be guaranteed in case of an ordinally irreducible $i$, i.e. a Frank t-norm.

**Theorem 6.** *Consider a generator triplet* $(p, i, j)$ *such that* $i$ *is a t-norm, then the following statements are equivalent:*

(i) *the mapping* $j(1-x, 1-y)$ *is a t-norm;*
(ii) *the mapping* $p(x, 1-y)$ *is symmetrical;*
(iii) *$i$ is a symmetrical ordinal sum of Frank t-norms.*

**Theorem 7.** *Consider a generator triplet* $(p, i, j)$ *such that* $i$ *is a t-norm, then the following statements are equivalent:*

(i) *the mapping* $p(x, 1-y)$ *is a t-norm;*
(ii) *$i$ is a Frank t-norm.*

In the latter case, i.e. when $i$ is a Frank t-norm, say $i = T^s$, $s \in [0, \infty]$, it holds that

$$p(x, y) = T^{1/s}(x, 1-y)$$
$$j(x, y) = T^s(1-x, 1-y).$$

This result closes the loop, and brings us back to the conclusions drawn from the axiomatic study of Fodor and Roubens expressed in Theorem 2.

## 7 Conclusion

We have reconsidered the construction of additive fuzzy preference structures. Generator triplets facilitate the construction of an additive fuzzy preference structure from a fuzzy large preference relation. A generator triplet is uniquely determined by its indifference generator. Specific indifference generators allow to characterize particular generator triplets, such as monotone triplets and t-norm triplets. However, not all additive fuzzy preference structures can be obtained by applying some monotone triplet to a large preference relation. This raises a number of questions: (i) As the notion of a large preference relation cannot be stated more generally, should we restrict the class of additive fuzzy preference structures? (ii) Should we devise algorithms to approximate a given additive fuzzy preference structure by a monotonically generated one? (iii) Or more philosophically, is indifference (similarity) really a monotone function of the large preference?

## References


1. B. De Baets, *Local and global characteristic behaviour of additive fuzzy preference structures*, Abstr. Nineteenth Linz Seminar on Fuzzy Set Theory (Linz, Austria), 1998, pp. 22–28.
2. B. De Baets and H. De Meyer, *On the equiponderate equation* $x^a + x^b + x = x^c + x^d + 1$ *and a representation of weight quadruplets*, J. Appl. Math. Decis. Sci. **2** (1998), 147–158.
3. B. De Baets and J. Fodor, *Twenty years of fuzzy preference structures (1978–1997)*, Riv. Mat. Sci. Econom. Social. **20** (1997), 45–66.
4. B. De Baets and B. Van De Walle, *Minimal definitions of classical and fuzzy preference structures*, Proc Annual Meeting of the North American Fuzzy Information Processing Society (Syracuse, New York, USA), 1997, pp. 299–304.
5. J. Fodor and M. Roubens, *Valued preference structures*, European J. Oper. Res. **79** (1994), 277–286.
6. J. Fodor and M. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support*, Kluwer Academic Publishers, Dordrecht, 1994.
7. C. Genest, J. Molina, J. Lallena and C. Sempi, *A characterization of quasi-copulas*, J. Multivariate Analysis **69** (1999), 193–205.
8. E.-P. Klement, R. Mesiar and E. Pap, *Triangular norms*, Kluwer Academic Publishers, 2000.
9. A. Kolesárová and J. Mordelová, *1-Lipschitz and kernel aggregation operators*, Proc. Internat. Summer School on Aggregation Operators and their Applications (Oviedo, Spain), 2001, pp. 71–75.
10. R. Nelsen, *An Introduction to Copulas*, Lecture Notes in Statistics, Vol. **139**, Springer-Verlag, New York, 1998.
11. M. Roubens and Ph. Vincke, *Preference modeling*, Lecture Notes in Economics and Mathematical Systems **250**, Springer-Verlag, Berlin, 1985.
12. B. Van de Walle, B. De Baets and E. Kerre, *Characterizable fuzzy preference structures*, Annals of Operations Research, Special Issue “Preference modelling” (D. Bouyssou and Ph. Vincke, eds.), **80** (1998), 105–136.
13. L. Zadeh, *Fuzzy sets*, Information and Control **8** (1965), 338–353.

# Preferences and Consistency Issues in Group Decision Making

**Francisco Chiclana, Enrique Herrera-Viedma, Sergio Alonso and Ricardo Alberto Marques Pereira**

**Abstract** A group selection of one alternative from a set of feasible ones should be based on the preferences of individuals in the group. Decision making procedures are usually based on pair comparisons, in the sense that processes are linked to some degree of credibility of preference. The main advantage of pairwise comparison is that of focusing exclusively on two alternatives at a time and on how they are related. However, it generates more information that needed and therefore inconsistent information may be generated. This paper addresses both preference representation and consistency of preferences issues in group decision making.

Different preference representation formats individuals can use to model or present their preferences on a set of alternatives in a group decision making situation are reviewed. The results regarding the relationships between these preference representation formats mean that the fuzzy preference relation "is preferred to" representing the strength of preference of one alternative over another in the scale [0, 1] can be used as the base element to integrate these different preference representation formats in group decision making situations.

Due to the complexity of most decision making problems, individuals' preferences may not satisfy formal properties that fuzzy preference relations are required to verify. Consistency is one of them, and it is associated with the **transitivity property.** Many properties have been suggested to model transitivity of fuzzy preference relations. As aforementioned, this paper provides an overview of the main results published in this area.

## 1 Introduction

Group Decision-Making (GDM) consists of multiple individuals interacting to reach a decision. Each decision maker (expert) may have unique motivations or goals and may approach the decision process from a different angle, but have a common interest in reaching eventual agreement on selecting the 'best' option(s) [11, 32].

Decisions depend, at least in part, on preferences [14]. Indeed, the following quotation from Fishburn [13] fully justifies the above in the context of GDM:

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

> DEMOCRATIC THEORY is based on the premise that the resolution of a matter of social policy, group choice or collective action should be based on the desires or preferences of the individuals in the society, group or collective

In order to reach a decision, experts have to express their preferences by means of a set of evaluations over a set of alternatives. It has been common practice in research to model GDM problems in which all the experts express their preferences using the same preference representation format. However, in real practice this is not always possible because each expert has his/her unique characteristics with regard to knowledge, skills, experience and personality, which implies that different experts may express their evaluations by means of different preference representation formats. In fact, this is an issue that recently has attracted the attention of many researchers in the area of GDM, and as a result different approaches to integrating different preference representation formats have been proposed [5, 6, 7, 12, 15, 49, 50].

In many situations decision processes are based on preference relations, in the sense that processes are linked to some degree of preference of any alternative over another. The main advantage of pairwise comparison is that of focusing exclusively on two alternatives at a time and on how they are related. However, it generates more information that needed and therefore inconsistent information may be generated.

This paper addresses both preference representation formats and the consistency of preferences issues in group decision making. A review of the main results on integration of preference representation structures will be given in Sect. 2, while the problem of consistency when working with fuzzy preference relations will be reviewed in Sect. 3. In Sect. 4 we make note the existence of a conflict between the definition of a consistent multiplicative preference relation and the scale proposed to provide a such preference relation. Obviously, the same problem exists when dealing with fuzzy preference relations. In order to overcome this problem, In Sect. 5 we propose a set of conditions to be verified by a function in order to model consistency of fuzzy preferences. Finally, in Sect. 6 we draw our conclusions.

## 2 Preference Representation Formats: Integration

Experts may provide preferences on a set of alternatives, $X = \{x_1, x_2, \ldots, x_n\}$ $(n \geq 2)$, by using many different representation formats. Among these formats we have:

A preference ordering of the alternatives

In this case, an expert, $e_k$, gives his preferences on $X$ as an individual preference ordering, $O^k = \{o^k(1), ..., o^k(n)\}$, where $o^k(\cdot)$ is a permutation function over the index set, $\{1, ..., n\}$, for the expert, $e_k$, [5, 42]. Therefore, according to this point of view , an ordered vector of alternatives, from best to worst, is given.

A utility function

In this case, an expert, $e_k$, gives his preferences on $X$ as a set of $n$ utility values, $U^k = \{u_i^k,\ i = 1, ..., n\}$, $u_i^k \in [0, 1]$, where $u_i^k$ represents the utility evaluation given by the expert $e_k$ to the alternative $x_i$ [30, 47].

A preference relation

In the classical preference modelling, given two alternatives, an expert judges them in one of the following ways: (i) one alternative is preferred to another; (ii) the two alternatives are indifferent to him/her; (iii) he/she is unable to compare them.

According to these cases, three binary relations can be defined: (i) the strict preference relation $P$: $(x, y) \in P$ if and only if the expert prefers $x$ to $y$; (ii) the indifference relation $I$: $(x, y) \in I$ if and only if the expert is indifferent between $x$ and $y$; (iii) the incomparability relation $J$: $(x, y) \in J$ if and only if the expert unable to compare $x$ and $y$.

Fishburn in [14] defines indifference as the absence of strict preference. He also points out that indifference might arise in three different ways: (a) when an expert truly feels that there is no real difference, in a preference sense, between the alternatives; (b) when the expert is uncertain as to his/her preference between the alternatives because 'he might find their comparison difficult and may decline to commit himself[/herself] to a strict preference judgement while not being sure that he[/she] regards [them] equally desirable (or undesirable)'; (c) or when both alternative are considered incomparable on a preference basis by the expert. It is obvious from the third case that Fishburn treats the incomparability relation as an indifference relation, i.e., $J$ is empty (there is no incomparability).

A preference structure on a set of alternatives $X$ is defined as a triplet $(P, I, J)$ of binary relation in $X$ that satisfy [38, 39]:

1. $P$ is irreflexive and asymmetrical
2. $I$ is reflexive and symmetrical
3. $J$ is irreflexive and symmetrical
4. $P \cap I = P \cap J = I \cap J = \emptyset$
5. $P \cup P^t \cup I \cup J = A^2$

where $P^t$ is the transpose (or inverse) of $P$: $(x, y) \in P \Leftrightarrow (y, x) \in P^t$. Condition 5 is called the completeness condition.

In [38] it is proved that a preference structure $(P, I, J)$ on a se of alternatives $X$ can be characterised by the single reflexive relation $R = P \cup I$: $(x, y) \in R$ if and only if "$x$ is as good as $y$". $R$ is called the large preference relation of $(P, I, J)$. Conversely, given any reflexive binary relation $R$ in $X$, a preference structure $(P, I, J)$ can be constructed on it as follows: $P = R \cap (R^t)^c$, $I = R \cap R^t$, $J = R^c \cap (R^t)^c$, where $R^c$ is the complement of $R$: $(x, y) \in R \Leftrightarrow (y, x) \notin R^c$.

When using numerical representations of preferences on a set of alternatives $X$, we have [13]:

$$r_{ij} = 1 \Leftrightarrow \text{the expert prefers } x_i \text{ to } x_j \Leftrightarrow x_i \succ x_j$$
$$r_{ij} = 0 \Leftrightarrow \text{the expert prefers } x_j \text{ to } x_i \Leftrightarrow x_j \succ x_i$$

Clearly, this can be extended by adding the indifference case:

$$r_{ij} = 0.5 \Leftrightarrow \text{the expert is indifferent between } x_i \text{ and } x_j \Leftrightarrow x_i \sim x_j$$

However, if $x_i$ is preferred to $x_j$ and $x_j$ to $x_k$, the question whether the "degree or strength of preference" of $x_i$ over $x_j$ exceeds, equals, or is less than the "degree or strength of preference" of $x_j$ over $x_k$ cannot be answered by the classical preference modelling. The implementation of the degree of preference between alternatives may be essential in many situations. Take for example the case of 3 alternatives $\{x, y, z\}$ and 2 experts. If one of the experts prefers $x$ to $y$ to $z$, and the other prefers $z$ to $y$ to $x$ then using the above values it may be difficult or impossible to decide which alternative is the best. This may be not the case if intensities of preferences are allowed in the above model. As Fishburn points out in [13], if alternative $y$ is closer to the best alternative than to the worst one for both experts then it might seem appropriate to "elect" it as the social choice, while if it is closer to the worst than to the best, then it might be excluded from the choice set. Intensity of preferences can be implemented when modelling preferences by using fuzzy preference relations [51] or multiplicative preference relations [40].

*A fuzzy preference relation* $R$ on a set of alternatives $X$ is a fuzzy set on the product set $X \times X$, that is characterized by a membership function

$$\mu_R : X \times X \longrightarrow [0, 1].$$

When cardinality of $X$ is small, the preference relation may be conveniently represented by the $n \times n$ matrix $R = (r_{ij})$ being $r_{ij} = \mu_R(x_i, x_j)\ \ \forall i, j \in \{1, \ldots, n\}$. The element $r_{ij} \in R$ is usually interpreted as the preference degree of the alternative $x_i$ over $x_j$, as follows [46]: $r_{ij} = 1/2$ indicates indifference between $x_i$ and $x_j$ $(x_i \sim x_j)$, $r_{ij} > 1/2$ represents an uncertain preference of $x_i$ over $x_j$ $(x_i \succ x_j)$ with $r_{ij} = 1$ when $x_i$ is definitely (certainly) preferred to $x_j$. In this case, the preference matrix, $R$, is usually assumed additive reciprocal, i.e.,

$$r_{ij} + r_{ji} = 1\ \ \forall i, j \in \{1, \ldots, n\}.$$

*A multiplicative preference relation* $A$ on a set of alternatives $X$ is represented by a matrix $A \subset X \text{ x } X$, $A = (a_{ij})$, being $a_{ij}$ interpreted as the ratio of the preference intensity of alternative $x_i$ to that of $x_j$, i.e., it is interpreted as $x_i$ is $a_{ij}$ times as good as $x_j$. Saaty suggests measuring $a_{ij}$ using a ratio-scale, and precisely the 1 to 9 scale: $a_{ij} = 1$ indicates indifference between $x_i$ and $x_j$, $a_{ij} = 9$ indicates that $x_i$ is absolutely preferred to $x_j$, and $a_{ij} \in \{1, \ldots, 9\}$ indicates intermediate preference evaluations. In this case, the preference relation, $A$, is usually assumed multiplicative reciprocal, i.e.,

$$a_{ij} \cdot a_{ji} = 1 \ \ \forall i, j \in \{1, \ldots, n\}.$$

In the context of GDM with heterogeneous information, an interesting question to answer is that of the relationship between the different preference representation formats.

*Preference Orderings and Utility Functions → Binary Preference Relations*

Clearly, from a preference ordering on $X$ we can derive a binary preference relation

$$x_i \succeq x_j \Leftrightarrow o(i) \leq o(j) \ \forall i, j = 1, \ldots, n,$$

Also, given an utility function on $X$, a preference ordering, and consequently a classical preference relation, can easily be derived as follows

$$o(i) \leq o(j) \Leftrightarrow u(x_i) \geq u(x_j) \ \forall i, j = 1, \ldots, n,$$

*Binary Preference Relations → Preference Orderings and Utility Functions*

Given a binary preference relation, it is not always possible to assure the existence of a unique preference ordering or an utility function verifying the above equivalence. In order to get a positive answer additional conditions have to be imposed to the preference relation.

Given the binary preference relation *is preferred to* ($\succ$) on a countable set $X$, with *is indifferent to* ($\sim$) defined as $x \sim y$ if neither $x \succ y$ nor $y \succ x$, a fundamental result is that there exists an utility function $u\colon X \to \mathbb{R}$ such that

$$x \succ y \Leftrightarrow u(x) > u(y)$$

if and only if $\succ$ on $X$ is a weak order, i.e., it is transitive ($x \succ y \wedge y \succ z \Rightarrow x \succ z$), irreflexive (we never have $x \succ x$) and $\sim$ is transitive ($x \sim y \wedge y \sim z \Rightarrow x \sim z$) [15]. The utility function $u$ is said to represent the preference relation $\succ$ . Obviously, in this case, any (positive) monotonic transformation of the utility function $u$ is also a utility function representing the same preference relation because such a transformation preserves the ranking order of the original utility numbers. So, if we begin with the utility function $u$ and then use the (positive) monotonic transformation $f$ to get a new function $v$ defined as $v(x) = f(u(x))$, then $v$ is also a utility function representing the same preference relation as $u$.

*Fuzzy Preference Relations → Preference Orderings and Utility Functions*

Given a fuzzy preference relation on a finite set of alternatives $X$, not necessarily reciprocal, Wang proved in [48] that if the following acyclic property was verified

$$\begin{aligned}&\forall i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, n\} :\\ &\qquad r_{i_1 i_2} > r_{i_2 i_1}, r_{i_2 i_3} > r_{i_3 i_2}, \ldots, r_{i_m i_{m-1}} > r_{i_{m-1} i_m} \Rightarrow r_{i_1 i_m} > r_{i_m i_1}\end{aligned}$$

then a total order can be produced in $X$, i.e, given any two arbitrary alternatives $x_i$ and $x_j$ in $X$, one of the following relations holds: $x_i \succ x_j$, $x_j \succ x_i$, $x_i \sim x_j$. A similar result was obtained in [2] when the fuzzy preference relation is reciprocal and is weakly transitive ($r_{ij} > 0.5 \wedge r_{jk} > 0.5 \Rightarrow r_{ik} > 0.5$).

With fuzzy preference relations, Orlovsky [26] proposed a rational criterion to produce a total order on $X$ based on the strict preference relation $R^s = (r^s_{ij})$ with $r^s_{ij} = \max\{r_{ij} - r_{ji}, 0\}$ and the concept of non-dominance. Conditions that guarantee the existence of un-fuzzy non-dominated alternatives were obtained by Montero and Tejada (see [35, 36]) and by Kołodziejczyk [29]. A quantifier non-dominance degree that extended Orlovsky's non-dominance degree was proposed by Chiclana et al. in [4].

*Preference Orderings and Utility Functions → Fuzzy Preference Relations*

In [5, 6, 22] the following results were obtained:

**Proposition 1.** *Let $X$ be a set of alternatives and $\lambda^k_i$ represents an evaluation associated to alternative $x_i$, indicating the performance of that alternative according to a point of view (expert or criteria) $e_k$. Then, the intensity of preference of alternative $x_i$ over alternative $x_j$, $r^k_{ij}$, for $e_k$ is given by the following transformation function*

$$r^k_{ij} = \varphi(\lambda^k_i, \lambda^k_j) = \frac{1}{2} \cdot [1 + \psi(\lambda^k_i, \lambda^k_j) - \psi(\lambda^k_j, \lambda^k_i)],$$

*where $\psi$ is a function verifying*

1. $\psi(z, z) = \frac{1}{2}$, $\forall z \in \mathbb{R}$.
2. *$\psi$ is non decreasing in the first argument and non increasing in the second argument.*

*Utility Values and Fuzzy Preference Relations*

**Corollary 1.** *If $\lambda^k_i = u^k_i$ and*

$$\psi(z, y) = \begin{cases} \dfrac{s(z)}{s(z) + s(y)} & \text{if } (z, y) \neq (0, 0) \\ \dfrac{1}{2} & \text{if } (z, y) = (0, 0) \end{cases}$$

*where $s : [0, 1] :\to \mathbb{R}^+$ is a non decreasing and continuous function, verifying $s(0) = 0$, then $\varphi$ transforms utility values given on the basis of a ratio scale into fuzzy preference relations. In particular, if $\psi(x, y) = \frac{x^2}{x^2 + y^2}$, then*

$$r^k_{ij} = f^1\left(u^k_i, u^k_j\right) = \frac{\left(u^k_i\right)^2}{\left(u^k_i\right)^2 + \left(u^k_j\right)^2}$$

*Preference Orderings and Fuzzy Preference Relations*

**Corollary 2.** *If* $\lambda_i^k = o^k(i)$*, and* $\psi(\lambda_i^k, \lambda_j^k) = F(\lambda_j^k - \lambda_i^k)$*, where F is any non decreasing function, then* $\varphi$ *transforms preference orderings into fuzzy preference relations.*

*In particular, if* $\psi(x, y) = \frac{y-x}{2(n-1)}$*, then*

$$r_{ij}^k = f^2\left(o_i^k, o_j^k\right) = \frac{1}{2}\left(1 + \frac{o_j^k - o_i^k}{n-1}\right)$$

*Multiplicative Preference Relations and Fuzzy Preference Relations*

**Proposition 2.** *Let X be a set of alternatives, and associated with it a multiplicative preference relation* $A^k = \left(a_{ij}^k\right)$*. Then, the corresponding additive fuzzy preference relation,* $R^k = \left(r_{ij}^k\right)$*, associated with* $A^k$ *is given as follows:*

$$r_{ij}^k = g\left(a_{ij}^k\right) = \frac{1}{2}\left(1 + \log_9 a_{ij}^k\right)$$

These results may justify the choice of fuzzy preference relations as the base element to integrate these different preference representation formats in GDM context. In the following section we deal with issue of consistency of preferences.

## 3 Consistency of Preferences

There are three fundamental and hierarchical levels of rationality assumptions when dealing with preference relations [18]:

- The first level of rationality requires indifference between any alternative and itself.
- The second one assumes the property of reciprocity in the pairwise comparison between any two alternatives.
- Finally, the third one is associated with the transitivity in the pairwise comparison among any three alternatives.

The mathematical modelling of all these rationality assumptions obviously depends on the scales used for providing the preference values [9, 16, 26, 40, 46].

A preference relation verifying the third level of rationality is usually called a *consistent preference relation* and any property that guarantees the transitivity of the preferences is called a consistency property. The lack of consistency in decision making can lead to inconsistent conclusions; that is why it is important, in fact crucial, to study conditions under which consistency is satisfied [16, 26, 40].

In a crisp context, where an expert provides his/her opinion on the set of alternatives $X$ by means of a binary preference relation, $R$, the concept of consistency it has traditionally been defined in terms of acyclicity [43], that is the absence of sequences such as $x_1, x_2, \ldots, x_k (x_{k+1} = x_1)$ with $x_j R x_{j+1} \forall j = 1, \ldots, k$. Clearly, this condition as said before is closely related to the transitivity of the binary relation and its corresponding binary indifference relation.

In a fuzzy context, where an expert expresses his/her opinions using fuzzy preference relations, $R$, or multiplicative preference relations, in the case of Saaty's method, $A$, the traditional requirement to characterise consistency has followed the way of extending the classical requirements of binary preference relations. Thus, in these cases consistency is also based on the notion of transitivity, in the sense that if alternative $x_i$ is preferred to alternative $x_j$ and this one to $x_k$ then alternative $x_i$ should be preferred to $x_k$. The main difference in these cases with respect to the classical one is that transitivity has been modelled in many different ways due to the role the intensities of preference have [16, 19, 20, 21, 26, 40, 46, 52].

Due to the hierarchical structure of the three rationality assumptions for a preference relation, the verification of a particular level of rationality should be a necessary condition in order to verify the next level of rationality. This means that the third level of rationality, transitivity of preferences, should imply or be compatible with the second level of rationality, reciprocity of preferences, and the second level with the first one, indifference of any alternative with itself.

This necessary compatibility between the rationality assumptions can be used as a criterion for considering a particular condition modelling any one of the rationality levels as adequate or inadequate. In the case of fuzzy (multiplicative) preference relations, the indifference between any alternative, $x_i$, and itself is modelled by associating the preference value $r_{ii} = 0.5$ ($a_{ii} = 1$). The reciprocity of fuzzy (multiplicative) preferences is modelled using the property $r_{ij} + r_{ji} = 1, \ \forall i, j$ ($a_{ij} \cdot a_{ji} = 1, \ \forall i, j$). A necessary condition for a preference relation to verify reciprocity should be that indifference between any alternative and itself holds. Because reciprocity property implies the indifference of preferences, we conclude that both properties are compatible.

In the case of multiplicative preference relations Saaty means by *consistency* what he calls *cardinal transitivity* in the strength of preferences, which is a stronger condition than the traditional requirement of the transitivity of preferences [40]:

**Definition 1.** *A reciprocal multiplicative preference relation $A = (a_{ij})$ is consistent if $a_{ij} \cdot a_{jk} = a_{ik} \ \ \forall i, j, k = 1, \ldots, n$.*

Inconsistency for Saaty is a violation of proportionality which may not entail violation of transitivity [40]. Furthermore, consistency implies reciprocity, and therefore, they are both compatible.

In [40] Saaty shows that a reciprocal multiplicative preference relation is consistent if and only if its maximum or principal eigenvalue $\lambda_{max}$ is equal to the number of alternatives $n$. Under this consistency property, Saaty proves that there exists a set of priorities (utilities) $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ such that $a_{ij} = \frac{\lambda_i}{\lambda_j}$. Moreover, this set of values is unique up to positive linear transformation $f(\lambda_i) = \beta \cdot \lambda_i$ with $\beta > 0$.

Thus, if a multiplicative preference relation is consistent then it can be represented by a unique (up to positive linear transformations) utility function.

For fuzzy preference relations, there exist many properties or conditions that have been suggested as rational conditions to be verified by a consistent relation. Among these, we can citethe following:

1. *Triangle condition [30]:* $r_{ij} + r_{jk} \geq r_{ik} \ \forall i, j, k$.
   This condition can be geometrically interpreted considering alternatives $x_i,\ x_j,\ x_k$ as the vertices of a triangle with length sides $r_{ij},\ r_{jk}$ and $r_{ik}$ [30], and therefore the length corresponding to the vertices $x_i,\ x_k$ should not exceed the sum of the lengths corresponding to the vertices $x_i,\ x_j$ and $x_j,\ x_k$.
2. *Weak transitivity [46]:* $r_{ij} \geq 0.5,\ r_{jk} \geq 0.5 \Rightarrow r_{ik} \geq 0.5 \ \forall i, j, k$.
   The interpretation of this condition is the following: If $x_i$ is preferred to $x_j$ and $x_j$ is preferred to $x_k$, then $x_i$ should be preferred to $x_k$. This kind of transitivity is the usual transitivity condition ($x_i$ is preferred to alternative $x_j$ and this one to $x_k$ then alternative $x_i$ should be preferred to $x_k$) a logical and consistent person should use if he/she does not want to express inconsistent opinions, and therefore it is the minimum requirement condition that a consistent fuzzy preference relation should verify.
3. *Max-min transitivity [9, 52]:* $r_{ik} \geq \min(r_{ij}, r_{jk}) \ \forall i, j, k$.
   The idea represented here is that the preference value obtained by a direct comparison between two alternatives should be equal to or greater than the minimum partial values obtained when comparing both alternatives with an intermediate one. This kind of transitivity has been the traditional requirement to characterise consistency in the case of fuzzy preference relations [52], although it is a very strong concept that it could not be verified even when a fuzzy preference relation is considered perfectly consistent from a practical point of view. For example, let us consider a set of three alternatives $X = \{x_1, x_2, x_3\}$, such that $x_1 \prec x_2 \prec x_3$. Suppose that the opinions about these alternatives are given by the following fuzzy preference relation

$$R = \begin{pmatrix} 0.5 & 0.1 & 0 \\ 0.9 & 0.5 & 0.4 \\ 1 & 0.6 & 0.5 \end{pmatrix}.$$

   On the one hand, this matrix reflects the fact that $x_1 \prec x_2 \prec x_3$; it verifies weak transitivity and the triangle condition. On the other hand, it does not verifies max-min transitivity because $r_{13} < \min\{r_{12}, r_{23}\}$.
4. *Max-max transitivity [9, 52]:* $r_{ik} \geq \max(r_{ij}, r_{jk}) \ \forall i, j, k$.
   This concept represents the idea that the preference value obtained by a direct comparison between two alternatives should be equal to or greater than the maximum partial values obtained when comparing both alternatives using an intermediate one. This is a stronger concept than max-min transitivity and therefore if a fuzzy preference relation does not verify the latter neither verifies the former.
5. *Restricted max-min transitivity [46]:* $r_{ij} \geq 0.5,\ r_{jk} \geq 0.5 \Rightarrow r_{ik} \geq \min(r_{ij}, r_{jk}) \ \forall i, j, k$.

When a fuzzy preference relation verifies this condition it is modelled the concept that when an alternative $x_i$ is preferred to $x_j$ with a value $p_{ij}$ and $x_j$ is preferred to $x_k$ with a value $r_{ij}$, then $x_i$ should be preferred to $x_k$ with at least an intensity of preference $r_{ik}$ equal to the minimum of the above values. The inequality should becomes equality only when there exist indifference between at least two of the three alternatives. A consistent fuzzy preference relation has to verify this condition, which goes a step further than weak transitivity because add an extra requirement about the degrees of preferences involved. This transitivity condition is therefore stronger than weak transitivity but it is milder than max-min transitivity. It is easy to prove that the above fuzzy preference relation $R$ verifies restricted max-min transitivity.

6. *Restricted max-max transitivity [46]:* $r_{ij} \geq 0.5,\ r_{jk} \geq 0.5 \Rightarrow r_{ik} \geq \max(r_{ij}, r_{jk})$ $\forall i, j, k$.
   In this case it is modelled the concept that when an alternative $x_i$ is preferred to $x_j$ with a value $r_{ij}$ and $x_j$ is preferred to $x_k$ with a value $r_{ij}$, then $x_i$ should be preferred to $x_k$ with at least an intensity of preference $r_{ik}$ equal to the maximum of the above values. As in the previous case, the the equality should hold only when there exist indifference between at least two of the three alternatives, in which case, restricted max-max transitivity and restricted max-min transitivity coincide. It is clear that this concept is, on the one hand, stronger than restricted max-min transitivity and, on the other hand, milder than max-max transitivity. This concept has been considered by Tanino [46] as a compulsory condition to be verified by a consistent fuzzy preference relation. It is easy to prove that the fuzzy reciprocal preference relation $R$, given above, verifies restricted max-max transitivity.
7. *Multiplicative transitivity [46]*: $\frac{r_{ji}}{r_{ij}} \cdot \frac{r_{kj}}{r_{jk}} = \frac{r_{ki}}{r_{ik}}$ $\forall i, j, k$.
   Tanino in [46] introduced this concept of transitivity only in the case of being $r_{ij} > 0$ $\forall i, j$, and interpreting $r_{ij}/r_{ji}$ as a ratio of the preference intensity for $x_i$ to that of $x_j$, i.e., $x_i$ is $r_{ij}/r_{ji}$ times as good as $x_j$. Multiplicative transitivity includes restricted max-max transitivity [45, 46], and rewritten as $r_{ij} \cdot r_{jk} \cdot r_{ki} = r_{ik} \cdot r_{kj} \cdot r_{ji}$ $\forall i, j, k$. In the case of a reciprocal fuzzy preference relation this expression can be expressed in the following form:

$$
\begin{aligned}
&\forall i, j, k:\\
&\qquad r_{ij} \cdot r_{jk} \cdot (1 - r_{ik}) = r_{ik} \cdot r_{kj} \cdot r_{ji} \quad \Leftrightarrow\\
&\qquad r_{ij} \cdot r_{jk} - r_{ij} \cdot r_{jk} \cdot r_{ik} = r_{ik} \cdot r_{kj} \cdot r_{ji} \quad \Leftrightarrow\\
&\qquad r_{ik} \cdot r_{kj} \cdot r_{ji} + r_{ij} \cdot r_{jk} \cdot r_{ik} = r_{ij} \cdot r_{jk} \quad \Leftrightarrow\\
&\qquad r_{ik} \cdot (r_{kj} \cdot r_{ji} + r_{ij} \cdot r_{jk}) = r_{ij} \cdot r_{jk} \quad \Leftrightarrow\\
&\qquad r_{ik} = \frac{r_{ij} \cdot r_{jk}}{r_{ij} \cdot r_{jk} + r_{ji} \cdot r_{jk}} \quad \Leftrightarrow\\
&\qquad r_{ik} = \frac{r_{ij} \cdot r_{jk}}{r_{ij} \cdot r_{jk} + (1 - r_{ij}) \cdot (1 - r_{jk})}
\end{aligned}
$$

   This expression is a well known andlike uninorm which is self-dual with respect to the negator operator $N(x) = 1 - x$ (for more details see [17, 28]).

This type of transitivity has also been studied by De Baets et al. in [8] within a general framework of transitivity of reciprocal fuzzy preference relations, the cycle-transitivity, under the name of 'isostochastic transitivity'. This is also a symmetric sum in the sense of Silvert [44] that has been applied for information combination in approximate reasoning (see [10] and cites within it).

8. *Additive transitivity [45, 46]:* $(r_{ij} - 0.5) + (r_{jk} - 0.5) = (r_{ik} - 0.5)\ \forall i, j, k$, or equivalently $r_{ij} + r_{jk} + r_{ki} = \frac{3}{2}\ \forall i, j, k$.
This kind of transitivity has the following interpretation: suppose we want to establish a ranking between three alternatives $x_i$, $x_j$ and $x_k$, and that the information available about these alternatives suggests that we are in an indifference situation, i.e. $x_i \sim x_j \sim x_k$. When giving preferences this situation would be represented by $r_{ij} = r_{jk} = r_{ik} = 0.5$. Suppose now that we have a piece of information that says $x_i \prec x_j$, i.e. $r_{ij} < 0.5$. This means that $r_{jk}$ or $r_{ik}$ have to change, otherwise there would be a contradiction, because we would have $x_i \prec x_j \sim x_k \sim x_i$. If we suppose that $r_{jk} = 0.5$ then we have the situation: $x_j$ is preferred to $x_i$ and there is no difference in preferring $x_j$ to $x_k$. We must then conclude that $x_k$ has to be preferred to $x_i$. Furthermore, as $x_j \sim x_k$ then $r_{ij} = r_{ik}$, and so $(r_{ij} - 0.5) + (r_{jk} - 0.5) = (r_{ij} - 0.5) = (r_{ik} - 0.5)$. We have the same conclusion if $r_{ik} = 0.5$. In the case of $r_{jk} < 0.5$, then we have that $x_k$ is preferred to $x_j$ and this to $x_i$, so $x_k$ should be preferred to $x_i$. On the other hand, the value $r_{ik}$ has to be equal to or lower than $r_{ij}$, being equal only in the case of $r_{jk} = 0.5$ as we have already shown. Interpreting the value $r_{ji} - 0.5$ as the intensity of strict preference of alternative $x_j$ over $x_i$, then it seems reasonable to suppose that the intensity of preference of $x_i$ over $x_k$ should be equal to the sum of the intensities of preferences when using an intermediate alternative $x_j$, that is, $r_{ik} - 0.5 = (r_{ij} - 0.5) + (r_{jk} - 0.5)$. The same reasoning can be applied in the case of $r_{jk} > 0.5$ [26, 45, 46]. The fuzzy preference relation $R$, given above, verifies additive transitivity.

The following diagram shows all logical relationships between the defined transitivity conditions. We note that there is no relationship between weak-transitivity and triangle condition [30],

$$\begin{array}{ccccccccc} & & & & (7) & & & & \\ & & & & \Downarrow & & & & \\ (1) & \Leftarrow & (8) & \Rightarrow & (6) & \Rightarrow & (5) & \Rightarrow & (2) \\ & & & & \Uparrow & & \Uparrow & & \\ & & & & (4) & \Rightarrow & (3) & & \end{array}$$

In the following, we will show that max-max transitivity is not compatible with the reciprocity property. If a fuzzy preference relation verifies max-max transitivity and reciprocity then $r_{ik} \geq \max\{r_{ij}, r_{jk}\}\ \forall i, j, k$ and $r_{ij} = 1 - r_{ji}\}\ \forall i, j$, which implies:

$$1 - r_{ik} \leq 1 - \max\{r_{ij}, r_{jk}\}\ \forall i, j, k \Rightarrow r_{ki} \leq \min\{r_{kj}, r_{ji}\}\ \forall i, j, k$$

which contradicts max-max transitivity. The same conclusion can be obtained regarding max-min transitivity. Therefore both properties are not adequate properties to model the transitivity for fuzzy preference relations.

If we examine the relationship between restricted max-max transitivity and reciprocity, then we conclude that the fuzzy preference relation also has to verify the complementary restricted min-min transitivity, that is,

$$\forall i, j, k : \min\{r_{ij}, r_{jk}\} \leq 0.5 \Rightarrow r_{ik} \geq \min\{r_{ij}, r_{jk}\}.$$

However, nor restricted max-max transitivity nor restricted min-min transitivity imply reciprocity. For example, the following fuzzy preference relation

$$R = \begin{pmatrix} 0.5 & 0.6 & 0.8 \\ 0.4 & 0.5 & 0.7 \\ 0.1 & 0.3 & 0.5 \end{pmatrix}$$

verifies both restricted transitivity properties but it is not reciprocal. This does not imply that they are incompatible with the reciprocity property. In fact, a fuzzy preference relation can be reciprocal and still verify both restricted transitivity properties, as the one we would have obtained by changing the values $r_{13}$ for 0.9 or the value $r_{31}$ for 0.2.

If we examine the compatibility between the additive consistency property and reciprocity then we conclude that the first one implies the second one. Firstly, we show that additive consistency property implies indifference of preferences. Indeed, when $i = j = k$ additive consistency property reduces to $r_{ii} + r_{ii} + r_{ii} = 1.5$ $\forall i$ which implies $r_{ii} = 0.5$ $\forall i$. Secondly, we show that additive consistency property implies reciprocity property. If $k = i$ then additive consistency reduces to $r_{ij} + r_{ji} + r_{ii} = 1.5$ $\forall i, j$ and because we already have that $r_{ii} = 0.5$ $\forall i$ then $r_{ij} + r_{ji} = 1$ $\forall i, j$.

As shown in [26], additive transitivity for fuzzy preference relations can be seen as the parallel concept of Saaty's consistency property for multiplicative preference relations [41]:

**Proposition 3.** *Let $A = (a_{ij})$ be a consistent multiplicative preference relation, then the corresponding reciprocal fuzzy preference relation, $R = g(A)$ verifies additive transitivity property.*

In such a way, the following definition of a consistent fuzzy preference relation may be given:

**Definition 2.** *A reciprocal fuzzy preference relation $R = (r_{ij})$ is additive consistent if $r_{ij} + r_{jk} + r_{ki} = \frac{3}{2}$ $\forall\, i, j, k = 1, \ldots, n$.*

In [26], Herrera-Viedma et. al. gave a characterisation of the consistency property defined by the additive transitivity property of a fuzzy preference relation $R^k = (r^k_{ij})$. Using this characterization method, a procedure was given to construct a consistent fuzzy preference relation $\widetilde{R}^k$ from a non-consistent fuzzy preference relation $R^k$. As in the case of multiplicative preference relations, if a fuzzy preference relation is additive transitivity then it can be represented by a unique (up to positive linear transformations) utility function. Additive transitivity has been used to obtain more consistent fuzzy preference relation from a given one (see [31]) and as shown in [1, 24, 25] it is also a valuable concept for incomplete fuzzy preference relations as it reduces experts' uncertainty when choosing values to estimate their unknown ones, which is not the case if other types of transitivity conditions were to be used.

## 4 Conflict Between Additive and Multiplicative Consistency Properties and Scales

There are many reasons that point in the direction of considering additive consistency as an adequate property to model transitivity of fuzzy preferences. However, a conflict between the additive consistency property and the scale used for providing the preference values, i.e., the closed interval [0, 1], can appear. To show this, we will use a simple example.

Let us suppose a set of three alternatives $\{x_1, x_2, x_3\}$ for which we have the following information: alternative $x_1$ is considerably more important than alternative $x_2$ and this one is demonstrably or very considerably more important than alternative $x_3$. Suppose that these statements are modelled using the following values $r_{12} = 0.75$ and $r_{23} = 9$ respectively. If we want to maintain the additive consistency property then we would obtain a negative value $r_{13} = 1.5 - r_{12} - r_{23} = -0.15$.

This conflict between the additive consistency property and the scale used for providing preference values suggests that a modification of this property where it acts incoherently has to be made. Because restricted max-max transitivity is the minimum condition required for a reciprocal fuzzy preference relation to be considered consistent, then the modification to introduce in the additive consistency property should maintain restricted max-max transitivity and, by reciprocity, the complementary restricted min-min transitivity.

Obviously, the same problem exists when dealing with multiplicative preference relations.The following simple example will show that there exists a conflict between the scales used to associate multiplicative preference values to judgements and the definition of consistency given by Saaty. Let us suppose a set of three alternatives $\{x_1, x_2, x_3\}$ on which an expert provides the following judgements: alternative $x_1$ is considerably more important than alternative $x_2$ and this one demonstrably or overwhelming more important than alternative $x_3$. In such a case, using Saaty's 1–9 scale, we would have the values $a_{12} = 5$ and $a_{23} = 7$.

On the one hand, if we want to maintain the multiplicative consistency property then, according to Saaty [40], we would have to assign the value $a_{13} = a_{12} \cdot a_{23} = 35$, and the only solution would be using the following consistent reciprocal multiplicative preference relation

$$A = \begin{pmatrix} 1 & 5 & 35 \\ 1/3 & 1 & 7 \\ 1/35 & 1/5 & 1 \end{pmatrix}.$$

Therefore, to avoid such a type of conflict we could proceed by choosing a different scale for providing judgements or by modifying the above definition. With respect to the first question, the use of any other scale of the form $[1/a, a],\ a \in \mathbb{R}^+$, would not make this conflict disappear, which means that the the only possible solution to overcome this conflict would consist of using the scale of pairwise comparison from 0 to $+\infty$. However, as Saaty points out in [40], this may not be useful at all because it assumes that the human judgement is capable of comparing the relative dominance of any two objects, which is not the case.

On the other hand, we note that if $a_{13} \in [7, 9]$ transitivity still holds. We analyze this fact by means of the measure of consistency proposed by Saaty. In [40] Saaty shows that a reciprocal multiplicative preference relation is consistent if and only if its maximum or principal eigenvalue $\lambda_{max}$ is equal to the number of alternatives $n$. However, because perfect consistency is difficult to obtain in practice, especially when measuring preferences on a set with a large number of alternatives, Saaty defined a *consistency index (*$CI = \lambda_{max} - n$*)* that reflects the deviation from consistency of all the $a_{ij}$ of a particular reciprocal multiplicative preference relation from the estimated ratio of priorities $w_i/w_j$.

A measure of inconsistency independent of the order of the reciprocal multiplicative preference relation is defined as the *consistency ratio (CR)*. This is obtained by taking the ratio of CI to the *random index (RI)*, which is an average consistency index of a sample set of randomly generated reciprocal matrices from the scale 1 to 9 (size 500 up to 11 by 11 matrices, and size 100 for squares matrices of orders 12, 13, 14 and 15). For this consistency measure, he proposed a threshold of 0.10 to accept the reciprocal multiplicative preference relation as consistent. When the *CR* is greater than 0.10 then, in order to improve consistency, those judgements with a greater difference $a_{ij}$ and $w_i/w_j$, are usually modified and a new priority vector is derived.

In our previous example we observe that the conflict between the multiplicative consistency property and the scale used by Saaty arises because if we impose consistency then we get values outside the range $[1/9, 9]$. If we restrict the possible values of $a_{13}$ to be in $[1/9, 9]$, then it is clear than in this case alternative $x_1$ should be considered as overwhelming more important than alternative $x_3$, and thus the value of $a_{13}$ should be greater or equal to 7. If $a_{13} = 7$ we get a *CR* value of 0.25412, with $a_{13} = 8$ a *CR* value of 0.212892 and with $a_{13} = 9$ a *CR* value of 0.179714, all of them greater than the minimum 0.10 for considering any reciprocal multiplicative preference relation consistent in this situation.

All these considerations mean that if we do not change the scale used to associate preference values to judgement or want to have a homogeneous scale when working in a group decision context, then the above definitions of consistency of preference relations should be modified.

In the next section, we set out the properties to be verified by a $f\colon [0,1] \times [0,1] \to [0,1]$ so that it can be used to obtained $r_{ik}$ from the pair of values $\left(r_{ij}, r_{jk}\right)$, that is, $r_{ik} = f\left(r_{ij}, r_{jk}\right)$.

## 5 Consistency Function of Preferences: Conditions to Verify

The assumption of experts being able to quantify their preferences in the domain [0,1] instead of {0, 1} or a set with finite cardinality, as it may be a set of linguistic labels [4, 27, 33, 34], underlies unlimited computational abilities and resources from the experts. Taking these unlimited computational abilities and resources into account we may formulate that an expert's preferences are consistent when for any three alternatives $x_i, x_j, x_k$ their preference values are related in the exact form

$$r_{ik} = f(r_{ij}, r_{jk})$$

being $f$ a function $f\colon [0,1] \times [0,1] \to [0,1]$. In what follows we will set out a set of conditions or properties to be verified by such a function $f$.

The above equality can be interpreted as the equation to solve in a a situation when we do want to compare the alternatives $x_i$ and $x_k$, but cannot do it directly, but we have an alternative $x_j$ of which we know the exact values of $r_{ij}$ and $r_{jk}$. In this situation, we can establish a broad comparison between alternatives $x_i$ and $x_k$ on the basis of the values $r_{ij}$ and $r_{jk}$. Indeed, we can distinguish the following cases:

*Case 1.* $r_{ij} = 0.5$ ($r_{jk} = 0.5$) which means that $x_i \sim x_j$ ($x_j \sim x_k$) and as a consequence the strength of preference between $x_i$ and $x_k$ should be the same as the one between $x_j$ and $x_k$. We then have: $r_{ik} = r_{jk}$ ($r_{ik} = r_{jk}$).

*Case 2.* $r_{ij} > 0.5$ and $r_{jk} > 0.5$. In this case, alternative $x_i$ is preferred to alternative $x_j$ ($x_i \succ x_j$) and alternative $x_j$ is preferred to alternative $x_k$ ($x_j \succ x_k$). We then have that $x_i \succ x_j \succ x_k$ which implies $x_i \succ x_k$ and therefore $r_{ik} > 0.5$. Furthermore, in these cases restricted max-max transitivity should be imposed, which means that $x_i$ should be preferred to $x_k$ with a degree of intensity at least equal to the maximum of the intensities $r_{ij}$ and $r_{jk}$: $r_{ik} \geq max\{r_{ij}, r_{jk}\}$, where the equality holds only when there exists indifference between at least one of the alternatives and $x_j$, i.e., $r_{ij} = 0.5$ or $r_{jk} = 0.5$, as we have said in case 1. As a result, in this case $r_{ik} > \max\{r_{ij}, r_{jk}\}$ should be verified.

*Case 3.* When $r_{ij} < 0.5$ and $r_{jk} < 0.5$, a similar argument to the one of case 2 leads to $r_{ik} < \min\{r_{ij}, r_{jk}\}$.

*Case 4.* One reference value is greater than 0.5 and the other is lower than 0.5. Suppose that $r_{ij} > 0.5$ and $r_{jk} < 0.5$. This is equivalent to $r_{ij} > 0.5$ and $r_{kj} = 1 - r_{jk} > 0.5$, that is: $x_i \succ x_j$ and $x_k \succ x_j$. The comparison of alternatives $x_i$ and $x_j$ is done by comparing the intensities of preferences of them over the alternative $x_j$. An indifference situation between $x_i$ and $x_k$ would exist only when both alternatives are preferred over $x_j$ with the same intensity, while the alternative with greater intensity of preference over $x_j$ should be preferred to the other one. This is summarized in the following way:

$$\left\{\begin{array}{l} x_i \sim x_k \text{ if } r_{ij} = r_{kj} \Leftrightarrow r_{ij} + r_{jk} = 1 \\ x_i \succ x_k \text{ if } r_{ij} > r_{kj} \Leftrightarrow r_{ij} + r_{jk} > 1 \\ x_i \prec x_k \text{ if } r_{ij} < r_{kj} \Leftrightarrow r_{ij} + r_{jk} < 1 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} r_{ik} = 0.5 \text{ if } r_{ij} + r_{jk} = 1 \\ r_{ik} > 0.5 \text{ if } r_{ij} + r_{jk} > 1 \\ r_{ik} < 0.5 \text{ if } r_{ij} + r_{jk} < 1 \end{array}\right\}$$

It is obvious that the greater the value $|r_{ij} + r_{jk} - 1|$ the greater $|r_{ik} - 0.5|$.

The following modification of the additive consistency property where it acts incoherently meet the above conditions:

$$f(x, y) = \begin{cases} \min\{x, y\} & x, y \in [0, 0.5] \\ \max\{x, y\} & x, y \in [0.5, 1] \\ x + y - 0.5 & \text{otherwise} \end{cases}$$

However, this function is not associative which is a necessary requirement for a function to be considered consistent in this context. Indeed, we have hat:

$$r_{ik} = f(r_{ij}, r_{jk}) = f\left(r_{ij}, f(r_{il}, r_{lk})\right); \quad r_{ik} = f(r_{il}, r_{lk}) = f\left(f(r_{ij}, r_{jl}), r_{lk}\right)$$

and therefore it is true that:

$$f\left(f(r_{ij}, r_{jl}), r_{lk}\right) = f\left(r_{ij}, f(r_{il}, r_{lk})\right)$$

In terms of function $f$, case 1 implies $f(0.5, x) = f(x, 0.5) = x \ \forall x \in [0, 1]$, which means that function $f$ has neutral element 0.5. In particular, $f(0.5, 0.5) = 0.5$ which means that the neutral element of $f$ is idempotent. This property in conjunction with case 2 mean that function $f$ behaves in $[0.5, 1] \times [0.5, 1]$ as a t-conorm, while in conjunction with case 3 mean that function $f$ behaves in $[0, 0.5] \times [0, 0.5]$ as a t-norm. Clearly, other properties desirable to be verified by such a function $f$ include that of being continuous except maybe in the points (0, 1) and (1,0); increasing with respect to both arguments $x$ and $y$; and commutative.

We make note that uninorm operators present all the above properties, which may suggest that function $f$ may belong to the class of uninorms operators. As said in Sect. 3, multiplicative transitivity is a uninorm, and therefore it may be taken as the condition to be verified for a fuzzy preference relation to be considered as consistent.

## 6 Conclusions

In a GDM problem experts may provide their preferences by means of different preference representation formats. The integration of heterogeneous information is therefore an important issue to be addresses in these situations. A review of the main results regarding the relationships between the numerical representation formats was provided, and it was suggested that the fuzzy preference relation "is preferred to" representing the strength of preference of one alternative over another in the scale [0, 1] could be used as the base element to integrate these different preference representation formats in group decision making situations.

Once preferences are provided by the expert, the problem of measuring the consistency of these preferences becomes crucial to get 'good' solutions. While for crisp and multiplicative preference relations there might exist an agreement on the properties to be satisfied in order to be considered consistent, this is not the case for fuzzy preference relations. Indeed, for a fuzzy preference relation to be considered consistent, many different properties have been suggested. Most of these properties are related in some way to the traditional concept of transitivity. One of this properties is the additive transitivity property, which is equivalent to Saaty's consistency property for multiplicative preference relations. However, both consistency properties are in conflict with the corresponding scale used for providing the preferences. I order to overcome this conflict, a set of conditions have been set for reciprocal fuzzy preference relations to be considered 'fully consistent.' These set of conditions suggest that consistency might be represented by a uninorm operator. One of the suggested properties to model consistency for fuzzy preference relations, the multiplicative consistency, introduced by Tanino in 1988, is an example of a such operator.

## References

1. S. Alonso, F. Chiclana, F. Herrera, E. Herrera-Viedma, J. Alcalá-Fernández and C. Porcel, "A consistency based procedure to estimate missing pairwise preference values," **Int. J. Intell. Syst.,** accepted for publication, 2005.
2. L. Basile, "Ranking alternatives by weak transitivity relations," in J. Kacprzyk, M. Fedrizzi (Eds.) **Multiperson Decision Making using Fuzzy Sets and Possibility Theory,** Kluwer Academic Publishers, Dordrecht, pp. 105 – 112, 1990.
3. P.P. Bonissone and K.S. Decker, "Selecting uncertainty calculi and granularity: an experiment in trading-off and complexity". In Kanal, Lemmer (Eds.) *Uncertainty in Artificial Intelligence* (North-Holland, 1986) 217–247.
4. F. Chiclana, F. Herrera, E. Herrera-Viedma and M.C. Poyatos, "A classification method of alternatives for preference ordering criteria based on fuzzy majority," **The Journal of Fuzzy Mathematics** Vol. 4, No. 4, pp. 801–813, 1996.
5. F. Chiclana, F. Herrera and E. Herrera-Viedma, "Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations," **Fuzzy Sets Syst.,** Vol. 97, No. 1, pp. 33–48, July 1998.
6. F. Chiclana, F. Herera and E. Herera-Viedma, "Integrating multiplicative preference relations in a multipurpose decision-making model based on fuzzy preference relations," **Fuzzy Sets Syst.,** Vol. 122, No. 2, pp. 277–291, Sept. 2001.
7. F. Chiclana, F. Herrera and E. Herrera-Viedma, "A note on the internal consistency of various preference representations," **Fuzzy Sets Syst.,** Vol. 131, No. 1, pp. 75–78, Oct. 2002.
8. B. De Baets, H. De Meyer and D. De Schuymer, "Cyclic evaluation of transitivity of reciprocal relations," **Soc. Choice Welfare,** Vol.26, pp. 217–238, 2006.
9. D. Dubois and H. Prade, **Fuzzy sets and systems: Theory and application.** Academic Press, New York, 1980.

10. D. Dubois and H. Prade, “A review of fuzzy sets aggregation connectives,” **Information Sciences,** Vol. 36,pp. 85–121, 1985.
11. T. Evangelos, **Multi-Criteria Decision Making Methods: A Comparative Study**, Dordrecht: Kluwer Academic Publishers, 2000.
12. Z.-P. Fan, S.-H. Xiao and G.-H Hu, “An optimization method for integrating two kinds of preference information in group decision-making,” **Comput. Ind. Eng.,** Vol. 46, pp. 329–335, 2004.
13. P.C. Fishburn, **The Theory of Social Choice,** Princeton University Press, 1973.
14. P.C. Fishburn, **Utility Theory for Decision Making,** Robert E. Krieger Publishing Company, 1979.
15. P.C. Fishburn, “Preference structures and their numerical representations,” **Theoretical Computer Science** Vol. 217, pp. 359–383, 1999.
16. J.C. Fodor and M. Roubens, “Fuzzy preference modelling and multicriteria decision support.” Kluwer, Dordrecht, 1994.
17. J.C. Fodor, R.R. Yager and A. Rybalov, “Structure of Uninorms,” **Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems,** Vol. 5, pp. 411–127, 1997.
18. J.L. García-Lapresta and R.A. Marques, “Constructing reciprocal and stable aggregation operators,” Proceedings of AGOP’2003, University of Alcalá de Henares, Madrid, Spain, pp. 73–78, 2003.
19. J.L. García-Lapresta L.C. Meneses, “An empirical analysis of transitivity with four scaled preferential judgment modalities,” **Review of Economic Design,** Vol. 8, pp. 335–346, 2003.
20. J.L. García-Lapresta L.C. Meneses, “ Individual valued preferences and their aggregation: Analysis of consistency in a real case,” **Fuzzy Sets Syst.,** Vol. 151, pp. 269–284, 2005.
21. J.L. García-Lapresta and J. Montero, “Consistency in preference modelling.” In: **Modern Information Processing: From Theory to Applications** (eds. B. Bouchon-Meunier, G. Coletti, R.R. Yager), Elsevier, pp. 87–97, 2006.
22. F. Herrera, E. Herrera-Viedma and F. Chiclana, “Multiperson decision making based on multiplicative preference relations,” **Eur. J. Oper. Res.,** Vol. 129, pp. 372–385, 2001.
23. F. Herera, L. Martínez and P.J. Sánchez, “Managing non-homogeneous information in group decision making,” **Eur. J. Oper. Res.,** Vol. 166, pp. 115–132, 2005.
24. E. Herrera-Viedma, S. Alonso, F. Chiclana and F. Herrera, “A consensus model for group decision making with incomplete fuzzy preference relations,” **IEEE Trans. Fuzzy Syst.,** accepted for publication 2006.
25. E. Herrera-Viedma, F. Chiclana, F. Herrera and S. Alonso, “A group decision-making model with incomplete fuzzy preference relations based on additive consistency,” **IEEE Trans. Syst., Man, Cybern. B: Cybernetics,** accepted for publication 2006.
26. E. Herera-Viedma, F. Herrera, F. Chiclana and M. Luque, “Some issues on consistency of fuzzy preference relations,” **Eur. J. Oper. Res.,** Vol. 154, No. 1, pp. 98–109, Apr. 2004.
27. E. Herrera-Viedma, L. Martínez, F. Mata and F. Chiclana, “A consensus support system model for group decision-making problems with multi-granular linguistic preference relations,” **IEEE Trans. Fuzzy Syst.,** Vol. 13, No. 5, pp. 644–658, 2005.
28. E.P. Klement, R. Mesiar and E. Pap, “On the relationship of associative compensatory operators to triangular norms and conorms,”**Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems,** Vol. 4, pp. 129–144, 1996.
29. W. Kołodziejczyk, “Orlovsky’s concept of decision-making with fuzzy preference relations – Further results,” **Fuzzy Sets Syst.,** Vol. 19, pp. 11–20, 1986.
30. R.D. Luce and P. Suppes, “Preferences, Utility and Subject Probability,” in: R.D. Luce et al. (Eds.) *Handbook of Mathematical Psychology*, Vol. III, Wiley, New York, pp. 249–410, 1965.
31. J. Ma, Z-P. Fan,Y-P. Jiang, J.Y. Mao and L. Ma, “A method for repairing the inconsistency of fuzzy preference relations,” **Fuzzy Sets Syst.,** Vol. 157, No. 1, pp. 20–33, 2006.
32. G.H. Marakas, **Decision Support Systems in the 21th Century,** 2nd edition, New Jersey: Pearson Education, Inc., 2003.
33. Marimin, M. Umano, I. Hatono and H. Tamure, “Linguistic labels for expressing fuzzy preference relations in fuzzy group decision making”, **IEEE Trans. Syst., Man, Cybern. B: Cybernetics,** Vol. 28, pp. 205–218, 1998.

34. G.A. Miller, "The magical number seven or minus two: some limits on our capacity of processing information", *Psychological Rev.,* Vol. 63, pp. 81–97, 1956.
35. F.J. Montero and J. Tejada, "Some problems on the definition of fuzzy preference relations," **Fuzzy Sets Syst.,** Vol. 20, pp. 45–53, 1986.
36. F.J. Montero and J. Tejada, "A necessary and sufficient condition for the existence of Orlovsky's choice set," **Fuzzy Sets Syst.,** Vol. 26, pp. 121–125, 1988.
37. S.A. Orlovski, "Decision-making with fuzzy preference relations," **Fuzzy Sets Syst.,** Vol. 1, pp. 155–167, 1978.
38. S. Ovchinnikov and M. Roubens, "On fuzzy stric preference, indifference, and incomparability relations," **Fuzzy Sets Syst.,** Vol. 49, No. 1, pp. 15–20, 1992.
39. P. Perny and M. Roubens, "Fuzzy preference modeling." In R. Slowinski (Edt.) **Fuzzy sets in decision analysis, operation research and statistics** (Massachusetts: Kluwer Academic Publishers, 1998) pp. 3–30.
40. Th. L. Saaty, **The Analytic Hierarchy Process,** McGraw-Hill, New York, 1980.
41. Th.L. Saaty, **Fundamentals of Decision Making and Priority Theory with the AHP,** Pittsburg: RWS Publications, 1994.
42. F. Seo and M. Sakawa, "Fuzzy Multiattribute Utility Analysis for Collective Choice," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 15, No. 1, pp. 45–53, 1985.
43. A.K. Sen, "Social Choice Theory : A Re-Examination," **Econometrica,** Vol. 45, pp. 53–89, 1977.
44. W. Silvert, "Symmetric summation: A class of operations on fuzzy sets," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 9, pp. 657–659, 1979.
45. T. Tanino, "Fuzzy preference orderings in group decision making," **Fuzzy Sets Syst.,** Vol. 12, pp. 117–131, 1984.
46. T. Tanino, "Fuzzy preference relations in group decision making," in J. Kacprzyk and M. Roubens (Eds.) **Non-Conventional Preference Relations in Decision Making,** Springer-Verlag, Berlin, 1988, pp. 54–71.
47. T. Tanino, "On Group Decision Making Under Fuzzy Preferences," in: J. Kacprzyk and M. Fedrizzi (Eds.) *Multiperson Decision Making Using Fuzzy Sets and Possibility Theory*, Kluwer Academic Publishers, Dordrech, pp. 172–185, 1990.
48. X. Wang, "An investigation into relations between some transitivity-related concepts," **Fuzzy Sets Syst.,** Vol. 89, pp. 257–262, 1997.
49. Q. Zhang, J.C.H. Chen, Y.-Q. He, J. Ma and D.-N Zhou, "Multiple attribute decision making: approach integrating subjective and objective information," **Int. J. Manuf. Technol. Manag.,** Vol. 5, pp. 338–361, 2003.
50. Q. Zhang, J.C.H. Chen and P.P. Chong, "Decision consolidation: criteria weight determination using multiple preference formats," **Decis. Support Syst.,** Vol. 38, pp. 247–258, 2004.
51. L. A. Zadeh, "Similarity relations and fuzzy orderings," **Information Sciences** Vol. 3, pp. 177–200, 1971.
52. H.-J. Zimmermann, **Fuzzy Set Theory and Its Applications.** Kluwer, Dordrecht, 1991.

# Fuzzy Set Extensions of the Dominance-Based Rough Set Approach

**Salvatore Greco, Benedetto Matarazzo and Roman Słowiński**

**Abstract** Rough set theory has been proposed by Pawlak in the early 80s to deal with inconsistency problems following from information granulation. It operates on an information table composed of a set $U$ of objects described by a set $Q$ of condition and decision attributes. Decision attributes make a partition of $U$ into decision classes. Basic concepts of rough set theory are: indiscernibility relation on $U$, lower and upper (rough) approximations of decision classes, dependence and reduction of attributes from $Q$, and decision rules induced from rough approximations of decision classes. The original rough set idea was failing, however, to handle preferential ordering of domains of attributes (scales of criteria), as well as preferential ordering of decision classes. In order to deal with multiple criteria decision problems a number of methodological changes to the original rough set theory were necessary. The main change is the substitution of the indiscernibility relation by a dominance relation, which permits approximation of ordered sets. In multiple criteria decision context, the information table is composed of decision examples given by a decision maker. The Dominance-based Rough Set Approach (DRSA) applied to this information table results with a set of decision rules, being a preference model of the decision maker. It is more general than the classical multiple attribute utility model or outranking model, and it is more understandable because of its natural syntax. In this chapter, after recalling the classical rough set approach and DRSA, we review their fuzzy set extensions. Moreover, we characterize the dominance-based rough approximation of a fuzzy set, and we show that the classical rough approximation of a crisp set is its particular case. In this sense, DRSA is also relevant in the case where preferences are not considered, but just a kind of monotonicity relating values of different attributes is meaningful for the analysis of data at hand. In general terms, monotonicity concerns relationship between different aspects of a phenomenon described by data: for example, the larger the house, the higher its price or the closer the house to the city centre, the higher its price. In this perspective, DRSA gives a very general framework for reasoning about data using only monotonicity relationships.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

# 1 Introduction

Rough set theory [19, 20] relies on the idea that some knowledge (data, information) is available about objects of a universe of discourse $U$. Thus, a subset of $U$ is defined on the base of the available knowledge about the objects and not on the base of information about membership or non-membership of the objects to the subset. For example, knowledge about patients suffering from a certain disease may contain information about body temperature, blood pressure, etc. All patients described by the same information are indiscernible in view of the available knowledge, and form groups of similar objects. These groups are called elementary sets, and can be considered as elementary building blocks of the available knowledge about patients. Elementary sets can be combined into compound concepts. For example, elementary sets of patients can be used to represent a set of patients suffering from a certain disease. Any union of elementary sets is called crisp set, while other sets are referred to as rough set. Each rough set has boundary line objects, i.e. objects which, in view of the available knowledge, cannot be classified with certainty as members of the set or of its complement. Therefore, in the rough set approach, any set is associated with a pair of crisp sets, called the lower and the upper approximation. Intuitively, the lower approximation consists of all objects which certainly belong to the set and the upper approximation contains all objects which possibly belong to the set. The difference between the upper and the lower approximation constitutes the boundary region of the rough set. Analogously, for a partition of universe $U$ into classes, one may consider rough approximation of the partition. It appeared to be particularly useful for analysis of classification problems, being the most common decision problems.

For algorithmic reasons, the rough set approach operates on an information table composed of a set $U$ of objects (actions) described by a set $Q$ of attributes. If in the set $Q$ disjoint sets ($C$ and $D$) of condition and decision attributes are distinguished, then the information table is called decision table. It is often assumed, without loss of generality, that set $D$ is a singleton $\{d\}$, and thus decision attribute $d$ makes a partition of set $U$ into decision classes. Data collected in such a decision table correspond to a multiple attribute classification problem. The classical rough set approach is naturally adapted to analysis of this type of decision problems, because the set of objects can be identified with examples of classification and it is possible to extract all the essential knowledge contained in the decision table using indiscernibility or similarity relations. However, as pointed out by Greco, Matarazzo and Słowiński (see e.g. [9, 12, 13, 22]), the classical rough set approach cannot extract all the essential knowledge contained in the decision table of multiple criteria classification (also called sorting) problems, i.e. problems of assigning a set of actions described by a set of criteria to one of predefined and ordered decision classes. Notwithstanding, in many real problems it is important to take into account the ordinal properties of the considered attributes.

For example, in bankruptcy risk evaluation, if the debt index (total debt/total activity) of company $A$ has a modest value, while the same index of company $B$ has a significant value, then, within the classical rough set approach, the two firms are just discernible, but no preference is given to one of them with reference to the attribute "debt ratio". In reality, from the point of view of the bankruptcy risk evaluation, it would be advisable to consider firm $A$ better than firm $B$, and not simply different (discernible). Therefore, the attribute "debt ratio" is a criterion. Let us observe that the rough set approach based on the use of indiscernibility or similarity relation is not able to capture a particular kind of inconsistency which may occur when in the decision table there is at least one criterion. For instance, in the bankruptcy risk evaluation, which is a classification (sorting) problem, if firm $A$ is better than firm $B$ with respect to all the considered criteria (e.g. debt ratio, return on equity, etc.) but firm $A$ is assigned to a class of a higher risk than firm $B$, then there is an inconsistency which cannot be captured by the classical rough set approach, because these firms are discernible. In order to detect this inconsistency, the rough approximation should handle the ordinal properties of criteria. This can be made by replacing the indiscernibility or similarity relation by the dominance relation, which is a very natural concept within multiple criteria decision analysis.

On the basis of these considerations, Greco, Matarazzo and Słowiński (see e.g. [9, 12, 13, 22]) have proposed a new rough set approach to multiple criteria classification problems, called the Dominance-based Rough Set Approach (DRSA). Even if DRSA has been proposed to deal with ordinal properties of data related to preferences in decision problems, the concept of dominance-based rough approximation can be used in a much more general context [14]. This is because the monotonicity, which is crucial for DRSA, is also meaningful for problems where preferences are not considered. Generally, monotonicity concerns relationship between different aspects of a phenomenon described by data. More specifically, it concerns mutual trends between different variables like distance and gravity in physics, or inflation rate and interest rate in economics. Whenever we discover a relationship between different aspects of a phenomenon, this relationship can be represented by a monotonicity with respect to some specific measures of these aspects. So, in general, monotonicity is a property translating in a formal language a primitive intuition of interaction between different concepts of our knowledge. Within classical rough set approach, the idea of mononicity is not evident, although it is also present there. Because of very coarse representation of considered concepts, monotonicity is taken into account in the sense of presence or absence of particular aspects characterizing the concepts. Thus, the classical rough set approach involves the idea of monotonicity related to a scale with only two values: "presence" and "absence". Monotonicity gains importance when a finer representation of the concepts is considered. A representation is finer when, for each aspect characterizing concepts, not only its presence or its absence is taken into account, but also the degree of its presence or absence is considered relevant. Due to graduality, the idea of monotonicity can be exploited in the whole range of its potential.

Graduality is typical for fuzzy set philosophy [37] and, therefore, a joint consideration of rough sets and fuzzy sets is worthwhile. In fact, rough sets and fuzzy sets capture the two basic complementary aspects of monotonicity: rough sets deal with relationships between different concepts and fuzzy sets deal with expression of different dimensions in which the concepts are considered. For this reason, many approaches have been proposed to combine fuzzy sets with rough sets (see for example [1, 3, 21, 4]). The main preoccupation in almost all the studies combining rough sets with fuzzy sets was related to a fuzzy extension of Pawlak's definition of lower and upper approximations using fuzzy connectives [5, 16]. DRSA can also be combined with fuzzy sets along this line, obtaining a rough set model permitting to deal with fuzziness in preference representation [6, 9, 11]. Let us remark, however, that in fact there is no rule for the choice of the "right" connective, so this choice is always arbitrary to some extent. Moreover, there is another drawback for fuzzy extensions of rough sets involving fuzzy connectives: they are based on cardinal properties of membership degrees. In consequence, the result of these extensions is sensitive to order preserving transformation of membership degrees.

For example, consider the t-conorm of Łukasiewicz as fuzzy connective; it may be used in the definition of both fuzzy lower approximation (to build fuzzy implication) and fuzzy upper approximation (as a fuzzy counterpart of a union). The t-conorm of Łukasiewicz is defined as:

$$T^*(\alpha, \beta) = min\{\alpha + \beta, 1\}, \quad \alpha, \beta \in [0, 1]$$

$T^*(\alpha, \beta)$ can be interpreted as follows. If $\alpha = \mu_X(z)$ represents the membership of $z$ in set $X$ and $\beta = \mu_Y(z)$ represents the membership of $z$ in set $Y$, then $T^*(\alpha, \beta)$ expresses the membership of $z$ in set $X \cup Y$. Given two fuzzy propositions $p$ and $q$, putting $v(p) = \alpha$ and $v(q) = \beta$, $T^*(\alpha, \beta)$ can be interpreted also as $v(p \vee q)$, the truth value of the proposition $p \vee q$. Let us consider the following values of arguments:

$$\alpha = 0.5, \quad \beta = 0.3, \quad \gamma = 0.2, \quad \delta = 0.1$$

and their order preserving transformation:

$$\alpha' = 0.4, \quad \beta' = 0.3, \quad \gamma' = 0.2, \quad \delta' = 0.05.$$

The values of the t-conorm are in the two cases as follows:

$$T^*(\alpha, \delta) = 0.6, \quad T^*(\beta, \gamma) = 0.5, \quad T^*(\alpha', \delta') = 0.45, \quad T^*(\beta', \gamma') = 0.5.$$

One can see that the order of the results has changed after the order preserving transformation of the arguments. This means that the Łukasiewicz t-conorm takes into account not only the ordinal properties of the membership degrees, but also their cardinal properties. A natural question arises: is it reasonable to expect from the membership degree a cardinal content instead of ordinal only? Or, in other words, is it realistic to claim that a human is able to say in a meaningful way not only that

a) "object $x$ belongs to fuzzy set $X$ more likely than object $y$"
   (or "proposition $p$ is more credible than proposition $q$"),
   but even something like
b) "object $x$ belongs to fuzzy set $X$ two times
   more likely than object $y$"
   (or "proposition $p$ is two times more credible than proposition $q$")?

It is safer, of course, to consider information of type a), because information of type b) is rather meaningless for a human (see [17]). The above doubt about the cardinal content of the fuzzy membership degree shows the need for methodologies which consider the imprecision in perception typical for fuzzy sets but avoid as much as possible meaningless transformation of information through fuzzy connectives.

The DRSA approach we are proposing for a fuzzy extension of rough sets takes into account the above request. It avoids arbitrary choice of fuzzy connectives and not meaningful operations on membership degrees. This approach belongs to the minority of fuzzy extensions of the rough set concept that do not involve fuzzy connectives and cardinal interpretation of membership degrees. Within this minority, it is related to the approach of Nakamura and Gao [18] using $\alpha$-cuts on fuzzy similarity relation between objects. The DRSA approach to fuzzy extension of rough sets proposes a methodology of fuzzy rough approximation that infers the most cautious conclusion from available imprecise information. In particular, we observe that any approximation of knowledge about $Y$ using knowledge about $X$ is based on positive or negative relationships between premises and conclusions, i.e.:

i) "the more $x$ is $X$, the more it is $Y$" (positive relationship),
ii) "the more $x$ is $X$, the less it is $Y$" (negative relationship).

The following simple relationships illustrate $i)$ and $ii)$:

- "the larger the market share of a company, the greater its profit" (positive relationship) and
- "the greater the debt of a company, the smaller its profit" (negative relationship).

These relationships have been already considered within fuzzy set theory under the name of gradual decision rules [2]. Recently, Greco, Inuiguchi and Słowiński [7, 8] proposed a rough set approach for induction of gradual decision rules handling ambiguity of information through fuzzy rough approximations. Examples of these decision rules are: "if a car is speedy with credibility at least 0.8 and it has high fuel consumption with credibility at most 0.7, then it is a good car with a credibility at least 0.9", and "if a car is speedy with credibility at most 0.5 and it has high fuel consumption with credibility at least 0.8, then it is a good car with a credibility at most 0.6". Remark that the syntax of gradual decision rules is based on monotonic relationship between degrees of credibility that can also be found in dominance-based decision rules induced from preference-ordered data. This explains why one can build a fuzzy rough approximation using DRSA. In this perspective, DRSA can be extended to analyze any relationship of monotonicity in reasoning about data. This extension results in knowledge representation model composed of a set of gradual decision rules. The general character of fuzzy rough approximation based on DRSA is confirmed by the fact that the classical rough set approach can be seen as one its special cases. On one hand this allows a deeper insight into fundamental properties of the classical rough set approach and, on the other hand, it permits a further generalization of the rough set approach.

This chapter is organized as follows. Sect. 2 introduces classical rough set approach and its fuzzy set extensions based on fuzzy connectives. Sect. 3 presents Dominance-based Rough Set Approach. Sect. 4 describes fuzzy extensions of DRSA based on fuzzy connectives. Sect. 5 presents rough approximation of a fuzzy set based on the property of monotonicity. In Sect. 6, the monotonic fuzzy rough set is compared with the classical rough set, showing that the latter is a particular case of the former. Sect. 7 contains conclusions.

## 2 Classical Rough Set Approach and its Fuzzy Set Extensions

Formally, an **information table** is the 4-tuple $S =< U, Q, V, f >$, where $U$ is a finite set of **objects** (universe), $Q = \{q_1, q_2, ..., q_m\}$ is a finite set of **attributes**, $V_q$ is the domain of the attribute $q$, $V = \bigcup_{q \in Q} V_q$ and $f : U \times Q \to V$ is a total function such that $f(x, q) \in V_q$ for each $x \in U$ and $q \in Q$, called **information function**. If in the set $Q$ disjoint sets ($C$ and $D$) of **condition** and **decision** attributes are distinguished, then the information table is called **decision table**. It is often assumed, without loss of generality, that set $D$ is a singleton $\{d\}$, and thus decision attribute $d$ defines a classification (partition) of set $U$ into **decision classes**. Therefore, each object $x$ from $U$ is described by a vector (string)

$$Des_Q(x) = [f(x, q_1), f(x, q_2), ..., f(x, q_m)],$$

called description of $x$ in terms of the evaluations on the attributes from $Q$; it represents the available information (qualitative or quantitative) about $x$. Obviously, $x \in U$ can be described in terms of any non-empty subset $P \subseteq Q$.

With every (non-empty) subset of attributes $P$ there is associated an indiscernibility relation on $U$, denoted by $I_P$:

$$I_P = \{(x, y) \in U \times U : f(x, q) = f(y, q), \forall q \in P\}.$$

If $(x, y) \in I_P$, it is said that objects $x$ and $y$ are $P$-indiscernible. Clearly, the indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric and transitive). The family of all the equivalence classes of relation $I_P$ is denoted by $U|I_P$, and the equivalence class containing object $x \in U$, by $I_P(x)$. The equivalence classes of relation $I_P$ are called $P$-elementary sets. Let $S$ be an information table, $X$ a non-empty subset of $U$ and $\emptyset \neq P \subseteq Q$. The $P$-lower approximation and the $P$-upper approximation of $X$ in $S$ are defined, respectively, as:

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\},$$
$$\overline{P}(X) = \{x \in U : I_P(x) \cap X \neq \emptyset\}.$$

The elements of $\underline{P}(X)$ are all and only those objects $x \in U$ which belong to the equivalence classes generated by the indiscernibility relation $I_P$, contained in $X$; the elements of $\overline{P}(X)$ are all and only those objects $x \in U$ which belong to the equivalence classes generated by the indiscernibility relation $I_P$, containing at least one object $x$ belonging to $X$. In other words, $\underline{P}(X)$ is the largest union of the $P$-elementary sets included in $X$, while $\overline{P}(X)$ is the smallest union of the $P$-elementary sets containing $X$. The $P$-**boundary** of $X$ in $S$, denoted by $Bn_P(X)$, is $Bn_P(X) = \overline{P}(X) - \underline{P}(X)$. The following **inclusion property** holds:

$$\underline{P}(X) \subseteq X \subseteq \overline{P}(X).$$

Therefore, if object $x$ belongs to $\underline{P}(X)$, it is certainly an element of $X$, while if $x$ belongs to $\overline{P}(X)$, it is only possibly an element of $X$. $Bn_P(X)$ constitutes the "doubtful region" of $X$: nothing can be said with certainty about the membership of its elements in set $X$. The following relation, called **complementarity property**, is satisfied:

$$\underline{P}(X) = U - \overline{P}(U - X), \quad \overline{P}(X) = U - \underline{P}(U - X).$$

Another important property of rough approximation is the following: for all $P \subseteq R \subseteq C$ and for all $X \subseteq U$:

$$\underline{P}(X) \subseteq \underline{R}(X), \quad \overline{P}(X) \subseteq \overline{R}(X).$$

This property of **monotonicity with respect to sets of attributes** says that enlarging the set of attributes, i.e. using more information, one can get better approximations characterized by not smaller lower approximations and not larger upper approximations. If the $P$-boundary of $X$ is empty, $Bn_P(X) = \emptyset$, then set $X$ is an ordinary (crisp) set with respect to $P$, that is, it may be expressed as a union of some $P$-elementary sets; otherwise, if $Bn_P(X) \neq \emptyset$, set $X$ is an approximate (rough) set with respect to $P$ and may be characterized by means of the lower and upper approximations, $\underline{P}(X)$ and $\overline{P}(X)$. The family of all sets $X \subseteq U$ having the same $P$-lower and $P$-upper approximations is called a rough set. The definition of approximations of a subset $X \subseteq U$ can be extended to a classification, i.e. a partition $\mathbf{Y} = \{Y_1, Y_2, ..., Y_n\}$ of $U$. Subsets $Y_i$, $i = 1, ..., n$, are disjunctive classes of $\mathbf{Y}$. By $P$-lower ($P$-upper) approximation of $\mathbf{Y}$ in $S$ we mean sets $\underline{P}(\mathbf{Y}) = \{\underline{P}(Y_1), \underline{P}(Y_2), ..., \underline{P}(Y_n)\}$ and $\overline{P}(\mathbf{Y}) = \{\overline{P}(Y_1), \overline{P}(Y_2), ..., \overline{P}(Y_n)\}$, respectively. Using rough approximation of a classification one can induce certain, or approximate knowledge from the decision table, represented in terms of **decision rules**, i.e. logical statements (consequence relations) of the type "if..., then...", where the antecedent (condition part) specifies values assumed by one or more condition attributes and the consequence (decision part) specifies an assignment to one or more decision classes. Therefore, the general syntax of a rule is the following:

"if $f(x, q_1)$ is equal to $r_{q_1}$ and $f(x, q_2)$ is equal to $r_{q_2}$ and ...
$f(x, q_p)$ is equal to $r_{q_p}$, then $x$ belongs to $Y_{j_1}$ or $Y_{j_2}$ or ... $Y_{j_k}$",

where $\{q_1, q_2, ..., q_p\} \subseteq C$, $(r_{q_1}, r_{q_2}, ..., r_{q_p}) \in V_{q_1} \times V_{q_2} \times ... \times V_{q_p}$ and $Y_{j_1}, Y_{j_2}, ..., Y_{j_k}$ are decision classes of the considered classification ($D$-elemen-tary sets). If the consequence is univocal, i.e. $k$=1, then the rule is certain, otherwise it is approximate.

A further step towards generalization of the rough approximations consists in considering fuzzy reflexive binary relation $R(x, y)$ defined on $U$, that is a relation of fuzzy similarity. To make this generalization, we need negation and some classical connectives of fuzzy logic. The following definitions of fuzzy logic are useful (see, e.g., [5, 16]). For each proposition $p$, we consider its truth value $v(p)$ ranging from $v(p) = 0$ ($p$ is definitely false) to $v(p) = 1$ ($p$ is definitely true); and for all intermediate values, the greater $v(p)$, the more credible is the truth of $p$. A negation is a non-increasing function $N : [0, 1] \to [0, 1]$ such that $N(0) = 1$ and $N(1) = 0$. Given proposition $p$, $N(v(p))$ states the credibility of the negation of $p$. A t-norm $T$ and a t-conorm $T^*$ are two functions $T : [0, 1] \times [0, 1] \to [0, 1]$ and $T^* : [0, 1] \times [0, 1] \to [0, 1]$, such that given two propositions, $p$ and $q$, $T(v(p), v(q))$ represents the credibility of the conjunction of $p$ and $q$, and $T^*(v(p), v(q))$ represents the credibility of the disjunction of $p$ and $q$. t-norm $T$ and t-conorm $T^*$ must satisfy the following properties:

$$T(\alpha, \beta) = T(\beta, \alpha) \text{ and } T^*(\alpha, \beta) = T^*(\beta, \alpha), \text{ for all } \alpha, \beta \in [0, 1],$$

$$T(\alpha, \beta) \leq T(\gamma, \delta) \text{ and } T^*(\alpha, \beta) \leq T^*(\gamma, \delta), \text{ for all } \alpha, \beta, \gamma, \delta \in [0, 1]$$

$$\text{such that } \alpha \leq \gamma \text{ and } \beta \leq \delta,$$

$$T(\alpha, T(\beta, \gamma)) = T(T(\alpha, \beta), \gamma) \text{ and } T^*(\alpha, T^*(\beta, \gamma)) = T^*(T^*(\alpha, \beta), \gamma),$$

$$\text{for all } \alpha, \beta, \gamma \in [0, 1],$$

$$T(1, \alpha) = \alpha \text{ and } T^*(0, \alpha) = \alpha, \text{ for all } \alpha \in [0, 1].$$

A negation is strict iff it is strictly decreasing and continuous. A negation $N$ is involutive iff, for all $\alpha \in [0, 1]$, $N(N(\alpha)) = \alpha$. $(T, T^*, N)$ is a De Morgan triplet iff $N(T^*(\alpha, \beta)) = T(N(\alpha), N(\beta))$, where $N$ is a strict negation. A fuzzy implication is a function $I : [0, 1] \times [0, 1] \to [0, 1]$ such that given two propositions, $p$ and $q$, $I(v(p), v(q))$ represents the credibility of the implication of $q$ by $p$. A fuzzy implication must satisfy the following properties ([5]):

$$I(\alpha, \beta) \geq I(\gamma, \beta) \text{ for all } \alpha, \beta, \gamma \in [0, 1], \text{ such that } \alpha \leq \gamma,$$

$$I(\alpha, \beta) \geq I(\alpha, \gamma) \text{ for all } \alpha, \beta, \gamma \in [0, 1], \text{ such that } \beta \geq \gamma,$$

$$I(0, \alpha) = 1, I(\alpha, 1) = 1 \text{ for all } \alpha \in [0, 1],$$

$$I(1,0) = 0.$$

An implication $I^{\rightarrow}_{N,T^*}$ is a $T^*$-implication if there is a t-conorm $T^*$ and a strong negation $N$ such that $I^{\rightarrow}_{N,T^*}(\alpha,\beta) = T^*(N(\alpha),\beta)$. A fuzzy similarity on the universe $U$ is a fuzzy binary relation (i.e. function $R : U \times U \rightarrow [0,1]$) reflexive ($R(x,x) = 1$ for all $x \in U$), symmetric ($R(x,y) = R(y,x)$ for all $x, y \in U$) and transitive (given t-norm $T$, $T(R(x,y), R(y,z)) = R(x,z)$ for all $x, y, z \in U$). Greco, Matarazzo and Słowiński [9, 10] considered fuzzy similarity being a reflexive fuzzy relation but not necessarily symmetric and transitive. Greco, Matarazzo and Słowiński [9] proposed the following definition of rough approximation based on fuzzy similarity. The lower and upper approximations of a fuzzy set $X$ on $U$ having membership function $\mu_X : U \rightarrow [0,1]$, are fuzzy sets $\underline{R}(X)$ and $\overline{R}(X)$ in $U$ with membership functions $\mu(x, \underline{R}(X))$ and $\mu(x, \overline{R}(X))$, specifying for every $x \in U$ its degree of membership to $\underline{R}(X)$ and $\overline{R}(X)$. These degrees are equivalent, respectively, to the credibility of the following statements: "for every $y \in U$, $x$ is not similar to $y$ or $y$ belongs to $X$" (since "$p$ implies $q$" is logically equivalent to "not $p$ or $q$", where $p$ = "for every $y \in U$, $x$ is similar to $y$" and $q$ = "$y$ belongs to $X$"), and "there exists at least one $y \in U$ such that $x$ is similar to $y$ and $y$ belongs to $X$". Formally, these two membership functions are defined as:

$$\mu(x, \underline{R}(X)) = T_{y\in U}\left(T^*(N(R(x,y)), \mu_X(y))\right),$$
$$\mu(x, \overline{R}(X)) = T^*_{y\in U}\left(T(R(x,y), \mu_X(y))\right). \quad (i)$$

Let us remark that using the definition of $T^*$-implication (i.e. $I^{\rightarrow}_{N,T^*}(\alpha,\beta) = T^*(N(\alpha),\beta))\ \forall \alpha, \beta \in [0,1]$), it is possible to rewrite the definition of $\mu(x, \underline{R}(X))$ and $\mu(x, \overline{R}(X))$ in the following way:

$$\mu(x, \underline{R}(X)) = T_{y\in U}\left(I^{\rightarrow}_{N,T^*}(R(x,y), \mu_X(y))\right),$$
$$\mu(x, \overline{R}(X)) = T^*_{y\in U}\left(N(I^{\rightarrow}_{N,T^*}(R(x,y), N(\mu_X(y))))\right).$$

Therefore, $\mu(x, \underline{R}(X))$ can be interpreted as the credibility of the statement "for each $y \in U$, the similarity of $x$ to $y$ implies that $y$ belongs to $X$", while $\mu(x, \overline{R}(X))$ can be interpreted as the credibility of the statement "for at least one $y \in U$, the similarity of $x$ to $y$ does not imply that $y$ does not belongs to $X$". Let us observe that the above definition can also be given using a general fuzzy implication $I$ instead of a $T^*$-implication in the lower approximation $\mu(x, \underline{R}(X))$, while keeping definition $(i)$ for the upper approximation $\mu(x, \overline{R}(X))$, i.e.

$$\mu(x, \underline{R}(X)) = T_{y\in U}\left(I(R(x,y), \mu_X(y))\right),$$
$$\mu(x, \overline{R}(X)) = T^*_{y\in U}\left(T(R(x,y), \mu_X(y))\right).$$

Another possible definition of fuzzy rough approximation has been proposed by Dubois and Prade [1, 3] (see also [21]):

$$\mu(x, \underline{R}(X)) = inf_{y \in U} \left(I(R(x, y), \mu_X(y))\right),$$
$$\mu(x, \overline{R}(X)) = sup_{y \in U} \left(T(R(x, y), \mu_X(y))\right).$$

Let us observe that if set $U$ is finite (and thus infimum over $U$ becomes minimum over $U$, and supremum over $U$ becomes maximum over $U$) and we consider $T(\alpha, \beta) = min(\alpha, \beta)$, $T^*(\alpha, \beta) = max(\alpha, \beta)$, our fuzzy rough approximations coincide with that of Dubois and Prade for $T^*$-implication, and boil down to

$$\mu(x, \underline{R}(X)) = min_{y \in U} \left(max(N(R(x, y)), \mu_X(y))\right),$$
$$\mu(x, \overline{R}(X)) = max_{y \in U} \left(min(R(x, y), \mu_X(y))\right).$$

It may be demonstrated, in particular, that the following properties of inclusion and complementarity hold:

1) $\mu(x, \underline{R}(X)) \le \mu_X(x) \le \mu(x, \overline{R}(X))$, for all $x \in U$,
2) if $(T, T*, N)$ is a De Morgan triplet, $N$ is involutive and the considered implication is a $T^*$-implication, then

$$\mu(x, \underline{R}(X)) = N(\mu(x, \overline{R}(U - X))), \quad \mu(x, \overline{R}(X)) = N(\mu(x, \underline{R}(U - X))).$$

where $U - X$ represents the fuzzy set whose membership function, $\forall x \in U$, has the form $\mu_{U-X}(x) = N(\mu(x))$.

The above results refer to fuzzy sets and can be read as follows: 1) means that $X$ includes its lower approximation and is included in its upper approximation (inclusion property); 2) means that the lower approximation of $X$ is the complement of the upper approximation of its complementary set (complementarity property). Given set $C$ of condition attributes, let us consider a fuzzy binary relation $R_q$ for each attribute $q \in C$, i.e. function $R_q : U \times U \rightarrow [0, 1]$, where, $\forall x, y \in U$, $R_q(x, y)$ represents the intensity or degree of similarity of $x$ to $y$ with respect to attribute $q$. More precisely, for $q \in C$ and $\forall x, y, w, z \in U$:

- $R_q(x, y) = 0$ means that $x$ is not similar at all to $y$,
- $R_q(x, y) = 1$ means that $x$ is absolutely similar to $y$, $(R_q(x, x) = 1)$,
- $R_q(x, y) \ge R_q(w, z)$ means that the similarity of $x$ to $y$ is at least as credible as the similarity of $w$ to $z$.

To model the comprehensive similarity of $x$ to $y$, $x, y \in U$ with respect to a subset of condition attributes $P = \{q_1, q_2, ..., q_p\} \subseteq C$, denoted by $R_P(x, y)$, one can use function $\Psi_P : [0, 1]^p \rightarrow [0, 1]$, non-decreasing in each of its arguments, verifying $\Psi_P(0, ..., 0) = 0$ and $\Psi_P(1, ..., 1) = 1$, such that, for all $x, y \in U$,

$$R_P(x, y) = \Psi_P(R_1(x, y), ..., R_p(x, y)).$$

Let us observe that in the above definition we considered function $\Psi_P$ for each $P \subseteq C$. The following **invariance property** (invariance with respect to similarity on attributes where maximal value is attained) relating functions $\Psi_{P_1}$ and $\Psi_{P_2}$, $\emptyset \neq P_1 \subseteq P_2 \subseteq C$ with $P_1 = \{q_1, q_2, ..., q_{p-k}\}$, $P_2 = \{q_1, q_2, ..., q_p\}$, $k > 0$, seems very reasonable,

$$\Psi_{P_1}(\alpha_1, ..., \alpha_{p-k}, 1, ..., 1) = \Psi_{P_2}(\alpha_1, ..., \alpha_{p-k}). \quad (ii)$$

This property says that, if for $x, y \in U$ the similarity with respect to attributes from $P_1 - P_2$ is maximal, that is $R_q(x, y) = 1$ for $q = p - k + 1, ..., p$, then the comprehensive similarity with respect to $P_2$ is equivalent to the similarity with respect to $P_1$, i.e.

$$R_{P_1}(x, y) = R_{P_2}(x, y).$$

A possible formulation of function $\Psi_P$ is the following. We consider the credibility of the statement "$x$ is similar to $y$ with respect to $q_1$, and $x$ is similar to $y$ with respect to $q_2$, and ... $x$ is similar to $y$ with respect to $q_p$". Using a t-norm, this credibility is calculated as:

$$R_P(x, y) = T_{q \in P}\left(R_q(x, y)\right),$$

i.e. for all $(\alpha_1, ..., \alpha_p) \in [0, 1]^p$, $\Psi_P(\alpha_1, ..., \alpha_p) = T_{q=1}^{p}\alpha_q$. This formulation of comprehensive similarity $R_P$ satisfies the property of invariance. If invariance property is satisfied, then the following property of monotonicity of approximations with respect to sets of attributes also holds: for all $P_1 \subseteq P_2 \subseteq C$, for all fuzzy sets $X$ on $U$ and for all $x \in U$

$$\mu(x, \underline{R}_{P_1}(X)) \geq \mu(x, \underline{R}_{P_2}(X)), \quad \mu(x, \overline{R}_{P_1}(X)) \leq \mu(x, \overline{R}_{P_2}(X)).$$

Using fuzzy rough approximation of a classification one can induce decision rules from the decision table, having the following general syntax:

"if $f(x, q_1)$ is similar to $r_{q_1}$ and $f(x, q_2)$ is similar to $r_{q_1}$ and ... $f(x, q_p)$ is similar to $r_{q_p}$, then $x$ belongs to $Y_{j_1}$ or $Y_{j_2}$ or ... $Y_{j_k}$", with a credibility $\alpha$,

where $q_1, q_2, ..., q_p \subseteq C$, $(r_{q_1}, r_{q_2}, ..., r_{q_p}) \in V_{q_1} \times V_{q_2} \times ... \times V_{q_p}$, $Y_{j_1}, Y_{j_2}, ... Y_{j_k}$ are some decision classes of the considered classification and $\alpha \in [0, 1]$.

## 3 Dominance-Based Rough Set Approach

In this section we present the main concepts of the Dominance-based Rough Set approach (DRSA) (for a more complete presentation see, for example, [9, 12, 13, 22].

We are considering condition attributes with domains (scales) ordered according to decreasing or increasing preference – such attributes are called *criteria*. For criterion $q \in Q$, $\succeq_q$ is a *weak preference* relation on $U$ such that $x \succeq_q y$ means "$x$ is at least as good as $y$ with respect to criterion $q$". We suppose that $\succeq_q$ is a complete preorder, i.e. a strongly complete and transitive binary relation, defined on $U$ on the basis of evaluations $f(\cdot, q)$. We assume, without loss of generality, that the preference is increasing with the value of $f(\cdot, q)$ for every criterion $q \in C$.

Furthermore, we assume that the set of decision attributes $D$ is a singleton $\{d\}$. Decision attribute $d$ makes a partition of $U$ into a finite number of decision classes, $\boldsymbol{Cl}=\{Cl_t, t \in T\}$, $T = \{1, ..., n\}$, such that each $x \in U$ belongs to one and only one class $Cl_t \in \boldsymbol{Cl}$. We suppose that the classes are preference-ordered, i.e. for all $r, s \in T$, such that $r > s$, the objects from $Cl_r$ are preferred to the objects from $Cl_s$. More formally, if $\succeq$ is a *comprehensive weak preference relation* on $U$, i.e. if for all $x,y \in U$, $x \succeq y$ means "$x$ is at least as good as $y$", we suppose: $[x \in Cl_r, y \in Cl_s, r>s] \Rightarrow [x \succeq y$ and *not* $y \succeq x]$. The above assumptions are typical for consideration of a *multiple criteria classification problem* (also called multiple criteria sorting problem).

The sets to be approximated are called *upward union* and *downward union* of classes, respectively:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \quad Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, \quad t = 1, ..., n.$$

The statement $x \in Cl_t^{\geq}$ means "$x$ belongs to at least class $Cl_t$", while $x \in Cl_t^{\leq}$ means "$x$ belongs to at most class $Cl_t$". Let us remark that $Cl_1^{\geq} = Cl_n^{\leq} = U$, $Cl_n^{\geq} = Cl_n$ and $Cl_1^{\leq} = Cl_1$. Furthermore, for $t = 2,...,n$, we have:

$$Cl_{t-1}^{\leq} = U - Cl_t^{\geq} \text{ and } Cl_t^{\geq} = U - Cl_{t-1}^{\leq}.$$

The key idea of the rough set approach is representation (approximation) of knowledge generated by decision attributes, by "*granules of knowledge*" generated by condition attributes.

In DRSA, where condition attributes are criteria and decision classes are preference ordered, the represented knowledge is a collection of *upward* and *downward unions of classes* and the "granules of knowledge" are sets of objects defined using a dominance relation.

We say that $x$ *dominates* $y$ with respect to $P \subseteq C$ (shortly, $x$ *P-dominates* $y$), denoted by $xD_P y$, if for every criterion $q \in P$, $f(x, q) \geq f(y, q)$. The relation of $P$-dominance is reflexive and transitive, that is it is a partial preorder.

Given a set of criteria $P \subseteq C$ and $x \in U$, the "granules of knowledge" used for approximation in DRSA are:

- a set of objects dominating $x$, called *$P$-dominating set*, $D_P^+(x)=\{y \in U: yD_P x\}$,
- a set of objects dominated by $x$, called *$P$-dominated set*, $D_P^-(x)=\{y \in U: xD_P y\}$.

Let us recall that the *dominance principle* (or Pareto principle) requires that an object $x$ dominating object $y$ on all considered criteria (i.e. $x$ having evaluations at least as good as $y$ on all considered criteria) should also dominate $y$ on the decision (i.e. $x$ should be assigned to at least as good decision class as $y$). This principle is the only objective principle that is widely agreed upon in the multiple criteria comparisons of objects.

The *$P$-lower approximation* of $Cl_t^{\geq}$, denoted by $\underline{P}\, Cl_t^{\geq}$, and the *$P$-upper approximation* of $Cl_t^{\geq}$, denoted by $\overline{P}(Cl_t^{\geq})$, are defined as follows ($t = 1,...,n$):

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\},$$
$$\overline{P}(Cl_t^{\geq}) = \{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}.$$

Analogously, one can define the *$P$-lower approximation* and the *$P$-upper approximation* of $Cl_t^{\leq}$ as follows ($t = 1,...,n$):

$$\underline{P}(Cl_t^{\leq}) = \{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\},$$
$$\overline{P}(Cl_t^{\leq}) = \{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}.$$

The $P$-lower and $P$-upper approximations so defined satisfy the following *inclusion properties* for each $t \in \{1,\ldots,n\}$ and for all $P \subseteq C$:

$$\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \quad \underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}).$$

The $P-$lower and $P-$upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ have an important *complementarity property*, according to which,

$$\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq}) \text{ and } \overline{P}(Cl_t^{\geq}) = U - \underline{P}(Cl_{t-1}^{\leq}),\ t = 2,...,n,$$

$$\underline{P}(Cl_t^{\leq}) = U - \overline{P}(Cl_{t+1}^{\geq}) \text{ and } \overline{P}(Cl_t^{\leq}) = U - \underline{P}(Cl_{t+1}^{\geq}),\ t = 1,...,n-1.$$

The *$P$-boundary* of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$, respectively, are defined as follows ($t$=1,..., $n$):

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \quad Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}).$$

The dominance-based rough approximations of upward and downward unions of classes can serve to induce "*if..., then...*" decision rules. It is meaningful to consider the following five types of decision rules:

1) Certain $D_{\geq}$-decision rules: *if* $x_{q1} \succeq_{q1} r_{q1}$ *and* $x_{q2} \succeq_{q2} r_{q2}$ *and* $\ldots$ $x_{qp} \succeq_{qp} r_{qp}$, *then* $x \in Cl_t^{\geq}$, where for each $w_q$,$z_q \in X_q$, "$w_q \succeq_q z_q$" means "$w_q$ is at least as good as $z_q$".
2) Possible $D_{\geq}$-decision rules: *if* $x_{q1} \succeq_{q1} r_{q1}$ *and* $x_{q2} \succeq_{q2} r_{q2}$ *and* $\ldots$ $x_{qp} \succeq_{qp} r_{qp}$, *then x possibly belongs to* $Cl_t^{\geq}$.
3) Certain $D_{\leq}$-decision rules: *if* $x_{q1} \preceq_{q1} r_{q1}$ *and* $x_{q2} \preceq_{q2} r_{q2}$ *and* $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, *then* $x \in Cl_t^{\leq}$, where for each $w_q$,$z_q \in X_q$, "$w_q \preceq_q z_q$" means "$w_q$ is at most as good as $z_q$".
4) Possible $D_{\leq}$-decision rules: *if* $x_{q1} \preceq_{q1} r_{q1}$ *and* $x_{q2} \preceq_{q2} r_{q2}$ *and* $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, *then x possibly belongs to* $Cl_t^{\leq}$.
5) Approximate $D_{\geq\leq}$-decision rules: *if* $x_{q1} \succeq_{q1} r_{q1}$ *and* $\ldots$ $x_{qk} \succeq_{qk} r_{qk}$ *and* $x_{q(k+1)} \preceq_{q(k+1)} r_{q(k+1)}$ *and* $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, *then* $x \in Cl_s^{\geq} \cap Cl_t^{\leq}$, where $s < t$.

The rules of type 1) and 3) represent certain knowledge extracted from the decision table, while the rules of type 2) and 4) represent possible knowledge. Rules of type 5) represent doubtful knowledge.

## 4 Fuzzy Set Extensions of the Dominance-Based Rough Set Approach

The concept of dominance can be refined by introducing gradedness through the use of fuzzy sets. Let $\succeq_q$ be a fuzzy weak preference relation on $U$ with respect to criterion $q \in C$, i.e. $\succeq_q$: $U \times U \to [0, 1]$, such that $\succeq_q(x, y)$ represents the credibility of the proposition "$x$ is at least as good as $y$ with respect to criterion $q$". Suppose that $\succeq_q$ is a fuzzy partial $T$-preorder, i.e. that it is reflexive ($\succeq_q(x, x) = 1$ for each $x \in U$) and $T$-transitive ($T(\succeq_q(x, y), \succeq_q(y, z)) \leq \succeq_q(x, z)$, for each $x, y, z \in U$) (see Fodor and Roubens 1992). Using the fuzzy weak preference (outranking) relations $\succeq_q$, $q \in C$, a fuzzy dominance relation on $U$ (denotation $D_P(x, y)$) can be defined for all $P \subseteq C$ as follows:

$$D_P(x, y) = T_{q \in P} \left( \succeq_q (x, y) \right).$$

Given $(x, y) \in U \times U$, $D_P(x, y)$ represents the credibility of the proposition "$x$ is at least as good as $y$ with respect to each criterion $q$ from $P$". Since the fuzzy weak preference relations $\succeq_q$ are supposed to be partial $T$-preorders, then also the fuzzy dominance relation $D_P$ is a partial $T$-preorder. Furthermore, let $Cl = \{Cl_t, t \in T\}$, $T = \{1, ..., n\}$, be a set of fuzzy classes in $U$, such that for each $x \in U$, $Cl_t(x)$ represents the membership function of $x$ to $Cl_t$. We suppose, as before, that the classes of $Cl$ are increasingly ordered, i.e. that for all $r, s \in T$, such that $r > s$, the objects from $Cl_r$ have a better comprehensive evaluation than the objects from $Cl_s$. On the basis of the membership functions of the fuzzy class $Cl_t$, we can define fuzzy membership functions of two other sets:

1) the upward cumulated fuzzy set $Cl_t^{\geq}$, whose membership function $Cl_t^{\geq}(x)$ represents the credibility of the proposition "$x$ is at least as good as the objects in $Cl_t$"

$$Cl_t^{\geq}(x) = \left\{ \begin{array}{cc} 1 & \text{if } \exists s \in T : Cl_s(x) > 0 \text{ and } s > t \\ Cl_t(x) & \text{otherwise} \end{array} \right\}$$

2) the downward cumulated fuzzy set $Cl_t^{\leq}$, whose membership function $Cl_t^{\leq}(x)$ represents the credibility of the proposition "$x$ is at most as good as the objects in $Cl_t$"

$$Cl_t^{\leq}(x) = \left\{ \begin{array}{cc} 1 & \text{if } \exists s \in T : Cl_s(x) > 0 \text{ and } s < t \\ Cl_t(x) & \text{otherwise} \end{array} \right\}$$

The $P$-lower and the $P$-upper approximations of $Cl_t^{\geq}$ with respect to $P \subseteq C$ are fuzzy sets in $U$, whose membership functions, denoted by $\underline{P}[Cl_t^{\geq}(x)]$ and $\overline{P}[Cl_t^{\geq}(x)]$, are defined as:

$$\underline{P}[Cl_t^{\geq}(x)] = T_{y \in U}(T^*(N(D_P(y,x)), Cl_t^{\geq}(y))),$$

$$\overline{P}[Cl_t^{\geq}(x)] = T^*_{y \in U}(T(D_P(x,y), Cl_t^{\geq}(y))).$$

$\underline{P}[Cl_t^{\geq}(x)]$ represents the credibility of the proposition "for all $y \in U$, $y$ does not dominate $x$ with respect to criteria from $P$ or $y$ belongs to $Cl_t^{\geq}$", while $\overline{P}[Cl_t^{\geq}(x)]$ represents the credibility of the proposition "there is at least one $y \in U$ dominated by $x$ with respect to criteria from $P$ which belongs to $Cl_t^{\geq}$".

The $P$-lower and $P$-upper approximations of $Cl_t^{\leq}$ with respect to $P \subseteq C$, denoted by $\underline{P}[Cl_t^{\leq}(x)]$ and $\overline{P}[Cl_t^{\leq}(x)]$, can be defined, analogously, as:

$$\underline{P}[Cl_t^{\leq}(x)] = T_{y \in U}(T^*(N(D_P(x,y)), Cl_t^{\leq}(y))),$$

$$\overline{P}[Cl_t^{\leq}(x)] = T^*_{y \in U}(T(D_P(y,x), Cl_t^{\leq}(y))).$$

$\underline{P}[Cl_t^{\leq}(x)]$ represents the credibility of the proposition "for all $y \in U$, $x$ does not dominate $y$ with respect to criteria from $P$ or $y$ belongs to $Cl_t^{\leq}$", while $\overline{P}[Cl_t^{\leq}(x)]$ represents the credibility of the proposition "there is at least one $y \in U$ dominating $x$ with respect to criteria from $P$ which belongs to $Cl_t^{\leq}$".

Let us remark that using the definition of the $T^*$-implication, it is possible to rewrite the definition of $\underline{P}[Cl_t^{\geq}(x)]$, $\overline{P}[Cl_t^{\geq}(x)]$, $\underline{P}[Cl_t^{\leq}(x)]$ and $\overline{P}[Cl_t^{\leq}(x)]$ in the following way:

$$\underline{P}[Cl_t^{\geq}(x)] = T_{y \in U}(I^{\rightarrow}_{T^*,N}(D_P(y,x), Cl_t^{\geq}(y))),$$

$$\overline{P}[Cl_t^{\geq}(x)] = T^*_{y\in U}(N(I^{\rightarrow}_{T^*,N}(D_P(x,y), N(Cl_t^{\geq}(y))))),$$

$$\underline{P}[Cl_t^{\leq}(x)] = T_{y\in U}(I^{\rightarrow}_{T^*,N}(D_P(x,y), Cl_t^{\leq}(y))),$$

$$\overline{P}[Cl_t^{\leq}(x)] = T^*_{y\in U}(N(I^{\rightarrow}_{T^*,N}(D_P(y,x), N(Cl_t^{\leq}(y))))).$$

The following results can be proved:

1) for each $x \in U$ and for each $t \in T$,
$\underline{P}[Cl_t^{\geq}(x)] \leq Cl_t^{\geq}(x) \leq \overline{P}[Cl_t^{\geq}(x)]$, $\quad \underline{P}[Cl_t^{\leq}(x)] \leq Cl_t^{\leq}(x) \leq \overline{P}[Cl_t^{\leq}(x)]$,
2) if $(T, T^*, N)$ constitute a De Morgan triplet, if negation $N$ is involutive and if $N[Cl_t^{\geq}(x)] = Cl_{t-1}^{\geq}(x)$ for each $x \in U$ and $t = 2, ..., n-1$, then

$$\underline{P}[Cl_t^{\geq}(x)] = N(\overline{P}[Cl_{t-1}^{\leq}(x)]),\ \overline{P}[Cl_t^{\geq}(x)] = N(\underline{P}[Cl_{t-1}^{\leq}(x)]),\ t = 2, ..., n$$

$$\underline{P}[Cl_t^{\leq}(x)] = N(\overline{P}[Cl_{t+1}^{\geq}(x)]),\ \overline{P}[Cl_t^{\leq}(x)] = N(\underline{P}[Cl_{t+1}^{\geq}(x)]),\ t = 1, ..., n-1$$

3) for all $P \subseteq R \subseteq C$, for all $x \in U$ and for each $t \in T$,

$$\underline{P}[Cl_t^{\geq}(x)] \leq \underline{R}[Cl_t^{\geq}(x)], \quad \overline{P}[Cl_t^{\geq}(x)] \geq \overline{R}[Cl_t^{\geq}(x)],$$

$$\underline{P}[Cl_t^{\leq}(x)] \leq \underline{R}[Cl_t^{\leq}(x)], \quad \overline{P}[Cl_t^{\leq}(x)] \geq \overline{R}[Cl_t^{\leq}(x)].$$

Results 1) to 3) can be read as fuzzy counterparts of the following results well-known within the classical rough set approach: 1) (inclusion property) says that $Cl_t^{\geq}$ and $Cl_t^{\leq}$ include their $P$-lower approximation and are included in their $P$-upper approximation; 2) (complementarity property) says that the $P$-lower ($P$-upper) approximation of $Cl_t^{\geq}$ is the complement of the $P$-upper ($P$-lower) approximation of its complementary set $Cl_{t-1}^{\leq}$, (analogous property holds for $Cl_t^{\leq}$ and $Cl_{t+1}^{\geq}$; 3) (monotonicity with respect to sets of attributes) says that enlarging the set of criteria the membership to the lower approximation does not decrease and the membership to the upper approximation does not increase.

Greco, Inuiguchi and Słowiński [6] proposed, moreover, the following fuzzy rough approximation based on dominance, which goes in line of the fuzzy rough approximation by Dubois and Prade [1, 3], concerning classical rough sets (see Sect. 2):

$$\underline{P}[Cl_t^{\geq}(x)] = inf_{y\in U}(I(D_P(y,x), Cl_t^{\geq}(y))),$$

$$\overline{P}[Cl_t^{\geq}(x)] = sup_{y\in U}(T(D_P(x,y), Cl_t^{\geq}(y))),$$

$$\underline{P}[Cl_t^{\leq}(x)] = inf_{y\in U}(I(D_P(x, y), Cl_t^{\leq}(y))),$$

$$\underline{P}[Cl_t^{\leq}(x)] = sup_{y\in U}(T(D_P(y, x), Cl_t^{\leq}(y))).$$

Using fuzzy rough approximations based on DRSA, one can induce decision rules having the same syntax as the decision rules obtained from crisp DRSA. In this case, however, each decision rule has a fuzzy credibility.

## 5 Dominance-Based Rough Approximation of a Fuzzy Set

In this section we show how the dominance-based rough set approach can be used for rough approximation of fuzzy sets.

A *fuzzy information base* is the 3-tuple $\boldsymbol{B} =< U, F, \varphi >$, where $U$ is a finite set of *objects* (universe), $F$={$f_1$,$f_2$,...,$f_m$} is a finite set of *features*, and $\varphi : U \times F \rightarrow [0, 1]$ is a function such that $\varphi(x, f_h) \in [0, 1]$ expresses the credibility that object $x$ has feature $f_h$. Each object $x$ from $U$ is described by a vector

$$Des_F(x)=[\varphi(x,f_1), \ldots, \varphi(x,f_m)]$$

called *description* of $x$ in terms of the evaluations of the features from $F$; it represents the available information about $x$. Obviously, $x \in U$ can be described in terms of any non-empty subset $E \subseteq F$ and in this case we have

$$Des_E(x)= [\varphi(x,f_h), f_h \in F].$$

For any $E \subseteq F$, we can define the dominance relation $D_E$ as follows: for any $x$,$y \in U$, $x$ dominates $y$ with respect to $E$ (denotation $xD_Ey$) if for any $f_h \in E$

$$\varphi(x, f_h) \geq \varphi(y, f_h).$$

Given $E \subseteq F$ and $x \in U$, let

$$D_E^+(x) = \{y \in U : yD_Ex\}, \quad D_E^-(x) = \{y \in U : xD_Ey\}.$$

Let us consider a fuzzy set $X$ in $U$, with its membership function $\mu_X : U \rightarrow [0, 1]$. For each cutting level $\alpha \in [0, 1]$ and for $* \in \{\geq, >\}$, we can define the $E$-lower and the $E$-upper approximation of $X^{*\alpha}$={$y \in U : \mu_X(y) * \alpha$} with respect to $E \subseteq F$ (denotation $\underline{E}(X^{*\alpha})$ and $\overline{E}(X^{*\alpha})$, respectively), as:

$$\underline{E}(X^{*\alpha}) = \{x \in U : D_E^+(x) \subseteq X^{*\alpha}\} \qquad = \bigcup_{x\in U}\{D_E^+(x) : D_E^+(x) \subseteq X^{*\alpha}\},$$

$$\overline{E}(X^{*\alpha}) = \{x \in U : D_E^-(x) \cap X^{*\alpha} \neq \emptyset\} = \bigcup_{x\in U}\{D_E^+(x) : D_E^-(x) \cap X^{*\alpha} \neq \emptyset\}.$$

Analogously, for each cutting level $\alpha \in [0,1]$ and for $\diamond \in \{\leq, <\}$, we define the $E$-lower and the $E$-upper approximation of $X^{\diamond\alpha}=\{y \in U : \mu_X(y) \diamond \alpha\}$, with respect to $E \subseteq F$ (denotation $\underline{E}(X^{\diamond\alpha})$ and $\overline{E}(X^{\diamond\alpha})$, respectively), as:

$$\underline{E}(X^{\diamond\alpha}) = \{x \in U : D_E^-(x) \subseteq X^{\diamond\alpha}\} \qquad = \bigcup_{x\in U}\{D_E^-(x) : D_E^-(x) \subseteq X^{\diamond\alpha}\},$$

$$\overline{E}(X^{\diamond\alpha}) = \{x \in U : D_E^+(x) \cap X^{\diamond\alpha} \neq \emptyset\} = \bigcup_{x\in U}\{D_E^-(x) : D_E^+(x) \cap X^{\diamond\alpha} \neq \emptyset\}.$$

Let us remark that we can rewrite the rough approximations $\underline{E}(X^{\geq\alpha})$, $\overline{E}(X^{\geq\alpha})$, $\underline{E}(X^{\leq\alpha})$ and $\overline{E}(X^{\leq\alpha})$ as follows:

$$\underline{E}(X^{\geq\alpha}) = \{x \in U : \forall w \in U, wD_E x \Rightarrow w \in X^{\geq\alpha}\},$$

$$\overline{E}(X^{\geq\alpha}) = \{x \in U : \exists w \in U \text{ such that } wD_E x \text{ and } w \in X^{\geq\alpha}\},$$

$$\underline{E}(X^{\leq\alpha}) = \{x \in U : \forall w \in U, xD_E w \Rightarrow w \in X^{\leq\alpha}\},$$

$$\overline{E}(X^{\leq\alpha}) = \{x \in U : \exists w \in U \text{ such that } xD_E w \text{ and } w \in X^{\leq\alpha}\}.$$

Rough approximations $\underline{E}(X^{>\alpha})$, $\overline{E}(X^{>\alpha})$, $\underline{E}(X^{<\alpha})$ and $\overline{E}(X^{<\alpha})$ can be rewritten analogously by the simple substitution of "$\geq$" with "$>$" and "$\leq$ "with "$<$".

Let us remark that in the above approximations, even if $X^{\geq\alpha}= Y^{\leq\alpha}$, their approximations are, in general, different due to the different directions of cutting the membership functions of $X$ and $Y$. Of course, a similar remark holds also for $X^{<\alpha}$ and $Y^{>\alpha}$. Considerations of the directions in the cuts $X^{\leq\alpha}$, $X^{<\alpha}$ and $X^{\geq\alpha}$, $X^{>\alpha}$ are important in the definition of the rough approximations of unions and intersection of cuts.

The rough approximations $\underline{E}(X^{\geq\alpha})$, $\overline{E}(X^{\geq\alpha})$, $\underline{E}(X^{\leq\alpha})$, $\overline{E}(X^{\leq\alpha})$ and $\underline{E}(X^{>\alpha})$, $\overline{E}(X^{>\alpha})$, $\underline{E}(X^{<\alpha})$, $\overline{E}(X^{<\alpha})$ satisfy the following inclusion properties: for any $0 \leq \alpha \leq 1$

$$\underline{E}(X^{\geq\alpha}) \subseteq X^{\geq\alpha} \subseteq \overline{E}(X^{\geq\alpha}), \quad \underline{E}(X^{\leq\alpha}) \subseteq X^{\leq\alpha} \subseteq \overline{E}(X^{\leq\alpha}),$$

$$\underline{E}(X^{>\alpha}) \subseteq X^{>\alpha}\overline{E}(X^{>\alpha}), \quad \underline{E}(X^{<\alpha}) \subseteq X^{<\alpha} \subseteq \overline{E}(X^{<\alpha}).$$

Furthermore, the following complementary properties hold: for any $0 \leq \alpha \leq 1$

$$\underline{E}(X^{\geq\alpha}) = U - \overline{E}(X^{<\alpha}), \quad \underline{E}(X^{\leq\alpha}) = U - \overline{E}(X^{>\alpha}),$$

$$\underline{E}(X^{>\alpha}) = U - \overline{E}(X^{\leq\alpha}), \quad \underline{E}(X^{<\alpha}) = U - \overline{E}(X^{\geq\alpha}).$$

The following property of monotonicity with respect to sets of features also holds: for any $E_1 \subseteq E_2 \subseteq F$ and for any $0 \leq \alpha \leq 1$

$$\underline{E}_1(X^{\geq\alpha}) \subseteq \underline{E}_2(X^{\geq\alpha}), \quad \underline{E}_1(X^{>\alpha}) \subseteq \underline{E}_2(X^{>\alpha}),$$

$$\underline{E}_1(X^{\leq\alpha}) \subseteq \underline{E}_2(X^{\leq\alpha}), \quad \underline{E}_1(X^{<\alpha}) \subseteq \underline{E}_2(X^{<\alpha}),$$

$$\overline{E}_1(X^{\geq\alpha}) \supseteq \overline{E}_2(X^{\geq\alpha}), \quad \overline{E}_1(X^{>\alpha}) \supseteq \overline{E}_2(X^{>\alpha}),$$

$$\overline{E}_1(X^{\leq\alpha}) \supseteq \overline{E}_2(X^{\leq\alpha}), \quad \overline{E}_1(X^{<\alpha}) \supseteq \overline{E}_2(X^{<\alpha}).$$

We consider also fuzzy rough approximations $\underline{X}_E^{\uparrow}$, $\underline{X}_E^{\downarrow}$ , $\overline{X}_E^{\uparrow}$, $\overline{X}_E^{\downarrow}$, which are fuzzy sets with membership functions defined, respectively, as follows: for any $y \in U$,

$$\mu_{\underline{X}_E^{\uparrow}}(y) = max\{\alpha \in [0,1] : y \in \underline{E}(X^{\geq\alpha})\},$$
$$\mu_{\underline{X}_E^{\downarrow}}(y) = min\{\alpha \in [0,1] : y \in \underline{E}(X^{\leq\alpha})\},$$
$$\mu_{\overline{X}_E^{\uparrow}}(y) = max\{\alpha \in [0,1] : y \in \overline{E}(X^{\geq\alpha})\},$$
$$\mu_{\overline{X}_E^{\downarrow}}(y) = min\{\alpha \in [0,1] : y \in \overline{E}(X^{\leq\alpha})\}.$$

$\mu_{\underline{X}_E^{\uparrow}}(y)$ is defined as the upward lower fuzzy rough approximation of $X$ with respect to $E$ and can be interpreted in the following way. For any $\alpha, \beta \in [0,1]$ we have that $\alpha < \beta$ implies $X^{\geq\alpha} \supseteq X^{\geq\beta}$. Therefore, the greater the cutting level $\alpha$, the smaller $X^{\geq\alpha}$ and, consequently, the smaller also its lower approximation $\underline{E}(X^{\geq\alpha})$. Thus, for each $y \in U$ and for each fuzzy set $X$, there is a threshold $k(y)$, $0 \leq k(y) \leq \mu_X(y)$, such that $y \in \underline{E}(X^{\geq\alpha})$ if $\alpha \leq k(y)$, and $y \notin \underline{E}(X^{\geq\alpha})$ if $\alpha > k(y)$. Since $k(y) = \mu_{\underline{X}_E^{\uparrow}}(y)$, this explains the interest of $\mu_{\underline{X}_E^{\uparrow}}(y)$. Analogous interpretation holds for $\mu_{\overline{X}_E^{\uparrow}}(y)$ defined as the upward upper fuzzy rough approximation of $X$ with respect to $E$.

$\mu_{\underline{X}_E^{\downarrow}}(y)$ is defined as the downward lower fuzzy rough approximation of $X$ with respect to $E$ and can be interpreted as follows. For any $\alpha, \beta \in [0,1]$ we have that $\alpha < \beta$ implies $X^{\leq\alpha} \subseteq X^{\leq\beta}$. Therefore, the greater the cutting level $\alpha$, the greater $X^{\leq\alpha}$ and, consequently, its lower approximation $\underline{E}(X^{\geq\alpha})$. Thus, for each $y \in U$ and for each fuzzy set $X$, there is a threshold $h(y)$, $\mu_X(y) \leq h(y) \leq 1$, such that $y \in \underline{E}(X^{\leq\alpha})$ if $\alpha \geq h(y)$, and $y \notin \underline{E}(X^{\leq\alpha})$ if $\alpha < h(y)$. We have that $h(y) = \mu_{\underline{X}_E^{\downarrow}}(y)$. Analogous interpretation holds for $\mu_{\overline{X}_E^{\downarrow}}(y)$ defined as the upward upper fuzzy rough approximation of $X$ with respect to $E$.

The upward and downward lower and upper fuzzy rough approximations can also be rewritten in the following equivalent formulation, which has been proposed and investigated by Greco, Inuiguchi and Słowiński [8]:

$$\mu_{\underline{X}_E^{\uparrow}}(y) = min\{\mu_X(z) : z \in D_E^+(y)\}, \quad \mu_{\overline{X}_E^{\uparrow}}(y) = max\{\mu_X(z) : z \in D_E^-(y)\},$$

$$\mu_{\underline{X}_E^{\downarrow}}(y) = max\{\mu_X(z) : z \in D_E^-(y)\}, \quad \mu_{\overline{X}_E^{\downarrow}}(y) = min\{\mu_X(z) : z \in D_E^+(y)\}.$$

The fuzzy rough approximations $\mu_{\underline{X}_E^{\uparrow}}(y)$, $\mu_{\overline{X}_E^{\uparrow}}(y)$, $\mu_{\underline{X}_E^{\downarrow}}(y)$ and $\mu_{\overline{X}_E^{\downarrow}}(y)$, satisfy the following inclusion properties: for any $y \in U$,

$$\mu_{\underline{X}_E^{\uparrow}}(y) \le \mu_X(y) \le \mu_{\overline{X}_E^{\uparrow}}(y), \quad \mu_{\underline{X}_E^{\downarrow}}(y) \le \mu_X(y) \le \mu_{\overline{X}_E^{\downarrow}}(y).$$

Furthermore, the following complementary properties hold: for any $y \in U$

$$\mu_{\underline{X}_E^{\uparrow}}(y) = \mu_{\overline{X}_E^{\downarrow}}(y), \quad \mu_{\underline{X}_E^{\downarrow}}(y) = \mu_{\overline{X}_E^{\uparrow}}(y).$$

The following property of monotonicity with respect to sets of features also holds: for any $E_1 \subseteq E_2 \subseteq F$ and for any $0 \le \alpha \le 1$

$$\mu_{\underline{X}_{E_1}^{\uparrow}}(y) \le \mu_{\underline{X}_{E_2}^{\uparrow}}(y), \quad \mu_{\underline{X}_{E_1}^{\downarrow}}(y) \ge \mu_{\underline{X}_{E_2}^{\downarrow}}(y),$$

$$\mu_{\overline{X}_{E_1}^{\uparrow}}(y) \ge \mu_{\overline{X}_{E_2}^{\uparrow}}(y), \quad \mu_{\overline{X}_{E_1}^{\downarrow}}(y) \le \mu_{\overline{X}_{E_2}^{\downarrow}}(y).$$

Using this way of fuzzy rough approximations based on DRSA, one can induce knowledge contained in the decision table in terms of decision rules with the following syntax:

*" if object y has feature $f_{i1}$ in degree at least $h_{i1}$ and has feature $f_{i2}$ in degree at least $h_{i2}$, ..., and has feature $f_{im}$ in degree at least $h_{im}$, then object y belongs to set X in degree at least $\alpha$",*

with $f_{i1}, f_{i2}, ..., f_{im} \in F$ and $h_{i1}, h_{i2}, ..., h_{im}, \alpha \in [0, 1]$.

# 6 Monotonic Rough Approximation of a Fuzzy Set Versus Classical Rough Set

What is the relationship between classical rough set and DRSA approximation of a fuzzy set? Greco, Matarazzo and Słowiński [14] proved that the former is a particular case of the latter. In the following we demonstrate this relationship.

Any information table can be expressed in terms of a specific type of an information base. An *information base* is called *Boolean* if $\varphi : U \times F \to \{0, 1\}$. A partition $\boldsymbol{F}=\{F_1,\dots,F_r\}$ of $F$, with *card*$(F_k) \ge 2$ for all $k=1,\dots,r$, is called *canonical* if, for each $x \in U$ and for each $F_k \subseteq F$, $k = 1,\dots,r$, there exists only one $f_j \in F_k$ for which $\varphi(x, f_j) = 1$ (and thus, for all $f_i \in F_k - \{f_j\}$, $\varphi(x, f_i) = 0$). The condition *card*$(F_k) \ge 2$ for all $k = 1,\dots,r$, is necessary because, otherwise, we would have at least one element of the partition $F_k = \{f'\}$ such that $\varphi(x, f') = 1$ for all $x \in U$, and this would mean that feature $f'$ gives no information and can be removed.

Now, we can see that any *information table* **S**=$< U, Q, V, f >$ can be interpreted as a Boolean information base **B**=$< U, F, \varphi >$ such that to each $v \in V_q$ corresponds one feature $f_{qv} \in F$ for which $\varphi(x, f_{qv}) = 1$ if $f(x, q) = v$, and $\varphi(x, f_{qv}) = 0$ otherwise. Let us remark that **F**=$\{F_1,\ldots,F_m\}$, with $F_q = \{f_{qv}, v \in V_q\}$, $q \in Q$, is a canonical partition of $F$. In other words, this means that each information system can be viewed as an information base where each possible value $v \in V$ of attribute $q$ corresponds to a specific feature $f_{qv}$. Let us remark that the vice versa is not true, i.e. there are Boolean information bases which cannot be transformed into information systems because their set of attributes does not admit any canonical partition, as shown by the following example.

*Example 1.* Let us consider a Boolean information base **B**, such that $U = \{x_1, x_2, x_3\}$, $F$=$\{f_1,f_2\}$ and function $\varphi$ is defined by Table 1. One can see that **F**=$\{\{f_1,f_2\}\}$ is not a canonical partition because $\varphi(x_3,f_1)$=$\varphi(x_3,f_2)$=1, while canonical partition **F** does not allow that for an object $x \in U$, $\varphi(x,f_1)$=$\varphi(x,f_2)$=1. Let us remark that also for the Boolean information base **B'**, such that $U$=$\{x_1,x_2,x_4\}$, $F$=$\{f_1,f_2\}$ and function $\varphi$ is defined by Table 2, **F**=$\{\{f_1,f_2\}\}$ is not a canonical partition because $\varphi(x_4,f_1)$=$\varphi(x_4,f_2)$=0, while canonical partition **F** does not allow that for an object $x \in U$, $\varphi(x,f_1)$=$\varphi(x,f_2)$=0.

**Table 1** Information base ***B***

| | $f_1$ | $f_2$ |
|---|---|---|
| $x_1$ | 0 | 1 |
| $x_2$ | 1 | 0 |
| $x_4$ | 1 | 1 |

**Table 2** Information base ***B'***

| | $f_1$ | $f_2$ |
|---|---|---|
| $x_1$ | 0 | 1 |
| $x_2$ | 1 | 0 |
| $x_4$ | 0 | 0 |

◇

The above means that the rough approximation in the context of a Boolean information base is more general than the rough set approximation in the context of an information system. Of course, the rough approximation in the context of a fuzzy information system is even more general than that in the context of a Boolean information base.

The equivalence between rough approximations in the context of a fuzzy information base and the classical definition of rough approximations in the context of an information system can be stated as follows. Let us consider an information system and the corresponding Boolean information base; for each $P \subseteq Q$, let $E^P$ be the set of all the features corresponding to values $v$ of attributes in $P$. Let $X$ be a crisp set in $U$ (i.e. $\mu_X : U \rightarrow \{0, 1\}$ and, therefore, for any $y \in U$, $\mu_X(y) = 1$ or $\mu_X(y) = 0$). Then, we have:

$$\underline{E^P}(X^{\geq 1}) = \underline{P}(X^{\geq 1}), \quad \overline{E^P}(X^{\geq 1}) = \overline{P}(X^{\geq 1}),$$

$$\underline{E^P}(X^{\leq 0}) = \underline{P}(U - X^{\geq 1}), \quad \overline{E^P}(X^{\leq 0}) = \overline{P}(U - X^{\geq 1}).$$

This result proves that the rough approximation of a crisp set $X$ within a Boolean information base admitting a canonical partition is equivalent to the classical rough approximation of set $X$ within the corresponding information system. Therefore, the classical rough approximation is a particular case of the rough approximation within a fuzzy information system.

## 7 Conclusions and Further Research Directions

In this chapter, after a brief review of the classical rough set approach and its fuzzy set extensions, we presented fuzzy set extensions of the Dominance-based Rough Set Approach (DRSA) and dominance-based rough approximations of fuzzy sets. The fuzzy set extensions of DRSA are based on fuzzy connectives, which is characteristic for almost all fuzzy rough set approaches. The dominance-based rough approximations of fuzzy sets infer, instead, the most cautious conclusions from available imprecise information, without using fuzzy connectives which are always arbitrary to some extent. Another advantage of dominance-based rough approximations of fuzzy sets is that they use only ordinal properties of membership degrees. Knowledge induced from dominance-based rough approximations of fuzzy sets is represented in terms of gradual decision rules. The dominance-based rough approximations of fuzzy sets generalize the classical rough approximations of crisp sets, as proved by showing that the classical rough set approach is one of its particular cases. We believe that, due to considering ordinal character of the graduality of fuzzy sets, without using any fuzzy connectives, dominance-based rough approximations of fuzzy sets give a new insight into both rough sets and fuzzy sets, and enable further generalizations of both of them. The recently proposed DRSA for fuzzy case-based reasoning is an example of this capacity [15].

**Acknowledgment** The research of the first two authors has been supported by the Italian Ministry of Education, University and Scientific Research (MIUR). The third author wishes to acknowledge financial support from the Ministry of Science and Higher Education (grant no. 3T11F 02127).

## References

1. Dubois, D., Prade, H., Rough fuzzy sets and fuzzy rough sets, *Internat. J. General Systems*, 17 (23) (1990) 191-209.
2. Dubois, D., Prade, H., Gradual inference rules in approximate reasoning, *Information Sciences*, 61 (1992) 103–122.

3. Dubois, D., Prade, H., Putting rough sets and fuzzy sets together, in R. Słowiński (ed.), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer, Dordrecht, 1992, pp. 203–232.
4. Dubois, D., Grzymala-Busse, J., Inuiguchi, M. and Polkowski, L. (eds.), *Transations on Rough Sets II: Rough Sets and Fuzzy Sets*, LNCS 3135, Springer- Verlag, Berlin, 2004.
5. Fodor, J., Roubens, M., *Fuzzy Preference Modelling and Multicriteria Decision Support*, Kluwer, Dordrecht, 1994.
6. Greco, S., Inuiguchi, M., Słowiński, R., Dominance-based rough set approach using possibility and necessity measures, in J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong (eds.), *Rough Sets and Current Trends in Computing*, LNAI 2475, Springer- Verlag, Berlin, 2002, pp. 85–92.
7. Greco, S. Inuiguchi, M., Słowiński, R., A new proposal for rough fuzzy approximations and decision rule representation, in D. Dubois, J. Grzymala-Busse, M. Inuiguchi and L. Polkowski (eds.), *Transations on Rough Sets II: Rough Sets and Fuzzy Sets*, LNCS 3135, Springer-Verlag, Berlin, 2004, pp. 156–164.
8. Greco, S., Inuiguchi, M., Słowiński, R., Fuzzy rough sets and multiple-premise gradual decision rules, *International Journal of Approximate Reasoning*, 41 (2006) 179–211.
9. Greco, S., Matarazzo, B., Słowiński, R., The use of rough sets and fuzzy sets in MCDM, chap. 14 in T.Gal, T.Stewart, T.Hanne (eds.), *Advances in Multiple Criteria Decision Making*, Kluwer Academic Publishers, Boston, 1999, pp. 14.1–14.59.
10. Greco, S., Matarazzo, B., Słowiński, R., Rough set processing of vague information using fuzzy similarity relations, in C. Calude and G. Paun (eds.), *From Finite to Infinite*, Springer-Verlag, Berlin, 2000, pp. 149–173.
11. Greco, S., Matarazzo, B., Słowiński, R., A fuzzy extension of the rough set approach to multicriteria and multiattribute sorting, in J. Fodor, B. De Baets and P. Perny, *Preferences and Decisions under Incomplete Information*, Physica-Verlag, Heidelberg, 2000, pp. 131–154.
12. Greco, S., Matarazzo, B., Słowiński R., Rough sets theory for multicriteria decision analysis, *European Journal of Operational Research*, 129 (2001) 1–47.
13. Greco, S., Matarazzo, B., Słowiński, R., Decision rule approach, in J. Figueira, S. Greco, M. Ehrgott (eds.) *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer-Verlag, Berlin, 2005, pp. 507–563.
14. Greco, S., Matarazzo, B., Słowiński, R., Generalizing rough set theory through Dominance-based Rough Set Approach, in D. Slezak, J. Yao, J. Peters, W. Ziarko, X. Hu (eds.), *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, LNAI 3642, Springer-Verlag, Berlin, 2005, pp. 1–11.
15. Greco, S., Matarazzo, B., Słowiński, R., Dominance-based Rough Set Approach to Case-Based Reasoning, in V. Torra, Y. Narukawa, A. Valls, J. Domingo-Ferrer (eds.), *Modelling Decisions for Artificial Intelligence*, LNAI 3885, Springer-Verlag, Berlin, 2006, pp. 7–18.
16. Klement, E. P., Mesiar, R., Pap, E., Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.
17. Marchant, T., The measurement of membership by comparisons, *Fuzzy Sets and Systems*, 148 (2004) 157–177.
18. Nakamura, A., Gao, J.M., A logic for fuzzy data analysis, *Fuzzy Sets and Systems*, 39 (1991) 127–132.
19. Pawlak, Z., Rough Sets, *International Journal of Computer and Information Sciences*, 11 (1982) 341–356.
20. Pawlak, Z., *Rough Sets*, Kluwer, Dordrecht, 1991.
21. Radzikowska, A. M., Kerre, E. E., A comparative study of fuzzy rough sets, *Fuzzy Sets and Systems*, 126 (2002) 137-155.
22. Słowiński, R., Greco, S., Matarazzo, B., Rough set based decision support, chap. 16 in E.K. Burke and G. Kendall (eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, Springer-Verlag, New York, 2005, pp. 475–527.
23. Zadeh, L., Fuzzy sets, *Information Control*, 8 (1965) 338–353.

# On Group Decision Making, Consensus Reaching, Voting and Voting Paradoxes under Fuzzy Preferences and a Fuzzy Majority: A Survey and some Perspectives

**Janusz Kacprzyk, Sławomir Zadrożny, Mario Fedrizzi and Hannu Nurmi**

**Abstract** Group decision making, as meant in this chapter, is the following choice problem which proceeds in a multiperson setting. There is a group of individuals (decisionmakers, experts, …) who provide their testimonies concerning an issue in question. These testimonies are assumed here to be individual preference relations over some set of option (alternatives, variants, …). The problem is to find a solution, i.e. an alternative or a set of alternatives, from among the feasible ones, which best reflects the preferences of the group of individuals as a whole. We will survey main developments in group decision making under fuzziness. First, we will briefly outline some basic inconsistencies and negative results of group decision making and social choice, and show how they can be alleviated by some plausible modifications of underlying assumptions, mainly by introducing fuzzy preference relations and, to a lesser extent, a fuzzy majority. Then, we will concentrate on how to derive solutions under individual fuzzy preference relations, and a fuzzy majority equated with a fuzzy linguistic quantifier (e.g., most, almost all, …) and dealt with in terms of a fuzzy logic based calculus of linguistically quantified statements or via the ordered weighted averaging (OWA) operators. We will briefly mention that one of solution concepts proposed can be a prototype for a wide class of group decision making choice functions. Then, we will discuss a related issue of how to define a "soft" degree of consensus in the group under individual fuzzy preference relations and a fuzzy majority. Finally, we will show how fuzzy preferences can help alleviate some voting paradoxes.

**Key words:** Fuzzy logic · Linguistic quantifier · Fuzzy preference relation · Fuzzy majority · Group decision making · Social choice · Consensus

## 1 Introduction

In this section we will first discus the very essence of group decision making and how fuzzy preferences and a fuzzy majority can help alleviate some inherent difficulties and make models more realistic. Then, we will briefly present some tools

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

to be used, notably how to deal with linguistically quantified statements, and with a linguistic quantifier driven aggregation.

### *1.1 Group Decision Making*

The essence of decision making, one of the most crucial and omnipresent human activities, is basically to find a best alternative (option, variant, ...) from among some feasible (relevant, available, ...) ones. It has been a subject of intensive research, notably formal, mathematical models have been devised to formalize the human rational behavior. Initially, this rationality has been equated with the maximization of some utility (value) function. Unfortunately, it has become more and more clear that the human behavior is rarely consistent with the maximization of a (expected) utility function, and some attempts to make decision making models more human consistent have been made, notably via a plausible modification of assumptions on, e.g., human preferences, axioms underlying the (expected) utility based approach, etc. – cf. Aizerman [1], many contributions in Kacprzyk and Fedrizzi [24], Kacprzyk and Roubens [51], Nurmi [34], etc. Potentials of fuzzy sets, in particular fuzzy relations, have been recognized quite early as well, cf. Blin [10], Blin and Whinston [11].

However, decision making in real world usually proceeds under multiple criteria, decisionmakers, stages, etc. In this paper we basically consider the case of multiperson decision making, more specifically group, practically from the perspective of social choice, under some fuzification of preferences and majority. We assume that there is a set of individuals who provide their testimonies assumed to be *preferences* over the set of alternatives. The problem is to find a *solution*, i.e. an alternative (or a set of alternatives) which is best acceptable by the group of individuals as a whole. For a different point of departure, involving choice sets or utility functions, we may refer the interested reader to, e.g., Kim [29], Salles [37], etc.

Since its very beginning group decision making has been plagued by negative results. Their essence is that no "rational" choice procedure satisfies all "natural", or plausible, requirements; so, each choice procedure has at least one serious drawback. By far the best known negative result is the so-called Arrow's impossibility theorem (cf. Arrow [2] or Kelly [56]). Another well known negative results are due to Gibbard and Satterthwaite, McKelvey, Schofield, etc. – cf. Nurmi [34]. Basically, all these negative results might be summarized as: no matter which group choice procedure we employed, it would satisfy one set of plausible conditions but not another set of equally plausible ones. Unfortunately, this general property pertains to all possible choice procedures, so that attempts to develop new, more sophisticated choice procedures do not seem very promising in this respect. Much more promising seems to be to modify some basic assumptions underlying the group decision making process. This line of reasoning is also basically assumed here.

A notable research direction is here based on the introduction of an *individual* and *social fuzzy preference relation*. Suppose that we have a set of $n \geq 2$ alternatives,

$S = \{s_1, \ldots, s_n\}$, and a set of $m \geq 2$ individuals, $E = \{1, \ldots, m\}$. Then, an individual's $k \in E$ individual fuzzy preference relation in $S \times S$ assigns a value in the unit interval for the preference of one alternative over another.

Normally, there are also some conditions to be satisfied, as, e.g., reflexivity, connectivity, (max-min) transitivity, etc. One should however note that it is not clear which of these "natural" properties of preference relations should be assumed. We will briefly discuss this issue in Sect. 2, but the interested reader should consult, e.g., Salles [37]. Moreover, a deep discussion is given in, e.g., Fodor and Roubens' [15], and also in De Baets et al.'s paper in this volume.

In this paper we assume that the individual and social fuzzy preference relations are defined in $S \times S$, i.e. assign to each pair of alternatives a strength of preference of one over another as a value from [0, 1]. Sometimes a better solution would be to assume the values of the strength of preference belonging to some ordered set (e.g. a set of linguistic values). This gives rise to some non-standard notions of soft preferences, orderings, etc. The best source for information on these and other related topics is Salles [37], and among the new approaches, the ones due to Herrera et al. [27, 28, 29, 30, 31, 32, 33] are particularly worth mentioning. The fuzzy preferences will be employed only instrumentally, and we will not discuss them and their properties in more detail.

Another basic element underlying group decision making is the concept of a *majority* – notice that a solution is to be an alternative (or alternatives) best acceptable by the group as a whole, that is by (at least!) *most* of its members since in practically no real nontrivial situation it would be accepted by all.

Some of the above mentioned problems, or negative result, with group decision making are closely related to too strict a representation of majority (e.g., at least a half, at least 2/3, ...). A natural line of reasoning is to somehow make that strict concept of majority closer to its real human perception by making it more vague. A good, often cited example in a biological context may be found in Loewer and Laddaga [62]:

> " ...It can correctly be said that there is a consensus among biologists that Darwinian natural selection is an important cause of evolution though there is currently no consensus concerning Gould's hypothesis of speciation. This means that there is a widespread agreement among biologists concerning the first matter but disagreement concerning the second ..."

and it is clear that a rigid majority as, e.g., more than 75% would evidently not reflect the essence of the above statement. However, it should be noted that there are naturally situations when a strict majority is necessary, for obvious reasons, as in all political elections.

A natural manifestations of such a "soft" majority are the so-called *linguistic quantifiers* as, e.g., most, almost all, much more than a half, etc. Such linguistic quantifiers can be, fortunately enough, dealt with by fuzzy-logic-based calculi of linguistically quantified statements as proposed by Zadeh [86]. Moreover, Yager's [84] ordered weighted averaging (OWA) operators can be used for this purpose (cf. Yager and Kacprzyk [85]), and also some other tools as, e.g., the Choquet integral.

In this paper we will present how fuzzy preference relations and fuzzy majorities can be employed for deriving solution of group decision making, and of degrees of consensus. We also mention some approaches to the alleviation of some voting paradoxes.

## *1.2 Fuzzy Linguistic Quantifiers and the Ordered Weighted Averaging (OWA) Operators for a Linguistic Quantifier Driven Aggregation*

Our notation is standard. A fuzzy set $A$ in $X = \{x\}$, will be characterized and equated with its membership function $\mu_A : X \longrightarrow [0, 1]$ such that $\mu_A(x) \in [0, 1]$ is the grade of membership of $x \in X$ in $A$, from full membership to full nonmembership, through all intermediate values. For a finite $X = \{x_1, \ldots, x_n\}$ we write $A = \mu_A(x_1)/x_1 + \cdots + \mu_A(x_n)/x_n$. Moreover, we denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. Other, more specific notation will be introduced when needed.

A *linguistically quantified statement*, e.g. "most experts are convinced", may be generally written as

$$Qy\text{'s are } F \tag{1}$$

where $Q$ is a linguistic quantifier (e.g., most), $Y = \{y\}$ is a set of objects (e.g., experts), and $F$ is a property (e.g., convinced).

We may assign to the particular $y$'s (objects) a different importance (relevance, competence, ...), $B$, which may be added to (1) yielding a *linguistically quantified statement with importance qualification* generally written as

$$QBy'\text{s are } F \tag{2}$$

which may be exemplified by "most ($Q$) of the important ($B$) experts ($y$'s) are convinced ($F$)".

From our point of view, the main problem is to find the truth of such linguistically quantified statements, i.e. truth($Qy$'s are $F$) or truth($QBy$'s are $F$) knowing truth($y$ is $F$), for each $y \in Y$. One can use different calculi but we will consider Zadeh's [86] and Yager's [84] OWA operators based calculi only.

### 1.2.1 A Fuzzy-logic-based Calculus of Linguistically Quantified Statements

In Zadeh's [86] method, a fuzzy linguistic quantifier $Q$ is assumed to be a fuzzy set defined in [0, 1]. For instance, $Q =$ "most" may be given as

$$\mu_Q(x) = \begin{cases} 1 & \text{for } x \geq 0.8 \\ 2x - 0.6 & \text{for } 0.3 < x < 0.8 \\ 0 & \text{for } x \leq 0.3 \end{cases} \tag{3}$$

which may be meant as that if at least 80% of some elements satisfy a property, then *most* of them certainly (to degree 1) satisfy it, when less than 30% of them satisfy it, then *most* of them certainly do not satisfy it (satisfy to degree 0), and between 30% and 80% – the more of them satisfy it the higher the degree of satisfaction by *most* of the elements.

This is a *proportional* fuzzy linguistic quantifier (e.g., most, almost all, etc.), and we will deal with such quantifiers only since they are obviously more important for modeling a fuzzy majority than the absolute quantifiers (e.g., about 5, much more than 10, etc.).

Property $F$ is defined as a fuzzy set in $Y$. For instance, if $Y = \{X, W, Z\}$ is the set of experts and $F$ is a property "convinced", then $F =$ "convinced" $= 0.1/X + 0.6/W + 0.8/Z$ which means that expert $X$ is convinced to degree 0.1, $W$ to degree 0.6 and $Z$ to degree 0.8. If now $Y = \{y_1, \ldots, y_p\}$, then it is assumed that truth($y_i$ is $F$) $= \mu_F(y_i)$, $i = 1, \ldots, p$.

Then, we follow the two steps:

$$r = \frac{1}{p} \sum_{i=1}^{p} \mu_F(y_i) \tag{4}$$

$$\text{truth}(Qy\text{'s are } F) = \mu_Q(r) \tag{5}$$

In the case of importance qualification, $B$ is defined as a fuzzy set in $Y$, and $\mu_B(y_i) \in [0, 1]$ is a degree of importance of $y_i$: from 1 for definitely important to 0 for definitely unimportant, through all intermediate values. We rewrite first "$QBy$'s are $F$" as "$Q(B$ and $F)y$'s are $B$" which leads to the following counterparts of (4) and (5):

$$r' = \frac{\sum_{i=1}^{p} [\mu_B(y_i) \wedge \mu_F(y_i)]}{\sum_{i=1}^{p} \mu_B(y_i)} \tag{6}$$

$$\text{truth}(QBY'\text{s are } F) = \mu_Q(r') \tag{7}$$

*Example 1.* Let $Y =$ "experts" $= \{X, Y, Z\}$, $F =$ "convinced" $= 0.1/X + 0.6/Y + 0.8/Z$, $Q =$ "most" be given by (3), $B =$ "important" $= 0.2/X + 0.5/Y + 0.6/Z$. Then: $r = 0.5$ and $r' = 0.92$, and truth("most experts are convinced")=0.4 and truth("most of the important experts are convinced")=1.

The method presented is simple and efficient, and has proven to be useful in a multitude of cases, also in this paper.

### 1.2.2 Ordered Weighted Averaging (OWA) Operators

Yager [84] (see also Yager and Kacprzyk's [85]) has proposed a special class of aggregation operators, called the *ordered weighted averaging* (or OWA, for short) operators, which seem to provide an even better and more general aggregation in the sense of being able to simply and uniformly model a large class of fuzzy linguistic quantifiers.

An OWA operator of dimension $p$ is a mapping $F : [0, 1]^p \rightarrow [0, 1]$ if associated with $F$ is a weighting vector $W = [w_1, \ldots, w_p]^T$ such that: $w_i \in [0, 1]$, $w_1 + \cdots + w_p = 1$, and

$$F(x_1, \ldots, x_p) = w_1 b_1 + \cdots + w_p b_p \tag{8}$$

where $b_i$ is the $i$-th largest element among $\{x_1, \ldots, x_p\}$. $B$ is called an ordered argument vector if each $b_i \in [0, 1]$, and $j > i$ implies $b_i \geq b_j$, $i = 1, \ldots, p$.

Then

$$F(x_1, \ldots, x_p) = W^T B \tag{9}$$

*Example 2.* Let $W^T = [0.2, 0.3, 0.1, 0.4]$, and calculate $F(0.6, 1.0, 0.3, 0.5)$. Thus, $B^T = [1.0, 0.6, 0.5, 0.3]$, and $F(0.6, 1.0, 0.3, 0.5) = W^T B = 0.55$; and $F(0.0, 0.7, 0.1, 0.2) = 0.21$.

For us it is relevant how the OWA weights are found from the membership function of a fuzzy linguistic quantifier $Q$; an early approach given in Yager [84] may be used here:

$$w_k = \mu_Q(k/p) - \mu_Q((k-1)/p) \quad \text{for} \quad k = 1, \ldots, p \tag{10}$$

Some examples of the $w_i$'s associated with the particular quantifiers are:

- If $w_p = 1$, and $w_i = 0$, for each $i \neq p$, then this corresponds to $Q =$ "all";
- If $w_i = 1$ for $i = 1$, and $w_i = 0$, for each $i \neq 1$, then this corresponds to $Q =$ "at least one",
- If $w_i = 1/p$, for each $i = 1, 2, \ldots, p$, then this corresponds to the arithmetic mean,

and the intermediate cases as, e.g., a half, most, much more than 75%, a few, almost all, etc. may be obtained by a suitable choice of the $w_i$'s between the above two extremes.

Thus, we will write

$$\text{truth}(Qy'\text{s are } F) = \text{OWA}_Q(\text{truth } (y_i \text{ is } F)) = W^T B \tag{11}$$

An important, yet difficult problem is the OWA operators with importance qualification. Suppose that we have a vector of data (pieces of evidence) $A = [a_1, \ldots, a_n]$, and a vector of importances $V = [v_1, \dots, v_n]$ such that $v_i \in [0, 1]$ is the importance of $a_i$, $i = 1, \dots, n$, $(v_1 + \dots + v_n \neq 1$, in general), and the OWA weights $W = [w_1, \dots, w_n]^T$ corresponding to $Q$ is determined via (10).

The case of an OWA operator with importance qualification, $\text{OWA}_V$, is unfortunately not trivial. In a popular Yager's [84] approach to be used here, the problem boils down to some redefinition of the OWA's weights $w_i$ into $\overline{w}_i$. Then, (8) becomes

$$F_V(a_1, \dots, a_n) = \overline{W}^T \cdot B = \sum_{j=1}^{n} \overline{w}_j b_j \tag{12}$$

We order first the pieces of evidence $a_i$, $i = 1, \dots, n$, in descending order to obtain $B$ such that $b_j$ is the $j$-th largest element of $\{a_1, \dots, a_n\}$. Next, we denote by $u_j$ the importance of $b_j$, i.e. of the $a_i$ which is the $j$-th largest; $i, j = 1, \dots, n$. Finally, the new weights $\overline{W}$ are defined as

$$\overline{w}_j = \mu_Q\left(\frac{\sum_{k=1}^{j} u_k}{\sum_{k=1}^{n} u_k}\right) - \mu_Q\left(\frac{\sum_{k=1}^{j-1} u_k}{\sum_{k=1}^{n} u_k}\right) \tag{13}$$

*Example 3.* Suppose that $A = [a_1, a_2, a_3, a_4]$ = [0.7, 1, 0.5, 0.6], and $V = [u_1, u_2, u_3, u_4]$ = [1, 0.6, 0.5, 0.9]. $Q$ = "most" is given by (3).

Then, $B = [b_1, b_2, b_3, b_4] = [1, 0.7, 0.6, 0.5]$, and $\overline{W} = [0.04, 0.24, 0.41, 0.31]$, and $F_I(A) = \sum_{j=1}^{4} \overline{w}_j b_j = 0.067 \cdot 1 + 0.4 \cdot 0.7 + 0.333 \cdot 0.6 + 0.2 \cdot 0.5 = 0.6468$.

We have now the necessary formal means to proceed to our discussion of group decision making and consensus formation models under fuzzy preferences and a fuzzy majority.

Finally, let us mention that OWA-like aggregation operators may be defined in an ordinal setting, i.e. for non-numeric data (which are only ordered), and we will refer the interested reader to, e.g., Delgado, Verdegay and Vila [17] or Herrera, Herrera-Viedma and Verdegay [28], and some other of their later papers.

## 2 Group Decision Making under Fuzzy Preferences and a Fuzzy Majority: General Remarks

*Group decision making* (equated here with social choice) proceeds in the following setting. We have a set of $n \geq 2$ alternatives, $S = \{s_1, \dots, s_n\}$, and a set of $m \geq 2$ individuals, $E = \{1, \dots, m\}$. Each individual $k \in E$ provides his or her testimony as to the alternatives in $S$, assumed to be individual fuzzy preference relations defined over $S$ (i.e. in $S \times S$). Fuzzy preference relations are employed to reflect an omnipresent fact that the preferences may be not clear-cut so that conventional non-fuzzy preference relations may be not adequate (see, e.g., many articles in Kacprzyk and Roubens [51] or Kacprzyk, Nurmi and Fedrizzi [27]).

An *individual fuzzy preference relation* of individual $k$, $R_k$, is given by its membership function $\mu_{R_k} : S \times S \longrightarrow [0, 1]$ such that

$$\mu_{R_k}(s_i, s_j) = \begin{cases} 1 & \text{if } s_i \text{ is definitely preferred to } s_j \\ c \in (0.5, 1) & \text{if } s_i \text{ is slightly preferred to } s_j \\ 0.5 & \text{in the case of indifference} \\ d \in (0, 0.5) & \text{if } s_j \text{ is slightly preferred to } s_i \\ 0 & \text{if } s_j \text{ is definitely preferred to } s_i \end{cases} \tag{14}$$

We will also use a special type of an individual fuzzy preference relation, a *fuzzy tournament*, but this will be explained later on.

If card $S$ is small enough (as assumed here), an individual fuzzy preference relation of individual $k$, $R_k$, may conveniently be represented by an $n \times n$ matrix $R_k = [r_{ij}^k]$, such that $r_{ij}^k = \mu_{R_k}(s_i, s_j)$; $i, j = 1, \ldots, n$; $k = 1, \ldots, m$. $R_k$ is commonly assumed (also here) to be reciprocal in that $r_{ij}^k + r_{ji}^k = 1$; moreover, it is also normally assumed that $r_{ii}^k = 0$, for all $i, k$; for a different, more justified convention, cf. García-Lapresta and Llamazares [16]. Notice that we do not mention here other properties of (individual) fuzzy preference relations which are often discussed (cf. Salles [37]) but which will not be relevant to our discussion. Moreover, we will not use here a more sophisticated concept of a fuzzy preference systems proposed by De Baets et al. which is presented in their paper in this volume. The reasoning is in this case principally the same.

Basically, two lines of reasoning may be followed here (cf. Kacprzyk [36, 37, 38, 39, 19, 41]):

- a direct approach: $\{R_1, \ldots, R_m\} \longrightarrow$ solution, that is, a solution is derived directly (without any intermediate steps) just from the set of individual fuzzy preference relations, and
- an indirect approach: $\{R_1, \ldots, R_m\} \longrightarrow R \longrightarrow$ solution, that is, from the set of individual fuzzy preference relations we form first a social fuzzy preference relation, $R$ (to be defined later), which is then used to find a solution.

A solution is here, unfortunately, not clearly understood – see, e.g., Nurmi [33, 69, 70, 71, 34] for diverse solution concepts. In this paper we will only sketch the derivation of some more popular solution concepts, and this will show to the reader not only the essence of the particular solution concept but how a fuzzification may be performed so that the reader can eventually fuzzify other crisp solution concepts that may be found in the literature. More specifically, we will show the derivation of some fuzzy cores and minimax sets for the direct approach, and some fuzzy consensus winners for the indirect approach. In addition to fuzzy preference relations, which are usually employed, we will also use a fuzzy majority represented by a linguistic quantifier as proposed by Kacprzyk [36, 37, 38, 39, 19, 41].

First, we will consider the case of fuzzy preferences only, and then we will add a fuzzy majority which is a more interesting case for our purposes.

# 3 Group Decision Making under Fuzzy Preferences

In this section we will only assume that we have individual fuzzy preferences and a non-fuzzy majority. We will present some solution concepts that are derived using the above mentioned direct and indirect approach, i.e. directly from individual fuzzy preference relations or via a social preference relation.

## *3.1 Solutions Based on Individual Fuzzy Preference Relations*

Let us first consider solution concepts that do not require any preference aggregation at all. One of the best solution concepts is that of a core or a set of undominated alternatives. Suppose that the nonfuzzy required majority be $r$ (e.g., at least 50%).

**Definition 1.** An alternative $x \in S$ belongs to the *core* iff there is no other alternative $y \in S$ that defeats $x$ by the required majority $r$.

We can extend the notion of a core to cover fuzzy individual preference relations by defining *the fuzzy $\alpha$-core* as follows (cf. Nurmi [33]):

**Definition 2.** An alternative $s_i \in S$ belongs to the *fuzzy $\alpha$-core $S_\alpha$* iff there exists no other alternative $s_j \in S$ such that $r_{ji} > \alpha$ for at least $r$ individuals.

It is easy to see that if the nonfuzzy core is nonempty, so is $S_\alpha$ for some $\alpha \in (0, 1]$. In other words, $\exists \alpha \in (0, 1]$: core $\subset S_\alpha$. Moreover, for any two values $\alpha_1, \alpha_2 \in (0, 1]$ such that $\alpha_1 < \alpha_2$, we have:

$$S_{\alpha_1} \subseteq S_{\alpha_2}$$

The intuitive interpretation of the fuzzy $\alpha$-core is obvious: an alternative is a member of $S_\alpha$ iff a sufficient majority of voters does not feel strongly enough against it.

Another nonfuzzy solution concept with much intuitive appeal is a minimax set. In a nonfuzzy setting it is defined as follows:

**Definition 3.** For each $x, y \in S$ denote the number of individuals preferring $x$ to $y$ by $n(x, y)$. Then define

$$v(x) = \max_y n(y, x)$$

and

$$n* = \min_x v(x)$$

Now the minimax set is

$$Q(n*) = \{x \mid v(x) = n*\}$$

Thus, $Q(n*)$ consists of those alternatives that in pairwise comparison with any other alternative are defeated by no more than $n*$ votes. Obviously, if $n* < m/2$ , where $m$ is the number of individuals, then $Q(n*)$ is singleton and $x \in Q(n*)$ is the core if the simple majority rule is being applied.

Analogously, we can define a *the minimax degree set* $Q(\beta)$ as follows. Given $s_i, s_j \in S$ and let, for individuals $k = 1, \ldots, m$:

$$v_D^k(x_j) = \max_i r_{ij}$$

We now define

$$v_D(x_j) = \max_k v_D^k(x_j)$$

Let $\min_j v_D(x_j) = \beta$. Then

$$Q(\beta) = \{x_j \mid v_D(x_j) = \beta\}$$

For properties of the minimax degree set, we refer to Nurmi [33, 69, 70].

Another concept that is analogous to the nonfuzzy minimax set is a *minimax opposition set*. Let $n_{ij}$ be the number of those individuals for whom $r_{ij} > r_{ji}$ and let $v_f(x_j) = \max_i n_{ij}$. Denote by $\bar{v}_f$ the minimum of $v_f(x_j)$ with respect to $j$, i.e.

$$\bar{v}_f = min_j v_f(x_j)$$

Then: $Q(v_f) = \{x_j \mid v_f(x_j) = \bar{v}_f\}$.

But, clearly, $Q(v_f) = Q(n*)$ since $r_{ij} > r_{ji}$ implies that the individual prefers alternative $x_i$ to $x_j$. Similarly, the preference of $x_i$ over $x_j$ implies that $r_{ij} > r_{ji}$. Consequently, the minimax opposition set does not take into account the intensity of preferences as expressed in the individual preference relation matrices.

A more general solution concept, the *α-minimax set* (cf. Nurmi [33]) denoted $Q^\alpha(v_f^\alpha)$, is defined as follows. Let $n_\alpha(x_i, x_j)$ be the number of individuals for whom $r_{ij} \leq \alpha$ for some value of $\alpha \in [0, 0.5)$. We now define $\forall x_i \in S : v_f^\alpha(x_i) = \max_j n_\alpha(x_i, x_j)$ and $\bar{v}_f^\alpha = \min_i v_f^\alpha(x_i)$. Then

$$Q^\alpha(v_f^\alpha) = \{x_i \mid v_f^\alpha(x_i) = \bar{v}_f^\alpha\}$$

It can be shown that $Q^\alpha(v_f^\alpha) \subseteq Q(n*)$ (see [33]).

### 3.1.1 Fuzzy Tournaments

One purpose of studying fuzzy tournaments is to overcome the difficulties inherent in the use of conventional solution concepts, namely the fact that the latter tend to produce too large solution sets and are therefore not decisive enough. Another purpose of our discussion is to apply analogues of the nonfuzzy solutions to contexts where the opinions of individuals can be represented by more general constructs than just connected and transitive preference relations (cf., e.g., [51]).

Let us take a look at a few solution concepts of nonfuzzy tournaments, mostly those proposed by Nurmi and Kacprzyk [35].

**Definition 4.** Given the alternative set $S$, a tournament $P$ on $S$ is a complete and asymmetric relation on $S$.

In the context of group decision making $P$ can be viewed as a strict preference relation. When $S$ is of small cardinality, $P$ can be expressed as a matrix $[p_{ij}]$, $p_{ij} \in \{0, 1\}$ so that $p_{ij} = 1$ if the alternative represented by row $i$ is preferred to that represented by column $j$, and $p_{ij} = 0$ if the alternative represented by column $j$ is preferred to that represented by row $i$.

Suppose that each individual has a complete, transitive and asymmetric preference relation over $S$, and that the number of individuals is odd. Then a tournament can be constructed through pairwise comparisons of alternatives. In the ensuing tournament alternative $s_i$ is preferred to $s_j$ iff the number of individuals preferring the former to the latter is larger than the number of individual preferring $s_j$ to $s_i$.

Perhaps the best-known solution concept of tournaments is the Condorcet winner.

**Definition 5.** The *Condorcet winner* is an alternative which is preferred to all other alternatives by a majority.

The main problem with this solution concept is that it does not always exist.

**Definition 6.** The *Copeland winning set* $UC_C$ consists of those alternatives that have the largest number of 1s in their corresponding rows in the tournament matrix.

In other words, the Copeland winners defeat more alternatives than any other alternatives do.

**Definition 7.** The uncovered set is defined by means of a binary relation of covering. An alternative $s_i$ *covers* another alternative $s_j$ iff $s_i$ defeats $s_j$ and everything that $s_j$ defeats. The *uncovered set* consists of those alternatives that are covered by no alternatives.

**Definition 8.** The *Banks set* is the set of end-points of Banks chains. Starting from any alternative $s_i$ the *Banks chain* is constructed as follows. First one looks for an alternative that defeats $s_i$. Suppose that such an alternative exists and is $s_j$ (if one does not exist, then of course $s_i$ is the Condorcet winner). Next one looks for another alternative that defeats both $s_i$ and $s_j$, etc. Eventually, no alternative can be found that would defeat all previous ones in the chain starting from $s_i$. The last alternative which defeats all previous ones is the end-point of the Banks chain starting from $s_i$. The Banks set is then the set of all those end points.

The following relationships hold between the above mentioned solutions (cf. [34]):

- all solutions converge to the Condorcet winner when one exists,
- the uncovered set includes the Copeland winning set and the Banks set,

- when $S$ contains less than 7 elements, the uncovered set and the Banks set coincide, and
- when the cardinality of $S$ exceeds 12, the Banks set and the Copeland winning set may be distinct; however, they both always belong to the uncovered set.

Given a group $E$ of $m$ individuals, a collective fuzzy tournament $F = [r_{ij}]$ can be obtained through pairwise comparisons of alternatives so that

$$r_{ij} = \frac{card\{k \in E \mid s_i P_k s_j\}}{m}$$

where $P_k$ is a nonfuzzy tournament representing the preferences of individual $k$.

Let us now define a *strong fuzzy covering relation* $C_S \subset S \times S$ as follows

$$\forall i, j, l \in \{1, \ldots, n\} : s_i C_S s_j \Leftrightarrow r_{il} \geq r_{jl} \quad \& \quad r_{ij} > r_{ji}$$

Clearly, the strong fuzzy covering relation implies the nonfuzzy covering relation, but not *vice versa*. The set of $C_S$-undominated alternatives is denoted by $UC_S$.

Let us first define:

**Definition 9.** The *weak fuzzy covering relation* $C_W \subset S \times S$ is defined as follows:

$$\begin{aligned} &\forall s_i, s_j \in S : \\ &\quad s_i C_W s_j \Leftrightarrow r_{ij} > r_{ji} \\ &\quad \& \quad card\{s_l \in S : r_{il} > r_{jl}\} \geq card\{s_p \in S : r_{jp} > r_{ip}\} \end{aligned}$$

Obviously, $s_i C_S s_j$ implies $s_i C_W s_j$, but not conversely. Thus, the set of $C_W$-undominated alternatives, $UC_W$, is always a subset of $UC_S$. Moreover, the Copeland winning set is always included in $UC_S$, but not necessarily in $UC_W$ (see [35]).

If one is looking for a solution that is a plausible subset of the uncovered set, then $UC_W$ is not appropriate since it is possible that $UC_C$ is not always a subset of the uncovered set, let alone the Banks set.

Another solution concept, the $\alpha$-uncovered set, is based on the individual fuzzy preference tournament matrices. One first defines the fuzzy domination relation $D$ and an $\alpha$-covering relation $C_\alpha \subseteq S \times S$ as follows.

**Definition 10.** $s_i D s_j$ iff at least 50% of the individuals prefer $s_i$ to $s_j$ to a degree of at least 0.5.

**Definition 11.** If $s_i C_\alpha s_j$, then $s_i D s_j$ and $s_i D_\alpha s_k$, for all $s_k \in S$ for which $s_j D_\alpha s_k$.

**Definition 12.** The $\alpha$-uncovered set consists of those alternatives that are not $\alpha$-covered by any other alternative.

An obvious candidate for a plausible solution concept for fuzzy tournaments is an $\alpha$-uncovered set with the smallest value of $\alpha$.

Other fuzzy solution concepts analogous to their nonfuzzy counterparts can be defined (see Nurmi and Kacprzyk [35]). For example, the *$\alpha$-Banks set* can be constructed by imposing the restriction that the majority of voters prefer the next alternative to the previous one in the Banks chain with intensity of at least $\alpha$.

### 3.2 Solutions Based on a Social Fuzzy Preference Relation

The derivation of these solution concepts requires first a derivation of a social fuzzy preference relation.

Bezdek, Spillman and Spillman [8, 9] discuss the problem of finding the set of undominated alternatives or other stable outcomes given a collective fuzzy preference ordering over the alternative set; see also Nurmi [33].

We now define a couple of solution concepts for voting games with fuzzy collective preference relation.

**Definition 13.** The set $S_\alpha$ of *α-consensus winners* is defined as: $s_i \in S_\alpha$ iff $\forall s_j \neq s_i : r_{ij} \geq \alpha$, with $0.5 < \alpha \leq 1$

Whenever $S_\alpha$ is nonempty, it is a singleton, but it does not always exist. Thus, it may be useful to find other solution concepts that specify a nonempty alternative sets even when $S_\alpha$ is empty. One possible candidate is a straightforward extension of Kramer's minimax set. We call it a set of *minimax consensus winners*, denote it by $S_M$ and define as follows.

**Definition 14.** Let $\bar{r}_j = \max_i r_{ij}$ and $\bar{r} = \min_j \max_i r_{ij}$. Then $s_i \in S_M$ (the set of minimax consensus winners) iff $\bar{r}_i = \bar{r}$.

Clearly $S_M$ is always nonempty, but not necessarily a singleton. As a solution set it has the same interpretation as Kramer's minimax set: it consists of those alternatives which, when confronted with their toughest competitors, fare best, i.e. win by the largest score (if $\bar{r} \leq 0.5$) or lose by the smallest one (if $\bar{r} > 0.5$).

These solution concepts are based on the social preference relation matrix. Other ones can be obtained in several ways. For instance, one may start from a preference profile over a set of alternatives and construct the $[r_{ij}]$ matrix as follows:

$$r_{ij} = \begin{cases} \frac{1}{m} \Sigma_{k=1}^{m} a_{ij}^{k} & \text{for } i \neq j \\ r_{ij} = 0 & \text{for } i = j \end{cases}$$

where $a_{ij}^k = 1$ if $s_i$ is strictly preferred to $s_j$ by voter $k$, and $a_{ij}^k = 0$ otherwise.

There is nothing "fuzzy" in the above solutions. As the method of constructing the social preference relation matrix suggests, the starting point can just be the ordinary preference profile as well.

## 4 Group Decision Making under Fuzzy Preferences and a Fuzzy Majority

In this section we will consider some solution concepts of group decision making but when we both have fuzzy preference relations and a fuzzy majority, We will also follow here the two directions, i.e. by using the individual fuzzy preference relations only (a direct approach), and by deriving first a social fuzzy preference relation, and using it to derive solutions (an indirect approach).

## 4.1 Direct Derivation of a Solution

We will first employ the direct approach, i.e. $\{R_1, \ldots, R_m\} \longrightarrow$ solution to derive two popular solution concepts: fuzzy cores and minimax sets.

### 4.1.1 Fuzzy Cores

The core is a very intuitively appealing and often used solution concept. Conventionally, the core is defined as a set of *undominated alternatives*, i.e. those not defeated in *pairwise comparisons* by a required majority (strict!) $r \leq m$, i.e.

$$C = \{s_j \in S : \neg\exists s_i \in S \text{ such that } r_{ij}^k > 0.5 \text{ for at least } r \text{ individuals}\} \tag{15}$$

The first attempt at a fuzzification of the core is due to Nurmi [33] who has extended it to the *fuzzy $\alpha$-core* defined as

$$C_\alpha = \{s_j \in S : \neg\exists s_i \in S \text{ such that } r_{ij}^k > \alpha \geq 0.5 \text{ for at least } r \text{ individuals}\} \tag{16}$$

that is, as a set of alternatives not sufficiently (at least to degree $\alpha$) defeated by the required (still strict!) majority $r \leq m$.

As we have already indicated, in many group decision making related situations is may be more adequate to assume that the required majority is imprecisely specified as, e.g., given by a fuzzy linguistic quantifier as, say, *most*. This concept of a fuzzy majority has been proposed by Kacprzyk [36, 37, 38, 39, 19, 41], and it has turned out that it can be quite useful and adequate.

To employ a fuzzy majority to extend (fuzzify) the core, we start by denoting

$$h_{ij}^k = \begin{cases} 1 \text{ if } r_{ij}^k < 0.5 \\ 0 \text{ otherwise} \end{cases} \tag{17}$$

where here and later on in this section, if not otherwise specified, $i, j = 1, \ldots, n$ and $k = 1, \ldots, m$.

Thus, $h_{ij}^k$ just reflects if alternative $s_j$ defeats (in pairwise comparison) alternative $s_i$ ($h_{ij}^k = 1$) or not ($h_{ij}^k = 0$).

Then, we calculate

$$h_j^k = \frac{1}{n-1} \sum_{i=1, i \neq j}^{n} h_{ij}^k \tag{18}$$

which is clearly the extent, from 0 to 1, to which individual $k$ is not against alternative $s_j$, where 0 standing for definitely against to 1 standing for definitely not against, through all intermediate values.

Next, we calculate

$$h_j = \frac{1}{m} \sum_{k=1}^{m} h_j^k \tag{19}$$

which expresses to what extent, from 0 to 1 as in the case of (18), *all* the individuals are not against alternative $s_j$.

And, finally, we calculate

$$v_Q^j = \mu_Q(h_j) \tag{20}$$

is to what extent, from 0 to 1 as before, $Q$ (say, most) individuals are not against alternative $s_j$.

The *fuzzy Q-core* is now defined (cf. Kacprzyk [36, 37, 38, 39, 19, 41]) as a fuzzy set

$$C_Q = v_Q^1/s_1 + \cdots + v_Q^n/s_n \tag{21}$$

i.e. as a fuzzy set of alternatives that are not defeated by $Q$ (say, most) individuals.

Notice that in the above basic definition of a fuzzy $Q$-core we do not take into consideration to what degrees those defeats of one alternative by another are. They can be accounted for in a couple of plausible ways.

First and most straightforward is the introduction of a threshold into the degree of defeat in (17), for instance by denoting

$$h_{ij}^k(\alpha) = \begin{cases} 1 \text{ if } r_{ij}^k < \alpha \leq 0.5 \\ 0 \text{ otherwise} \end{cases} \tag{22}$$

where, again, $i, j = 1, \ldots, n$ and $k = 1, \ldots, m$. Thus, $h_{ij}^k(\alpha)$ just reflects if alternative $s_j$ sufficiently (i.e. at least to degree $1 - \alpha$) defeats (in pairwise comparison) alternative $s_i$ or not.

We can also explicitly introduce the strength of defeat into (17). Namely, we can introduce a function exemplified by

$$\hat{h}_{ij}^k = \begin{cases} 2(0.5 - r_{ij}^k) & \text{if } r_{ij}^k < 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where, again, $i, j = 1, \ldots, n$ and $k = 1, \ldots, m$. Thus, $\hat{h}_{ij}^k$ just reflects how strongly (from 0 to 1) alternative $s_j$ defeats (in pairwise comparison) alternative $s_i$.

Then, by following the same steps (18) – (21), we can derive an $\alpha/Q$-fuzzy core and an $s/Q$-fuzzy core.

*Example 4.* Suppose that we have four individuals, $k = 1, 2, 3, 4$, whose individual fuzzy preference relations are:

$$R_1 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.3 & 0.7 & 0.1 \\ 2 & 0.7 & 0 & 0.6 & 0.6 \\ 3 & 0.3 & 0.4 & 0 & 0.2 \\ 4 & 0.9 & 0.4 & 0.8 & 0 \end{array} \qquad R_2 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.4 & 0.6 & 0.2 \\ 2 & 0.6 & 0 & 0.7 & 0.4 \\ 3 & 0.4 & 0.3 & 0 & 0.1 \\ 4 & 0.8 & 0.6 & 0.9 & 0 \end{array}$$

$$R_3 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.5 & 0.7 & 0.1 \\ 2 & 0.5 & 0 & 0.8 & 0.4 \\ 3 & 0.3 & 0.2 & 0 & 0.2 \\ 4 & 1 & 0.6 & 0.8 & 0 \end{array} \qquad R_4 = \begin{array}{r|cccc} & j=1 & 2 & 3 & 4 \\ \hline i=1 & 0 & 0.4 & 0.7 & 0.8 \\ 2 & 0.6 & 0 & 0.4 & 0.3 \\ 3 & 0.3 & 0.6 & 0 & 0.1 \\ 4 & 0.7 & 0.7 & 0.9 & 0 \end{array}$$

Suppose now that the fuzzy linguistic quantifier is $Q =$ "most" defined by (3). Then, say:

$$C_{\text{"most"}} \cong 0.06/s_1 + 0.56/s_2 + 1/s_4$$
$$C_{0.3/\text{"most"}} \cong 0.56/s_4$$
$$C_{s/\text{"most"}} \cong 0.36/s_4$$

to be meant as follows: in case of $C_{\text{"most"}}$ alternative $s_1$ belongs to to the fuzzy $Q$-core to the extent 0.06. $s_2$ to the extent 0.56, and $s_4$ to the extent 1, and analogously for the $C_{0.3/\text{"most"}}$ and $C_{s/\text{"most"}}$. Notice that though the results obtained for the particular cores are different, for obvious reasons, $s_4$ is clearly the best choice which is evident if we examine the given individual fuzzy preference relations.

Clearly, the fuzzy linguistic quantifier based aggregation of partial scores in the above definitions of the fuzzy $Q$-core, $\alpha/Q$-core and $s/Q$-core, may be replaced by an ordered weighted averaging (OWA) operator based aggregation given by (10) and (11). This was proposed by Fedrizzi, Kacprzyk and Nurmi [19], and then followed by some other authors. The results obtained by using the OWA operators are similar to those for the usual fuzzy linguistic quantifiers.

Finally, let us notice that the individuals and alternatives may be assigned variable importance (competence) and relevance, respectively, and then the OWA based aggregation with importance qualification may be used. This will not change however the essence of the fuzzy cores defined above, and will not be discussed here for lack of space.

### 4.1.2 Minimax Sets

Another intuitively justified solution concept may be the minimax (opposition) set which may be defined for our purposes as follows.

Let $w(s_i, s_j) \in \{1, 2, \ldots, m\}$ be the number of individuals who prefer alternative $s_j$ to alternative $s_i$, i.e. for whom $r^k_{ij} < 0.5$.

If now

$$v(s_i) = \max_{j=1,\ldots,n} w(s_i, s_j) \tag{24}$$

and

$$v^* = \min_{i=1,\ldots,n} v(s_i) \tag{25}$$

then the *minimax set* is defined as

$$M(v^*) = \{s_i \in S : v(s_i) = v^*\} \tag{26}$$

i.e. as a (nonfuzzy) set of alternatives which in pairwise comparisons with any other alternative are defeated by no more than $v^*$ individuals, hence by the least number of individuals.

Nurmi [33] extends the minimax set, similarly in spirit to his extension of the core (16), to the $\alpha$*-minimax set* as follows. Let $w_\alpha(s_i, s_j) \in \{1, 2, \ldots, m\}$ be the number of individuals who prefer alternative $s_j$ to alternative $s_i$ at least to degree $1 - \alpha$, i.e. for whom $r^k_{ij} < \alpha \leq 0.5$.

If now

$$v_\alpha(s_i) = \max_{j=1,\ldots,n} w_\alpha(s_i, s_j) \tag{27}$$

and

$$v^*_\alpha = \min_{i=1,\ldots,n} v_\alpha(s_i) \tag{28}$$

then the $\alpha$*-minimax set* is defined as

$$M_\alpha(v^*_\alpha) = \{s_i \in S : v_\alpha(s_i) = v^*_\alpha\} \tag{29}$$

i.e. as a (nonfuzzy) set of alternatives which in pairwise comparisons with any other alternative are defeated (at least to degree $1 - \alpha$) by no more than $v^*$ individuals, hence by the least number of individuals.

A fuzzy majority was introduced into the above definitions of minimax sets by Kacprzyk [36, 37, 38, 39, 19] as follows.

We start with (17), i.e.

$$h^k_{ij} = \begin{cases} 1 \text{ if } r^k_{ij} < 0.5 \\ 0 \text{ otherwise} \end{cases} \tag{30}$$

and

$$h_i^k = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} h_{ij}^k \tag{31}$$

is the extent, between 0 and 1, to which individual $k$ is against alternative $s_i$.

Then

$$h_i = \frac{1}{m} \sum_{k=1}^{m} h_i^k \tag{32}$$

is the extent, between 0 and 1, to which all the individuals are against alternative $s_i$.

Next

$$t_i^Q = \mu_Q(h_i) \tag{33}$$

is the extent, from 0 to 1, to which $Q$ (say, most) individuals are against alternative $s_i$, and

$$t_Q^* = \min_{i=1,\ldots,n} t_i^Q \tag{34}$$

is the least defeat of any alternative by $Q$ individuals.

Finally, the $Q$-minimax set is

$$M_Q(t_Q^*) = \{s_i \in S : t_i^Q = t_Q^*\} \tag{35}$$

And analogously as for the $\alpha/Q$-core and $s/Q$-core, we can explicitly introduce the degree of defeat $\alpha < 0.5$ and $s$ into the definition of the $Q$-minimax set.

*Example 5.* For the same four individual fuzzy preference relations $R_1, \ldots, R_4$ as in Example 4, we obtain for instance:

$$M_{\text{“most”}}(0) = \{s_4\}$$
$$M_{0.3/\text{“most”}}(0) = \{s_1, s_2, s_4\}$$
$$M_{s/\text{“most”}} = \{s_1, s_2, s_4\}$$

The OWA based aggregation can also be employed for the derivation of fuzzy minimax sets given above. And, again, the results obtained by using the OWA based aggregation are similar to those obtained by directly employing Zadeh's [86] calculus of linguistically quantified statements.

## 4.2 Indirect Derivation of a Solution – the Consensus Winner

Now we follow the scheme: $\{R_1, \ldots, R_m\} \longrightarrow R \longrightarrow$ solution i.e. from the individual fuzzy preference relations we determine first a social fuzzy preference relation, $R$, which is similar in spirit to its individual counterpart but concerns the whole group of individuals, and then find a solution from such a social fuzzy preference relation.

It is easy to notice that the above direct derivation scheme involves in fact two problems:

- how to find a social fuzzy preference relation from the individual fuzzy preference relations, i.e.

$$\{R_1, \ldots, R_m\} \longrightarrow R$$

- how to find a solution from the social fuzzy preference relation, i.e.

$$R \longrightarrow \text{solution}$$

In this paper we will not deal in more detail with the first step, i.e. $\{R_1, \ldots, R_m\} \longrightarrow R$, and assume a (most) straightforward alternative that the social fuzzy preference relation $R = [r_{ij}]$ is given by

$$r_{ij} = \begin{cases} \frac{1}{m} \sum_{k=1}^{m} a_{ij}^k & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

where

$$a_{ij}^k = \begin{cases} 1 \text{ if } r_{ij}^k > 0.5 \\ 0 \text{ otherwise} \end{cases} \tag{37}$$

Notice that $R$ obtained via (36) need not be reciprocal, i.e. $r_{ij} \neq 1 - r_{ji}$, but it can be shown that $r_{ij} \leq 1 - r_{ji}$, for each $i, j = 1, \ldots, n$.

We will discuss now the second step, i.e. $R \longrightarrow$ solution, i.e. how to determine a solution from a social fuzzy preference relation.

A solution concept of much intuitive appeal is here the consensus winner (cf. Nurmi [33]) which will be extended under a social fuzzy preference relation and a fuzzy majority.

We start with

$$g_{ij} = \begin{cases} 1 \text{ if } r_{ij} > 0.5 \\ 0 \text{ otherwise} \end{cases} \tag{38}$$

which expresses whether alternative $s_i$ defeats (in the whole group's opinion!) alternative $s_j$ or not.

Next

$$g_i = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} g_{ij} \tag{39}$$

which is a mean degree to which alternative $s_i$ is preferred, by the whole group, over all the other alternatives.

Then

$$z_Q^i = \mu_Q(g_i) \tag{40}$$

is the extent to which alternative $s_i$ is preferred, by the whole group, over $Q$ (e.g., most) other alternatives.

Finally, we define the *fuzzy Q-consensus winner* as

$$W_Q = z_Q^1/s_1 + \cdots + z_Q^n/s_n \tag{41}$$

i.e. as a fuzzy set of alternatives that are preferred, by the whole group, over $Q$ other alternatives.

And analogously as in the case of the core, we can introduce a threshold $\alpha \geq 0.5$ and $s$ into (38) and obtain a *fuzzy $\alpha/Q$-consensus winner* and a *fuzzy $s/Q$-consensus winner*, respectively.

*Example 6.* For the same individual fuzzy preference relations as in Example 4, and using (36) and (37), we obtain the following social fuzzy preference relation

$R =$

| | $j=1$ | 2 | 3 | 4 |
|---|---|---|---|---|
| $i=1$ | 0 | 0 | 1 | 0.25 |
| 2 | 0.75 | 0 | 0.75 | 0.25 |
| 3 | 0 | 0.25 | 0 | 0 |
| 4 | 1 | 0.75 | 1 | 0 |

If now the fuzzy majority is given by $Q =$ "most" defined by (3) and $\alpha = 0.8$, then we obtain

$$W_{\text{“most”}} = \tfrac{1}{15}/s_1 + \tfrac{11}{15}/s_2 + 1/s_4$$
$$W_{0.8/\text{“most”}} = \tfrac{1}{15}/s_1 + \tfrac{11}{15}/s_4$$
$$W_{s/\text{“most”}} = \tfrac{1}{15}/s_1 + \tfrac{1}{15}/s_2 + 1/s_4$$

which is to be read similarly as for the fuzzy cores in Example 4. Notice that here once again alternative $s_4$ is clearly the best choice which is obvious by examining the social fuzzy preference relation.

One can also use here an OWA based aggregation defined by (10) and (11) as proposed by proposed by Fedrizzi and Kacprzyk [19] and Kacprzyk and Fedrizzi [47].

This concludes our brief exposition of how to employ fuzzy linguistic quantifiers to model the fuzzy majority in group decision making. We did not present some other solution concepts as, e.g., minimax consensus winners (cf. Nurmi [33], Kacprzyk [38]) or those based on fuzzy tournaments which have been proposed by Nurmi and Kacprzyk [35] and are mentioned earlier in this paper.

We will finish this section with a remark that in a number of recent papers by Kacprzyk and Zadrożny [52, 53] it has been shown that the concept of Kacprzyk's [36, 37] fuzzy $Q$-core can be a general (prototypical) choice function in group decision making and voting, for instance those of: a "consensus solution", Borda's rule, the minimax degree set, the plurality voting, the qualified plurality voting, the approval voting-like, the "consensus + approval voting", Condorcet's rule, the Pareto rule, Copeland's rule, Nurmi's minimax set, Kacprzyk's $Q$-minimax, the Condorcet looser, the Pareto inferior alternatives, etc. This result, as interesting as it is, is however beyond the scope of this paper.

## 5 Degrees of Consensus under Fuzzy Preferences and a Fuzzy Majority

In this section fuzzy linguistic quantifiers as representations of a fuzzy majority will be employed to define a degree of consensus as proposed in Kacprzyk [19]), and then advanced in Kacprzyk and Fedrizzi [21, 22], and Kacprzyk, Fedrizzi and Nurmi [46, 47], see also Kacprzyk, Nurmi and Fedrizzi [27, 50] and Zadrożny [65]. This degree is meant to overcome some "rigidness" of the conventional concept of consensus in which (full) consensus occurs only when "all the individuals agree as to all the issues". This may often be counterintuitive, and not consistent with a real human perception of the very essence of consensus (see, e.g., the citation from a biological context given in the beginning of the paper).

The new degree of consensus proposed can be therefore equal to 1, which stands for full consensus, when, say, "most of the individuals agree as to almost all (of the relevant) issues (alternatives, options)".

Our point of departure is again a set of individual fuzzy preference relations which are meant analogously as in Sect. 2 [see, e.g., (17)].

The degree of consensus is now derived in three steps:

- first, for each pair of individuals we derive a degree of agreement as to their preferences between *all* the pairs of alternatives,
- second, we aggregate these degrees to obtain a degree of agreement of each pair of individuals as to their preferences between $Q_1$ (a linguistic quantifier as, e.g., "most", "almost all", "much more than 50%", ...) pairs of relevant alternatives, and
- third, we aggregate these degrees to obtain a degree of agreement of $Q_2$ (a linguistic quantifier similar to $Q_1$) pairs of important individuals as to their preferences between $Q_1$ pairs of relevant alternatives, and this is meant to be the *degree of consensus* sought.

Notice that we assume here, as opposed to Sect. 2, that both the individuals and alternatives are assigned different degrees of importance and relevance. However, this may be useful in the context of consensus reaching, and a basic case with the same importance and relevance for all the individuals and alternatives will just be a special case of the one adopted in this paper.

The above derivation process of a degree of consensus may be formalized by using Zadeh's [86] calculus of linguistically quantified statements and Yager's [84] OWA based aggregation.

We start with the degree of strict agreement between individuals $k_1$ and $k_2$ as to their preferences between alternatives $s_i$ and $s_j$

$$v_{ij}(k_1, k_2) = \begin{cases} 1 \text{ if } r_{ij}^{k_1} = r_{ij}^{k_2} \\ 0 \text{ otherwise} \end{cases} \tag{42}$$

where here and later on in this section, if not otherwise specified, $k_1 = 1, \ldots, m-1$; $k_2 = k_1 + 1, \ldots, m$; $i = 1, \ldots, n-1$; $j = i + 1, \ldots, n$.

The relevance of alternatives is assumed to be given as a fuzzy set defined in the set of alternatives $S$ such that $\mu_B(s_i) \in [0, 1]$ is a *degree of relevance* of alternative $s_i$, from 0 for fully irrelevant to 1 for fully relevant, through all intermediate values.

The relevance of a pair of alternatives, $(s_i, s_j) \in S \times S$, may be defined, say, as

$$b_{ij}^B = \frac{1}{2}[\mu_B(s_i) + \mu_B(s_j)] \tag{43}$$

which is clearly the most straightforward option; evidently, $b_{ij}^B = b_{ji}^B$, and $b_{ii}^B$ do not matter; for each $i, j$.

And analogously, the *importance of individuals*, $I$, is defined as a fuzzy set in the set of individuals such that $\mu_I(k) \in [0, 1]$ is a *degree of importance* of individual $k$, from 0 for fully unimportant to 1 for fully important, through all intermediate values.

Then, the importance of a pair of individuals, $(k_1, k_2)$, $b_{k_1,k_2}^I$, may be defined in various ways, e.g., analogously as (19), i.e.

$$b_{k_1,k_2}^I = \frac{1}{2}[\mu_I(k_1) + \mu_I(k_2)] \tag{44}$$

The degree of agreement between individuals $k_1$ and $k_2$ as to their preferences between *all* the relevant pairs of alternatives is [cf. (6)]

$$v_B(k_1, k_2) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [v_{ij}(k_1, k_2) \wedge b_{ij}^B]}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{ij}^B} \tag{45}$$

The degree of agreement between individuals $k_1$ and $k_2$ as to their preferences between $Q_1$ relevant pairs of alternatives is

$$v_{Q_1}^B(k_1,k_2) = \mu_{Q_1}[v_B(k_1,k_2)] \tag{46}$$

In turn, the degree of agreement of *all* the pairs of important individuals as to their preferences between $Q_1$ pairs of relevant alternatives is

$$v_{Q_1}^{I,B} = \frac{2}{m(m-1)} \frac{\sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} [v_{Q_1}^B(k_1,k_2) \wedge b_{k_1,k_2}^I]}{\sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} b_{k_1,k_2}^I} \tag{47}$$

and, finally, the degree of agreement of $Q_2$ pairs of important individuals as to their preferences between $Q_1$ pairs of relevant alternatives, called the *degree of* $Q1/Q2/I/B$*-consensus*, is

$$con(Q_1, Q_2, I, B) = \mu_{Q_2}(v_{Q_1}^{I,B}) \tag{48}$$

Since the strict agreement (42) may be viewed too rigid, we can use the degree of sufficient agreement (at least to degree $\alpha \in (0, 1]$) of individuals $k_1$ and $k_2$ as to their preferences between alternatives $s_i$ and $s_j$, defined by

$$v_{ij}^{\alpha}(k_1,k_2) = \begin{cases} 1 \text{ if } \mid r_{ij}^{k_1} - r_{ij}^{k_2} \mid \leq 1-\alpha \leq 1 \\ 0 \text{ otherwise} \end{cases} \tag{49}$$

where, $k_1 = 1, \ldots, m-1; k_2 = k_1+1, \ldots, m; i = 1, \ldots, n-1; j = i+1, \ldots, n.$

The degree of sufficient (at least to degree $\alpha$) agreement between individuals $k_1$ and $k_2$ as to their preferences between all the relevant pairs of alternatives is

$$v_B^{\alpha}(k_1,k_2) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [v_{ij}^{\alpha}(k_1,k_2) \wedge b_{ij}^B]}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{ij}^B} \tag{50}$$

The degree of sufficient (at least to degree $\alpha$) agreement between the individuals $k_1$ and $k_2$ as to their preferences between $Q_1$ relevant pairs of alternatives is

$$v_{Q_1}^{B,\alpha}(k_1,k_2) = \mu_{Q_1}[v_B^{\alpha}(k_1,k_2)] \tag{51}$$

In turn, the degree of sufficient (at least to degree $\alpha$) agreement of *all* the pairs of important individuals as to their preferences between $Q_1$ relevant pairs of alternatives is

$$v_{Q_1}^{I,B,\alpha} = \frac{2}{m(m-1)} \frac{\sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} [v_{Q_1}^{B\alpha}(k_1,k_2) \wedge b_{k_1,k_2}^I]}{\sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} b_{k_1,k_2}^I} \tag{52}$$

and, finally, the degree of sufficient (at least to degree $\alpha$) agreement of $Q_2$ pairs of important individuals as to their preferences between $Q_1$ relevant pairs of alternatives, called the *degree of* $\alpha/Q1/Q2/I/B$*-consensus*, is

$$con^{\alpha}(Q_1, Q_2, I, B) = \mu_{Q_2}(v_{Q_1}^{I,B,\alpha}) \tag{53}$$

We can also explicitly introduce the strength of agreement into (42), and analogously define the degree of strong agreement of individuals $k_1$ and $k_2$ as to their preferences between alternatives $s_i$ and $s_j$, e.g., as

$$v_{ij}^{s}(k1, k2) = s(|\, r_{ij}^{k_1} - r_{ij}^{k_2} \,|) \tag{54}$$

where $s : [0, 1] \longrightarrow [0, 1]$ is some function representing the degree of strong agreements as, e.g.,

$$s(x) = \begin{cases} 1 & \text{for } x \leq 0.05 \\ -10x + 1.5 & \text{for } 0.05 < x < 0.15 \\ 0 & \text{for } x \geq 0.15 \end{cases} \tag{55}$$

such that $x' < x'' \Longrightarrow s(x') \geq s(x'')$, for each $x', x'' \in [0, 1]$, and there is such an $x \in [0, 1]$ that $s(x) = 1$.

The degree of strong agreement between individuals $k_1$ and $k_2$ as to their preferences between *all* the pairs of alternatives is then

$$v_{B}^{s}(k_1, k_2) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} [v_{ij}^{s}(k_1, k_2) \wedge b_{ij}^{B}]}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{ij}^{B}} \tag{56}$$

The degree of strong agreement between individuals $k_1$ and $k_2$ as to their preferences between $Q_1$ relevant pairs of alternatives is

$$v_{Q_1}^{B,s}(k_1, k_2) = \mu_{Q_1}[v_{B}^{s}(k_1, k_2)] \tag{57}$$

In turn, the degree of strong agreement of *all* the pairs of important individuals as to their preferences between $Q_1$ relevant pairs of alternatives is

$$v_{Q_1}^{I,B,s} = \frac{2}{m(m-1)} \frac{\sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} [v_{Q_1}^{B,s}(k_1, k_2) \wedge b_{k_1,k_2}^{I}]}{\sum_{k_1=1}^{m-1} \sum_{k_2=k_1+1}^{m} b_{k_1,k_2}^{I}} \tag{58}$$

and, finally, the degree of agreement of $Q_2$ pairs of important individuals as to their preferences between $Q_1$ relevant pairs of alternatives, called the *degree of $s/Q1/Q2/I/B$-consensus*, is

$$con^{s}(Q_1, Q_2, I, B) = \mu_{Q_2}(v_{Q_1}^{I,B,s}) \tag{59}$$

*Example 7.* Suppose that $n = m = 3$, $Q_1 = Q_2 =$ "most" are given by (3), $\alpha = 0.9$, $s(x)$ is defined by (55), and the individual preference relations are:

$$R_1 = [r^1_{ij}] = \begin{array}{r|ccc} & j=1 & 2 & 3 \\ \hline i=1 & 0 & 0.1 & 0.6 \\ 2 & 0.9 & 0 & 0.7 \\ 3 & 0.4 & 0.3 & 0 \end{array} \qquad R_2 = [r^2_{ij}] = \begin{array}{r|ccc} & j=1 & 2 & 3 \\ \hline i=1 & 0 & 0.1 & 0.7 \\ 2 & 0.9 & 0 & 0.7 \\ 3 & 0.3 & 0.3 & 0 \end{array}$$

$$R_3 = [r^3_{ij}] = \begin{array}{r|ccc} & j=1 & 2 & 3 \\ \hline i=1 & 0 & 0.2 & 0.6 \\ 2 & 0.8 & 0 & 0.7 \\ 3 & 0.4 & 0.3 & 0 \end{array}$$

If we assume the relevance of the alternatives to be $B = \{b^B_i/s_i\} = 1/s_1 + 0.6/s_2 + 0.2/s_3$, the importance of the individuals to be $I = \{b^I_k/k\} = 0.8/1 + 1/2 + 0.4/3$, $\alpha = 0.9$ and $Q =$ "most$''$ given by (3), then we obtain the following degrees of consensus:

$$con(\text{"most", "most"}, I, B) \cong 0.35$$
$$con^{0.9}(\text{"most", "most"}, I, B) \cong 0.06$$
$$con^{s}(\text{"most", "most"}, I, B) \cong 0.06$$

And, similarly as for the group decision making solutions shown in Sect. 2, the aggregation via Zadeh's [86] calculus of linguistically quantified propositions employed above may be replaced by the OWA based aggregation given by (10) and (11). The procedure is analogous as that presented in Sect. 2, and will not be repeated here.

For more information on these degrees of consensus, see, e.g., works by Kacprzyk, Fedrizzi, Nurmi and Zadrożny [19, 19, 41, 21, 22, 23, 24, 46], etc.

## 6 Remarks on some Voting Paradoxes and their Alleviation

Voting paradoxes are an interesting and very relevant topic that has a considerable theoretical and practical relevance. In this paper we will just give some simple examples of well known paradoxes and indicate some possibilities of how to alleviate them by using some elements of fuzzy preferences and a fuzzy majority. The paper is based on the works by Nurmi [74, 75], and Nurmi and Kacprzyk [77].

Table 1 presents an instance of Condorcet's paradox where there are three voter groups of equal size having preferences over alternatives $A$, $B$ and $C$ as indicated

**Table 1** Condorcet's paradox

| *Group I* | *Group II* | *Group III* |
|---|---|---|
| *A* | *B* | *C* |
| *B* | *C* | *A* |
| *C* | *A* | *B* |

by the rank order indicated below each group. In fact, the groups need not be of equal size. What is essential for the paradox is that any two of them constitutes a majority. Clearly, a collective preference relation formed on the basis of comparing alternatives in pairs and using majority rule, results in a cycle: $A$ is preferred to $B$, $B$ is preferred to $C$ and $C$ is preferred to $A$.

An instance of Borda's paradox, in turn, is given in Table 2, where alternative $A$ would win by a plurality of votes and, yet, both $B$ and $C$ would beat $A$, should pairwise majority comparisons be conducted.

A common feature in these classic paradoxes is an incompatibility of several intuitively plausible requirements regarding social choices. In the case of Condorcet's paradox the result obtained by using majority rule on a set of complete and transitive preferences is intransitive. In the case of Borda's paradox, the winner in the plurality sense is different from the winner in another sense, i.e. in the sense that requires the winner to beat all the other alternatives in binary contests.

Let us try to solve the above paradoxes using some fuzzy tools. The solutions presented are very much in the spirit of Sen's idea of broadening the amount of information about individuals. In particular, we shall take our point of departure in the notion of fuzzy individual preference relation. We consider the set $E$ of individuals and the set $S$ of decision alternatives. Each individual $i \in E$ is assumed to possess a fuzzy preference relation $R_i(x, y)$ over $S$. For each $x, y \in S$ the value $R_i(x, y)$ indicates the degree in which $x$ is preferred to $y$ by $i$ with 1 indicating the strongest preference of $x$ to $y$, 0.5 indifference between the two and value 0 the strongest preference of $y$ to $x$. Obviously, the assumption that the voters be endowed with fuzzy preference relations is precisely the kind of broadening of the information about individuals that Sen discusses. Some properties of fuzzy preference relations are defined in the following [15, 81].

**Definition 15.** *Connectedness. A fuzzy preference relation R is connected if an only if* $R(x, y) + R(y, x) \geq 1, \forall x, y \in S$.

**Definition 16.** *Reflexivity. A fuzzy preference relation R is reflexive if an only if* $R(x, x) = 1, \forall x \in S$.

**Definition 17.** *Max-min transitivity. A fuzzy connected and reflexive relation R is max-min transitive if and only if* $R(x, z) \geq min[R(x, y), R(y, z)], \forall x, y, z \in S$.

For the case of the Condorcet paradox, given the broadening of information concerning voter preferences represented by fuzzy preference relations, we can solve it very much in the spirit of its "father", Marquis de Condorcet (cf. Nurmi [75]). A way out of cyclical collective preferences is to look at the sizes of majorities supporting various collective preferences. For example, if the number of voters preferring $a$ to $b$ is 5 out of 9, while that of voters preferring $b$ to $c$ is 7 out of 9, then, according

**Table 2** Borda's paradox

| *voters 1–4* | *voters 5–7* | *voters 8,9* |
|---|---|---|
| *A* | *B* | *C* |
| *B* | *C* | *B* |
| *C* | *A* | *A* |

to Condorcet, the latter preference is stronger than the former. By cutting the cycle of collective majority preferences at its weakest link, one ends up with a complete and transitive relation. Clearly, with nonfuzzy preference relation this method works only in cases where not all of the majorities supporting various links in the cycle are of same size. With fuzzy preferences one can form the collective preference between any $x$ and $y \in S$ using a variation of the average rule (cf. Intilligator [34]), i.e.

$$R(x, y) = \frac{\sum_i R_i(x, y)}{m} \tag{60}$$

where $R(x, y)$ is the degree of collective fuzzy preference of $x$ over $y$.

Now, supposing that a preference cycle is formed on the basis of collective fuzzy preferences, one could simply ignore the link with weakest degree of preference and thus possibly end up with a ranking. In general one can proceed by eliminating weakest links in collective preference cycles until a ranking results.

The above method of successive elimination of weakest links in preference cycles thus works with fuzzy and nonfuzzy preferences. When individual preferences are fuzzy each voter is assumed to report his/her preferences so that the following matrix can be formed:

$$R_i = \begin{pmatrix} - & r_{12} & \dots & r_{1n} \\ r_{21} & - & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & - \end{pmatrix} \tag{61}$$

Here $r_{ij}$ indicates the degree in which $i$ prefers the $i$-th alternative to the $j$-th one. By averaging over the voters we obtain:

$$\bar{R} = \begin{pmatrix} - & \bar{r}_{12} & \dots & \bar{r}_{1n} \\ \bar{r}_{21} & - & \dots & \bar{r}_{2n} \\ \dots & \dots & \dots & \dots \\ \bar{r}_{n1} & \bar{r}_{n2} & \dots & - \end{pmatrix} \tag{62}$$

Apart from the successive elimination method one can use another straightforward method to resolve Condorcet's paradox, once the $\bar{R}$-matrix is given. It proceeds as follows. One first computes the row sums of the matrix:

$$\bar{r}_i = \sum_j \bar{r}_{ij} \tag{63}$$

These represent the total fuzzy preference weight assigned to the $i$-th alternative in all pairwise preference comparisons, when the weight in each comparison is the average fuzzy preference value. Let now

**Table 3** Fuzzy Condorcet's paradox

| *voter 1* | | | | *voter 2* | | | | *voter 3* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | | *A* | *B* | *C* | | *A* | *B* | *C* |
| *A* | - | .6 | .8 | *A* | - | .9 | .3 | *A* | - | .6 | .3 |
| *B* | .4 | - | .6 | *B* | .1 | - | .7 | *B* | .4 | - | .1 |
| *C* | .2 | .4 | - | *C* | .7 | .3 | - | *C* | .7 | .9 | - |

$$p_i = \frac{\bar{r}_i}{\sum_i \bar{r}_i}. \tag{64}$$

Clearly $p_i \geq 0$ and $\sum_i p_i = 1$. Thus, $p_i$ has the natural interpretation of choice probability. An obvious way to utilize this is to form the collective preference ordering on the basis of these choice probabilities. The result is necessarily a complete and transitive relation. Hence we can use the information broadening provided by fuzzy preferences to solve Condorcet's paradox (cf. Table 3).

For illustration, consider the example of Table 1 again and assume that each group consists of just one voter. Assume, furthermore, that the fuzzy preferences underlying the preference rankings are as follows:

The $\bar{R}$- matrix is now:

$$\bar{R} = \begin{pmatrix} - & .7 & .5 \\ .3 & - & .5 \\ .5 & .5 & - \end{pmatrix}$$

Now, $P_A = 0.4$, $P_B = 0.3$, $P_C = 0.3$.

Obviously, the solution is based on somewhat different fuzzy preference relations over the three alternatives. Should the preference relations be identical, we would necessarily end up with identical choice probabilities.

With fuzzy individual preference relations we can resolve Borda's paradox. To do that, we simply apply the same procedure as in the resolution of Condorcet's paradox.

Let us take a look at a fuzzy Borda's paradox for illustration. Assume that the fuzzy preferences underlying Table 2 are those indicated in Table 4.

**Table 4** A fuzzy Borda's paradox

| *4 voters* | | | | *3 voters* | | | | *2 voters* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *A* | *B* | *C* | | *A* | *B* | *C* | | *A* | *B* | *C* |
| *A* | - | .6 | .8 | *A* | - | .9 | .3 | *A* | - | .2 | .1 |
| *B* | .4 | - | .6 | *B* | .1 | - | .7 | *B* | .8 | - | .3 |
| *C* | .2 | .4 | - | *C* | .7 | .3 | - | *C* | .9 | .7 | - |

The matrix of average preference degrees is then the following:

$$\bar{R} = \begin{pmatrix} - & .6 & .5 \\ .4 & - & .6 \\ .5 & .4 & - \end{pmatrix}$$

The choice probabilities of $A$, $B$ and $C$ are, thus, $0.37, 0.33, 0.30$. We see that the choice probability of $A$ is the largest. In a sense, then, the method does not solve Borda's paradox in the same way as the Borda count does since also plurality method ends up with $A$ being chosen instead of the Condorcet winner alternative $B$. Note, however, that fuzzy preference relations give a richer picture of voter preferences than the ordinary preference rankings. In particular, $A$ is strongly preferred to $B$ and $C$ by both the 4 and 3 voter groups. Hence, it is to be expected that its choice probability is the largest.

For additional information on voting paradoxes and some ways to solve them using fuzzy logic, we refer the reader to Nurmi and Kacprzyk [77].

## 7 Concluding Remarks

In this paper we have briefly presented the use of fuzzy preference relations and fuzzy majorities in the derivation of group decision making (social choice) solution concepts and degrees of consensus. First, we briefly discussed some more general issues related to the role fuzzy preference relations and a fuzzy majority may play as a tool to alleviate difficulties related to negative results in group decision making exemplified by Arrow's impossibility theorem. Though very important for a conceptual point of view, these analyses are of a lesser practical relevance to the user who wishes to employ those fuzzy tools to constructively solve the problems considered.

Therefore, emphasis has been on the use of fuzzy preference relations and fuzzy majorities to derive more realistic and human-consistent solutions of group decision making. Reference has been given to other approaches and works in this area, as well as to the authors' previous, more foundational works in which an analysis of basic issues underlying group decision making and consensus formation has been included.

It is hoped that this work will provide the interested reader with some tools to constructively solve group decision making and consensus formation problems when both preferences and majorities are imprecisely specified or perceived, and may be modeled by fuzzy relations and fuzzy sets.

## References

1. Aizerman, M.A. (1985). New problems in the general choice theory, *Social Choice and Welfare*, 2, 235–282.
2. Arrow, K.J. (1963). *Social Choice and Individual Values*. 2nd Edition. Wiley, New York.

3. Barrett, C.R., Pattanaik, P.K. and Salles, M. (1986). On the structure of fuzzy social welfare functions. *Fuzzy Sets and Systems*, 19, 1–10.
4. Barrett, C.R., Pattanaik, P.K. and Salles, M. (1990). On choosing rationally when preferences are fuzzy. *Fuzzy Sets and Systems*, 34, 197–212.
5. Barrett, C.R., Pattanaik, P.K. and Salles, M. (1992). Rationality and aggregation of preferences in an ordinally fuzzy framework. *Fuzzy Sets and Systems* 49, 9–13.
6. Basu,K, Deb, R. and Pattanaik, P.K. (1992) Soft sets: An ordinal formulation of vagueness with some applications to the theory of choice. *Fuzzy Sets and Systems* 45, 45–58.
7. Bezdek, J.C., Spillman, B. and Spillman, R. (1978). A fuzzy relation space for group decision theory, *Fuzzy Sets and Systems*, 1, 255–268.
8. Bezdek, J.C., Spillman, B. and Spillman, R. (1979). Fuzzy relation space for group decision theory: An application, *Fuzzy Sets and Systems*, 2, 5–14.
9. Blin, J.M. (1974). Fuzzy relations in group decision theory, *J. of Cybernetics*, 4, 17–22.
10. Blin, J.M. and Whinston, A.P. (1973). Fuzzy sets and social choice, *J. of Cybernetics*, 4, 17–22.
11. Bordogna, G., Fedrizzi, M. and Pasi, G. (1997) A linguistic modelling of consensus in group decision making based on OWA operators, *IEEE Trans. on Systems, Man and Cybernetics*, SMC-27, 126–132.
12. Chiclana, F, Herrera, F. and Herrera-Viedma, E. (2001) Integrating multiplicative preference relations in a multipurpose decision making model based on fuzzy preference relations. *Fuzzy Sets and Systems*, 122, 277–291.
13. Chiclana, F, Herrera, F. and Herrera-Viedma, E. (2001a) Multiperson decision making based on multiplicative preference relations. *European Journal of Operational Research*, 129, 372–385.
14. Cutello, V. and Montero, J. (1993) A characterization of rational amalgamation operations, *International J. of Approximate Reasoning*, 8, 325–344.
15. Dasgupta, M. and Deb, R. (1996), Transitivity and fuzzy preferences, *Social Choice and Welfare*, 13, 305–318.
16. DeGrazia, A. (1953), Mathematical Derivation of an Election System, *Isis*, 44, 42–51.
17. Delgado, M., Verdegay, J.L. and Vila, M.A. (1993). On aggregation operations of linguistic labels, *Int. J. of Intelligent Systems*, 8, 351–370.
18. Delgado,M., Herrera, F., Herrera-Viedma, E. and Martınez, L. (1998) Combining numerical and linguistic information in group decision making. *Information Sciences*, 107, 177–194.
19. Fedrizzi, M., Kacprzyk, J. and Nurmi, H. (1993). Consensus degrees under fuzzy majorities and fuzzy preferences using OWA (ordered weighted average) operators, *Control and Cybernetics*, 22, 71–80.
20. Fedrizzi, M., Kacprzyk, J. and Nurmi, H. (1996). How different are social choice functions: a rough sets approach, *Quality and Quantity*, 30, 87–99.
21. Fedrizzi, M., Kacprzyk, J. and Zadrożny, S. (1988). An interactive multi-user decision support system for consensus reaching processes using fuzzy logic with linguistic quantifiers, *Decision Support Systems*, 4, 313–327.
22. Fishburn, P.C. (1990). Multiperson decision making: a selective review. In J. Kacprzyk and M. Fedrizzi (Eds.): *Multiperson Decision Making Models using Fuzzy Sets and Possibility Theory*, Kluwer, Dordrecht, pp. 3–27.
23. Fodor, J. and Roubens, M. (1994) *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, Dordrecht.
24. García-Lapresta, J.L. and Llamazares, B. (2000). Aggregation of fuzzy preferences: Some rules of the mean. *Social Choice and Welfare*, 17, 673–690.
25. González-Pachón, J., Gómez, D., Montero, J. and Yáñez, J. (2003) Searching for the dimension of valued preference relations, *International Journal of Approximate Reasoning*, 33, 133–157.
26. Gonzáalez-Pachón, J., Gómez, D., Montero, J. and Yáñez, J. (2003a) Soft dimension theory, *Fuzzy Sets and Systems*, 137, 137–149.
27. Herrera, F., and Herrera-Viedma, E. (2000) Choice functions and mechanisms for linguistic preference relations. *European Journal of Operational Research*, 120, 144–161.

28. Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1996). A model of consensus in group decision making under linguistic assessments, *Fuzzy Sets and Systems*, 78, 73–88.
29. Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1998) Choice processes for non-homogeneous group decision making in linguistic setting. *Fuzzy Sets and Systems*, 94, 297–308.
30. Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1997) Linguistic measures based on fuzzy coincidence for reaching consensus in group decision making. *International Journal of Approximate Reasoning*, 16, 309–334.
31. Herrera, F., Herrera-Viedma, E. and Verdegay, J.L. (1997a) A rational consensus model in group decision making using linguistic assessments. *Fuzzy Sets and Systems*, 88, 31–49.
32. Herrera, F., Martnez, L. (2000) An approach for combining numerical and linguistic information based on the 2-tuple fuzzy linguistic representation model in decision making. *International J. of Uncertainty , Fuzziness and Knowledge-Based Systems* 8, 539–562.
33. Herrera, F. and Verdegay, J.L. (1995). On group decision making under linguistic preferences and fuzzy linguistic quantifiers. In B. Bouchon-Meunier, R.R. Yager and L.A. Zadeh (Eds.): *Fuzzy Logic and Soft Computing*, World Scientific, Singapore, pp. 173–180.
34. Intrilligator, M.D. (1973). A probabilistic model of social choice, *Review of Economic Studies*, 40, 553–560.
35. Intrilligator, M.D. (1982). Probabilistic models of choice, *Mathematical Social Sciences*, 2, 157–166.
36. Kacprzyk, J. (1984). Collective decision making with a fuzzy majority rule, *Proc. of WOGSC Congress*, AFCET, Paris, pp. 153–159.
37. Kacprzyk, J. (1985). Zadeh's commonsense knowledge and its use in multicriteria, multistage and multiperson decision making. In M.M. Gupta et al. (Eds.): *Approximate Reasoning in Expert Systems*, North–Holland, Amsterdam, pp. 105–121.
38. Kacprzyk, J. (1985). Group decision-making with a fuzzy majority via linguistic quantifiers. Part I: A consensory-like pooling; Part II: A competitive-like pooling, *Cybernetics and Systems: an International J.*, 16, 119–129 (Part I), 131–144 (Part II).
39. Kacprzyk, J. (1986). Group decision making with a fuzzy linguistic majority, *Fuzzy Sets and Systems*, 18, 105–118.
40. Kacprzyk, J. (1987). On some fuzzy cores and "soft" consensus measures in group decision making. In J.C. Bezdek (Ed.): *The Analysis of Fuzzy Information*, Vol. 2, CRC Press, Boca Raton, pp. 119–130.
41. Kacprzyk, J. (1987). Towards 'human consistent' decision support systems through commonsense-knowledge-based decision making and control models: a fuzzy logic approach, *Computers and Artificial Intelligence*, 6, 97–122.
42. Kacprzyk, J. and Fedrizzi, M. (1986). "Soft" consensus measures for monitoring real consensus reaching processes under fuzzy preferences, *Control and Cybernetics*, 15, 309–323.
43. Kacprzyk, J. and Fedrizzi, M. (1988). A "soft" measure of consensus in the setting of partial (fuzzy) preferences, *Europ. J. of Operational Research*, 34, 315–325.
44. Kacprzyk, J. and Fedrizzi, M. (1989). A 'human-consistent' degree of consensus based on fuzzy logic with linguistic quantifiers, *Mathematical Social Sciences*, 18, 275–290.
45. Kacprzyk, J. and Fedrizzi, M., Eds. (1990). *Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory*, Kluwer, Dordrecht.
46. Kacprzyk, J., Fedrizzi, M. and Nurmi, H. (1992). Group decision making and consensus under fuzzy preferences and fuzzy majority, *Fuzzy Sets and Systems*, 49, 21–31.
47. Kacprzyk, J., Fedrizzi, M. and Nurmi, H. (1997). OWA operators in group decision making and consensus reaching under fuzzy preferences and fuzzy majority. In R.R. Yager and J. Kacprzyk (Eds.): *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Boston, pp. 193–206.
48. Kacprzyk, J. and Nurmi, H. (1998) Group decision making under fuzziness, in R. Słowiński (Ed.): *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, Kluwer, Boston, pp. 103–136.
49. Kacprzyk, J., Nurmi, H. and Fedrizzi, M., Eds. (1996). *Consensus under Fuzziness*, Kluwer, Boston.

50. Kacprzyk, J., Nurmi H. and Fedrizzi, M. (1999) Group decision making and a measure of consensus under fuzzy preferences and a fuzzy linguistic majority, In L.A. Zadeh and J. Kacprzyk (Eds.): *Computing with Words in Information/Intelligent Systems. Part 2. Foundations*, Physica–Verlag (Springer–Verlag), Heidelberg and New York, pp. 233-243.
51. Kacprzyk, J. and Roubens, M., Eds. (1988). *Non-Conventional Preference Relations in Decision Making*, Springer–Verlag, Heidelberg.
52. Kacprzyk, J. and Zadrożny, S. (2002) Collective choice rules in group decision making under fuzzy preferences and fuzzy majority: a unified OWA operator based approach. *Control and Cybernetics*, 31, 937–948.
53. Kacprzyk, J. and Zadrożny (2003) An Internet-based group decision support system, *Management*, VII (28), 4–10.
54. Kacprzyk J. and Zadrożny S. (2003) Dealing with imprecise knowledge on preferences and majority in group decision making: towards a unified characterization of individual and collective choice functions, *Bull. of the Polish Academy of Sciences. Tech. Sci.*, 3, 286–302.
55. Kacprzyk, J., Zadrożny, S. and Fedrizzi, M. (1997). An interactive GDSS for consensus reaching using fuzzy logic with linguistic quantifiers. In D. Dubois, H. Prade and R.R. Yager (Eds.): *Fuzzy Information Engineering–A Guided Tour of Applications*, Wiley, New York, pp. 567–574.
56. Kelly, J.S. (1978) *Arrow Impossibility Theorems.* Academic Press, New York.
57. Kelly, J.S. (1978) *Social Choice Theory: An Introduction*, Academic Press, New York.
58. Kim, J.B. (1983). Fuzzy rational choice functions, *Fuzzy Sets and Systems*, 10, 37–43.
59. Kuzmin, V.B. and Ovchinnikov, S.V. (1980a). Group decisions I: In arbitrary spaces of fuzzy binary relations, *Fuzzy Sets and Systems*, 4, 53–62.
60. Kuzmin, V.B. and Ovchinnikov, S.V. (1980b). Design of group decisions II: In spaces of partial order fuzzy relations, *Fuzzy Sets and Systems*, 4, 153–165.
61. Lagerspetz, E. (1995), Paradoxes and representation. *Electoral Studies*, 15, 83–92.
62. Loewer, B. and Laddaga, R. (1985). Destroying the consensus, in Loewer B., Guest Ed., Special Issue on Consensus, *Synthese*, 62 (1), pp. 79–96.
63. Montero, J. (1985) A note on Fung-Fu‘s theorem”, *Fuzzy Sets and Systems*, 13, 259–269.
64. Montero, J. (1987) Arrow‘s theorem under fuzzy rationality, *Behavioral Science*, 32, 267–273.
65. Montero, J. (1988) Aggregation of fuzzy opinions in a non-homogeneous group, *Fuzzy Sets and Systems*, 25, 15–20.
66. Montero, J. (1990) Single-peakedness in weighted aggregation of fuzzy opinions in a fuzzy group, in: Kacprzyk J and Fedrizzi, M. Eds., *Multiperson Decision Making Models*, Kluwer, Dordrecht, pp. 163–171.
67. Montero, J., Tejada, J. and Cutello, V. (1997) A general model for deriving preference structures from data, *European J. of Operational Research*, 98, 98–110.
68. Nurmi, H. (1981). Approaches to collective decision making with fuzzy preference relations, *Fuzzy Sets and Systems*, 6, 249–259.
69. Nurmi, H. (1982). Imprecise notions in individual and group decision theory: resolution of Allais paradox and related problems, *Stochastica*, VI, 283–303.
70. Nurmi, H. (1983). Voting procedures: a summary analysis, *British J. of Political Science*, 13, 181–208.
71. Nurmi, H. (1984). Probabilistic voting, *Political Methodology*, 10, 81–95.
72. Nurmi, H. (1987). *Comparing Voting Systems*, Reidel, Dordrecht.
73. Nurmi, H. (1997), Referendum design: an exercise in applied social choice theory, *Scandinavian Political Studies*, 20, 33–52.
74. Nurmi, H. (1998), Voting paradoxes and referenda, *Social Choice and Welfare*, 15, 333–350.
75. Nurmi, H. (1999), *Voting Paradoxes and How to Deal with Them*. Springer–Verlag, Berlin-Heidelberg-New York.
76. Nurmi, H. and Kacprzyk, J. (1991). On fuzzy tournaments and their solution concepts in group decision making, *Europ. J. of Operational Research*, 51, 223–232.
77. Nurmi, H. and Kacprzyk, J. (2000) Social choice under fuzziness: a perspective. In: J. Fodor, B. De Baets and P. Perny (Eds.): *Preferences and Decisions under Incomplete Knowledge*. PhysicaVerlag (SpringerVerlag), Heidelberg and New York, pp. 107–130.

78. Nurmi, H., Kacprzyk, J. and Fedrizzi, M. (1996). Probabilistic, fuzzy and rough concepts in social choice, *Europ. J. of Operational Research*, 95, 264–277.
79. Roubens, M. and Vincke, Ph. (1985). *Preference Modelling*, Springer–Verlag, Berlin.
80. Salles, M. (1996). Fuzzy utility. In S. Barberá, P.J. Hammond and C. Seidl (Eds.): *Handbook of Utility Theory*, Kluwer, Boston.
81. Sengupta, K. (1999), Choice rules with fuzzy preferences: some characterizations, *Social Choice and Welfare*, 16, 259–272.
82. Szmidt, E. and Kacprzyk, J. (1996). Intuitionistic fuzzy sets in group decision making, *Notes on Intuitionistic Fuzzy Sets*, 2, 15–32.
83. Tanino, T. (1984). Fuzzy preference orderings in group decision making, *Fuzzy Sets and Systems*, 12, 117–131.
84. Yager, R.R (1988). On ordered weighted averaging aggregation operators in multicriteria decision making, *IEEE Trans. on Systems, Man and Cybernetics*, SMC-18, 183–190.
85. Yager, R.R. and Kacprzyk, J. (Eds.) (1997). *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Boston.
86. Zadeh, L.A. (1983). A computational approach to fuzzy quantifiers in natural languages, *Computers and Maths. with Appls.*, 9, 149–184.
87. Zadrożny, S. (1997). An approach to the consensus reaching support in fuzzy environment. In: J. Kacprzyk, H. Nurmi and M. Fedrizzi (Eds.): *Consensus under Fuzziness*. Kluwer, Boston, pp. 83–109.

# Extension of Some Voting Systems to the Field of Gradual Preferences

**Bonifacio Llamazares and José Luis García-Lapresta**

**Abstract** In the classical theory of social choice, there exist many voting procedures for determining a collective preference on a set of alternatives. The simplest situation happens when a group of individuals has to choose between two alternatives. In this context, some voting procedures such as simple and absolute special majorities are frequently used. However, these voting procedures do not take into account the intensity with which individuals prefer one alternative to the other. In order to consider this situation, one possibility is to allow individuals showing their preferences through values located between 0 and 1. In this case, the collective preference can be obtained by means of an aggregation operator. One of the most important matter in this context is how to choose such aggregation operator. When we consider the class of OWA operators, it is necessary to determine the associated weights. In this contribution we survey several methods for obtaining the OWA operator weights. We pay special attention to the way the weights are chosen, regarding the concrete voting system we want to obtain when individuals do not grade their preferences between the alternatives.

## 1 Introduction

When a group of agents have to choose between two alternatives taking into account the individual opinions, there exist two main features for reaching this decision: how the agents show their preferences and how to aggregate the information they provide. With respect to the first aspect, individuals can declare their opinions in a dichotomous manner by showing which alternative is the best or by declaring indifference between these alternatives (this is the common way in classic voting systems); another possibility consists on allowing individuals to show gradual preferences in someway. In this sense, valued and fuzzy preferences consider numerical values for declaring intensities of preference (see Nurmi [57], Tanino [63], Fodor and Roubens [25], De Baets and Fodor [12], García-Lapresta and Llamazares [29, 30], Llamazares and García-Lapresta [44], Llamazares [41, 43] and Fodor and De Baets [23], among others). After Zadeh [83], linguistic preferences have been very used in the Decision Theory framework (see Delgado et al. [14, 15], Yager [76],

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008

Herrera et al. [36, 37, 38], Bordogna et al. [5], Herrera and Herrera-Viedma [35]) and in voting systems (see García-Lapresta [28] and García-Lapresta et al. [31]).

It is worth emphasizing that the outcome of a vote depends not only on the aggregation procedure, but on the way the individuals show their opinions. In fact, individuals usually feel different intensities of preference when they compare pairs of alternatives. However, in the classic voting procedures they can not declare these intensities, and different modalities of preference and identified in a unique way—for instance, as if they feel extreme preference. So, individual opinions are truncated and misrepresented. In order to avoid this drawback, an interesting problem is to extend classic voting systems in such a way that they could aggregate intensities of preference. A possibility is to consider the following procedure: Once individuals show their preferences through a value between 0 and 1, we obtain the collective intensity of preference by means of an OWA operator. From this value and through a kind of strong $\alpha$-cut, where $\alpha \in [\frac{1}{2}, 1)$, we can decide if an alternative is chosen or if both alternatives are collectively indifferent.

When individuals do not grade their preferences, the previous procedure allows us to obtain a voting system in the classic way. Hence, once fixed $\alpha$, it is possible to know what class of voting systems underlies in the aggregation process according to the used OWA operator. In this sense, we show that we can determine —by means of the aforementioned procedure—the OWA operators which generalize simple, absolute, Pareto, unanimous and absolute special majorities. Moreover, because there exist multiple OWA operators generating a specific classic voting system, we also propose a way for choosing the best-suited OWA operator. On the other hand, it is worth emphasizing that the induced extensions maintain some good properties of the genuine voting procedures such as symmetry, self-duality, monotonicity and unanimity.

Although in this chapter we focus on OWA operators, it is worth emphasizing that the previous procedure has been already used to characterize some classes of aggregation operators which extend some voting systems. So, García-Lapresta and Llamazares [30] generalize two classes of majorities based on difference of votes by using quasiarithmetic means and window (olympic) OWA operators as aggregation operators.

The chapter is organized as follows. Section 2 is devoted to crisp aggregation operators, their properties and some classic voting systems that can be defined with them. Moreover, we point out some drawbacks of some usual voting systems. In Sect. 3, we deal with general aggregation operators and particularly with OWAs. We also survey some methods appeared in the literature in order to determine the OWA operator weights. In Sect. 4, we extend several classic voting systems through OWA operators in the case that individuals show intensities of preference by means of numerical values within the unit interval, and we provide some characterization results. Moreover, we propose a method for choosing the best-suited OWA operators which extend some voting systems. We also obtain the crisp aggregation operators associated with these best-suited OWAs. Finally, some conclusions are included in Sect. 5.

## 2 Crisp Aggregation Operators

The simplest situation in the collective decision making procedures happens when a group of individuals has to choose between two alternatives $x$ and $y$. When individuals do not grade their preferences, some authors, such as May [49] and Fishburn [21], have used an index $d$ in order to distinguish among the three possible cases of ordinary preference and indifference between $x$ and $y$:

$$d = \begin{cases} 1, & \text{if } x \text{ is preferred to } y, \\ 0, & \text{if } x \text{ is indifferent to } y, \\ -1, & \text{if } y \text{ is preferred to } x. \end{cases}$$

In order to extend classical voting systems to the field of gradual preferences we define an equivalent index $r = (d + 1)/2$. In this way, we have:

$$r = \begin{cases} 1, & \text{if } x \text{ is preferred to } y, \\ \frac{1}{2}, & \text{if } x \text{ is indifferent to } y, \\ 0, & \text{if } y \text{ is preferred to } x. \end{cases}$$

We consider $m$ voters, $m \geq 2$, who show their preferences between $x$ and $y$. A *crisp profile* is a vector $\boldsymbol{r} = (r_1, \ldots, r_m) \in \{0, \frac{1}{2}, 1\}^m$ which describes the voters' preferences between $x$ and $y$. For each crisp profile, the collective preference will be obtained by means of a crisp aggregation operator.

**Definition 1.** *A* **crisp aggregation operator (CAO)** *is a mapping* $H : \{0, \frac{1}{2}, 1\}^m \longrightarrow \{0, \frac{1}{2}, 1\}$.

The interpretation of collective preference is consistent with the foregoing interpretation for individual preferences. So, $H(\boldsymbol{r})$ shows us if an alternative is collectively preferred to the other or the alternatives are collectively indifferent, according to whether $H(\boldsymbol{r})$ is 1 ($x$ defeats $y$), 0 ($y$ defeats $x$) or $\frac{1}{2}$ ($x$ and $y$ tie).

We now consider some properties of CAOs that are well known in the literature: **Symmetry**, **self-duality**, **monotonicity** and **unanimity**. Symmetry, also referred to as **anonymity** and **equality**, means that the collective preference depends only on the set of individual preferences, but not on which individuals have these preferences; i.e., all voters are treated equally. Self-duality, also referred to as **neutrality** (May [49]), says that if everyone reverses his/her preferences between both alternatives, then the collective preference is also reversed; i.e., the alternatives are treated equally. Monotonicity means that if the individual support for an alternative increases, then the outcome for this alternative can not be worse than in the first case. And unanimity says that the collective preference coincides with individual preferences when these are the same. A characterization of the CAOs which simultaneously satisfy the three first properties can be found in Fishburn [21, p. 56].

On the sequel we will use the following notation: Given $a \in \mathbb{R}$, we denote by $[a]$ the *integer part of* $a$, i.e., the largest integer smaller than or equal to $a$. Given $\boldsymbol{r}, \boldsymbol{s} \in \{0, \frac{1}{2}, 1\}^m$ and $\sigma$ a permutation on $\{1, \ldots, m\}$, we denote $\boldsymbol{r}_\sigma = (r_{\sigma(1)}, \ldots, r_{\sigma(m)})$;

$\mathbf{1} = (1, \ldots, 1)$; $t\,\mathbf{1} = (t, \ldots, t)$; $\boldsymbol{r} \geq \boldsymbol{s}$ will mean $r_i \geq s_i$ for all $i \in \{1, \ldots, m\}$; and $\boldsymbol{r} > \boldsymbol{s}$ will denote $\boldsymbol{r} \geq \boldsymbol{s}$ and $\boldsymbol{r} \neq \boldsymbol{s}$.

**Definition 2.** *Let $H$ be a CAO.*

1. *$H$ is* **symmetric** *if for all crisp profile $\boldsymbol{r}$ and all permutation $\sigma$ of $\{1, \ldots, m\}$ it holds*

$$H(\boldsymbol{r}_\sigma) = H(\boldsymbol{r}).$$

2. *$H$ is* **self-dual** *if for all crisp profile $\boldsymbol{r}$ it holds*

$$H(\mathbf{1} - \boldsymbol{r}) = 1 - H(\boldsymbol{r}).$$

3. *$H$ is* **monotonic** *if for all pair of crisp profiles $\boldsymbol{r}$ and $\boldsymbol{s}$ it holds*

$$\boldsymbol{r} \geq \boldsymbol{s} \;\Rightarrow\; H(\boldsymbol{r}) \geq H(\boldsymbol{s}).$$

4. *$H$ is* **unanimous** *if for all $t \in \{0, \frac{1}{2}, 1\}$ it holds*

$$H(t\,\mathbf{1}) = t.$$

Starting from the previous properties we can obtain some interesting consequences. We will use the following notation: the cardinal of a set will be denoted by #; given a crisp profile $\boldsymbol{r}$, we denote

$$n_x(\boldsymbol{r}) = \#\{i \mid r_i = 1\}, \qquad n_y(\boldsymbol{r}) = \#\{i \mid r_i = 0\};$$

i.e., $n_x(\boldsymbol{r})$ is the number of individuals who prefer $x$ to $y$, while $n_y(\boldsymbol{r})$ is the number of individuals who prefer $y$ to $x$.

For each crisp profile $\boldsymbol{r}$, if $H$ is a symmetric CAO, then $H(\boldsymbol{r})$ depends only on $n_x(\boldsymbol{r})$ and $n_y(\boldsymbol{r})$. If $H$ is a self-dual CAO, then it is characterized by the set

$$H^{-1}(\{1\}) = \{\boldsymbol{r} \in \{0, \tfrac{1}{2}, 1\}^m \mid H(\boldsymbol{r}) = 1\},$$

since

$$H^{-1}(\{0\}) = \{\boldsymbol{r} \in \{0, \tfrac{1}{2}, 1\}^m \mid \mathbf{1} - \boldsymbol{r} \in H^{-1}(\{1\})\},$$

$$H^{-1}(\{\tfrac{1}{2}\}) = \{0, \tfrac{1}{2}, 1\}^m \setminus \left(H^{-1}(\{1\}) \cup H^{-1}(\{0\})\right).$$

Therefore, any self-dual CAO can be defined by means of the crisp profiles where the CAO takes the value 1. Based on this approach, we now show some of the most popular voting systems.

**Definition 3.**

1. *The* **simple majority***, $H_S$, is the self-dual CAO defined by*

$$H(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) > n_y(\boldsymbol{r}).$$

2. *The* **absolute majority***, $H_A$, is the self-dual CAO defined by*

$$H(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) > \frac{m}{2}.$$

3. *The* **Pareto majority***, $H_P$, is the self-dual CAO defined by*

$$H(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) > 0 \;\; and \;\; n_y(\boldsymbol{r}) = 0.$$

4. *The* **unanimous majority***, $H_U$, is the self-dual CAO defined by*

$$H(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) = m.$$

5. *Given $\beta \in [\frac{1}{2}, 1)$, the* **absolute special majority** *$Q_\beta$ is the self-dual CAO defined by*

$$H(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) > \beta m.$$

Notice that all the previous voting systems have been defined by means of only $n_x(\boldsymbol{r})$, $n_y(\boldsymbol{r})$ and $m$. Then, their associated CAOs are symmetric.

Simple majority has been widely studied in the literature. It is worth emphasizing that the first axiomatic characterization of it was given by May [49]. Other characterizations of simple majority can be found in Fishburn [21, 22], Campbell [8, 9], Maskin [48], Campbell and Kelly [10, 11], Aşan and Sanver [1], Woeginger [68], Miroiu [51], Yi [82] and Llamazares [42].

In relation to absolute majority, it has been characterized by Fishburn [21, p. 60], while Pareto majority has been characterized by Sen [60, p. 76]. On the other hand, unanimous majority has been characterized by Woeginger [68] and Llamazares [42].

Clearly, absolute special majorities are located between absolute majority, for $\beta = 1/2$, and unanimous majority, for $\beta \geq (m-1)/m$. They have been studied by Fishburn [21, p. 67] (without self-duality assumption) and Ferejohn and Grether [18].

It is worth mentioning that the choice of a voting system is not trivial because all of them have several drawbacks. In order to explain this matter, we consider 1001 voters. The ordered pair $(n_x(\boldsymbol{r}), n_y(\boldsymbol{r}))$ represents the result of a ballot (obviously, $1001 - n_x(\boldsymbol{r}) - n_y(\boldsymbol{r})$ is the number of voters who are indifferent between $x$ and $y$). Suppose that the result of a ballot is $(1, 0)$. In this case, under simple majority voting, $x$ wins. So, in simple majority an alternative can be elected with very poor support. Moreover, in this situation —in fact, this happens when the result is $(n + 1, n)$ or $(n, n+1)$, for some $n \geq 0$ such that $2n+1 \leq m$ — the winning alternative can change when a single turncoat alters his preference between $x$ and $y$.

In order to avoid the problem of minimum support, we would be able to use absolute majority. In this case, the problem of minimum support disappears but we continue to have a problem of stability when the result of the ballot is $(501, 500)$ or $(500, 501)$. Furthermore, a new drawback appears because under absolute majority the winning alternative needs a hight quantity of votes. Consequently, there is a loss of decisiveness and, in many instances, there is no winning alternative. On the other hand, if we use absolute majority so that the winning alternative has a wide support, it is paradoxical that $x$ wins if the result of the ballot is $(501, 500)$ but not when the result is $(500, 0)$.

Pareto majority shares with simple and absolute majorities similar drawbacks. Since $x$ wins when the result of a ballot is $(1, 0)$, there are problems of minimum support and stability. Moreover, there exists a winning alternative only if the result of the ballot is $(n, 0)$ or $(0, n)$, with $n \geq 1$; therefore, there is a problem of decisiveness. For its part, unanimous majority has a large problem of decisiveness: It is the less decisive among self-dual, monotonic and no constant CAOs.

In relation to absolute special majorities, they have a problem of decisiveness because the winning alternative needs at least $[1001\,\beta] + 1$ votes even the other alternative lacks support. Moreover, this loss of decisiveness increases as $\beta$ increases.

The previous analysis shows that decisiveness and stability are conflicting concepts. Therefore, it seems very interesting to look for a balance between them. In order to achieve this objective, it is necessary to formalize these concepts.

The notion of decisiveness was used by Ferejohn and Grether [18], under the **strongness** name, for analyzing absolute special majorities.

**Definition 4.** *Given two CAOs $H_1$ and $H_2$, we say that $H_1$ is* **as decisive as** *$H_2$ if for all $\boldsymbol{r} \in \{0, \frac{1}{2}, 1\}^m$ it holds*

$$H_2(\boldsymbol{r}) = 1 \;\Rightarrow\; H_1(\boldsymbol{r}) = 1, \qquad\qquad H_2(\boldsymbol{r}) = 0 \;\Rightarrow\; H_1(\boldsymbol{r}) = 0.$$

Obviously, if $H_1$ is as decisive as $H_2$, then $H_1 = H_2$ or there exists some profile $\boldsymbol{r} \in \{0, \frac{1}{2}, 1\}^m$ such that $H_2(\boldsymbol{r}) = \frac{1}{2}$ and $H_1(\boldsymbol{r}) \neq \frac{1}{2}$, i.e., $H_1$ is **more decisive** than $H_2$.

When we come to define the notion of stability, it is necessary to bear in mind that for any no constant CAO there exist crisp profiles where an alternative stops winning when a single voter changes his/her preference. Consequently, we consider that a CAO is stable of grade $q$ ($q$-stable) when given any profile where there exists a winning alternative, $q$ voters can change their preferences without the other alternative becomes the winner. Since the case where an alternative can never win lacks interest, we also ask in the definition of $q$-stability that there be profiles where the change in the opinion of $q + 1$ individuals produces the switch of the winning alternative.

**Definition 5.** *Given $q \in \{0, 1, \ldots, m - 1\}$, a CAO $H$ is* **$q$-stable** *if it satisfies the following conditions:*

1. *For all $\boldsymbol{r}, \boldsymbol{s} \in \{0, \frac{1}{2}, 1\}^m$ such that $\#\{i \mid r_i \neq s_i\} \leq q$,*

$$H(\boldsymbol{r}) = 1 \;\Rightarrow\; H(\boldsymbol{s}) \geq \tfrac{1}{2}, \qquad\qquad H(\boldsymbol{r}) = -1 \;\Rightarrow\; H(\boldsymbol{s}) \leq \tfrac{1}{2}.$$

2. *There exist* $\boldsymbol{r}, \boldsymbol{s} \in \{0, \frac{1}{2}, 1\}^m$ *such that* $\#\{i \mid r_i \neq s_i\} = q + 1$ *satisfying* $H(\boldsymbol{r}) = 1$ *and* $H(\boldsymbol{s}) = 0$.

We note that Theorem 7 in the Subsection 4.2 shows what is the best voting system regarding a balance between decisiveness and stability.

## 3 Aggregation Operators

In order that voters can show different levels of intensity between the alternatives, we consider $r_i \in [0, 1]$ instead of $r_i \in \{0, \frac{1}{2}, 1\}$. In this way, $r_i$ denotes the *intensity* with which voter $i$ prefers $x$ to $y$. Under this assumption, it is usual to suppose that the preferences are *reciprocal*, i.e., $1 - r_i$ is the intensity with which individual $i$ prefers $y$ to $x$ (on this see Bezdek et al. [4], Nurmi [57], Tanino [63], Nakamura [55], Świtalski [61, 62], García-Lapresta and Llamazares [29] and De Baets et al. [13], among others).

Similarly to the crisp case, a *profile* is a vector $\boldsymbol{r} = (r_1, \ldots, r_m) \in [0, 1]^m$ which describes the voters' preferences between $x$ and $y$. For each profile, the collective preference will be obtained by means of an aggregation operator. These functions have been widely studied in the literature (see e.g. Dubois and Prade [17], Mizumoto [52, 53], Dubois and Koning [16], Yager [74], Fodor and Roubens [25], Grabisch et al. [34], Marichal [46], Calvo et al. [6], Xu and Da [72], Beliakov and Calvo [3], Mesiar et al. [50] and Torra [66], among other contributions).

**Definition 6.** *An* **aggregation operator** *is a mapping* $F : [0, 1]^m \longrightarrow [0, 1]$.

The properties introduced in Definition 2 for CAO's are also valid for aggregation operators, by considering general profiles instead of crisp profiles, and they have a similar interpretation. Moreover, we present an additional property for aggregation operators: **strict monotonicity**. This property means that the collective intensity of preference increases if no individual intensity decreases and some individual intensity increases.

**Definition 7.** *An aggregation operator* $F$ *is* **strictly monotonic** *if for every pair of different profiles* $\boldsymbol{r}, \boldsymbol{s} \in [0, 1]^m$ *it holds*

$$\boldsymbol{r} > \boldsymbol{s} \;\Rightarrow\; F(\boldsymbol{r}) > F(\boldsymbol{s}).$$

Although there exist numerous classes of aggregation operators, on the sequel we only focus our study on OWA operators. Nevertheless, other aggregation operators such as quasiarithmetic means have been also considered in this context by García-Lapresta and Llamazares [29, 30] and Llamazares and García-Lapresta [44].

On the other hand, since discrete Sugeno and Choquet integrals have been widely used as aggregation operators in multi-criteria decision making problems, we next do some considerations. Kandel and Byatt [39] have proven that the discrete Sugeno integral can be expressed in terms of the median. Moreover, it satisfies a similar property to stability under positive linear transformations, but for ordinal values. For these reasons, as Grabisch [33] points out, the discrete Sugeno integral seems to be more suitable for ordinal aggregation.

In relation to discrete Choquet integral, it is well-known that these aggregation functions generalize OWA operators (see e.g. Murofushi and Sugeno [54] and Fodor et al. [24]). Furthermore Grabisch [32] has proven that the class of OWA operators coincides with the class of symmetric discrete Choquet integrals. Therefore, since symmetry is an essential property in Social Choice Theory, it is sufficient to consider OWA operators.

## *3.1 OWA Operators*

Yager [28] introduced the OWA operators as a tool for aggregation procedures in multi-criteria decision making. An OWA operator is similar to a weighted mean, but with the values of the variables previously ordered in a decreasing way. Thus, contrary to the weighted means, the weights are not associated with concrete variables and, therefore, they are symmetric. Moreover, they verify other interesting properties, such as monotonicity, unanimity, continuity and compensativeness, i.e., the value of an OWA operator is always located between the minimum and the maximum values of the variables. On the other hand, OWA operators generalize the maximum and the minimum operators, the arithmetic mean, the median and the $k$-order statistic. For these reasons, OWA operators have been widely used in the literature (see, for instance, Yager and Kacprzyk [81] and Calvo et al. [7]).

OWA operators have been characterized by Fodor et al. [24], Ovchinnikov [59], Marichal and Mathonet [47]. On the other hand, there exist numerous generalizations of OWA operators: for instance, the quasi-OWA operators (Fodor et al. [24]), the Weighted Ordered Weighted Averaging (WOWA) Operators (Torra [64]), the Weighted Order Statistic Averaging (WOSA) operators (Ovchinnikov [59]), the Induced Ordered Weighted Averaging (IOWA) operators (Yager and Filev [80]), and the Heavy Ordered Weighted Averaging (HOWA) operators (Yager [78]).

OWA operators are usually defined as functions whose domain is $\mathbb{R}^m$. However, since individual intensities of preference vary between 0 and 1, we have restricted their domain to $[0, 1]^m$.

**Definition 8.** *Let* $\boldsymbol{w} = (w_1, \ldots, w_m) \in [0, 1]^m$ *satisfying* $\sum_{i=1}^m w_i = 1$. *The* **OWA operator associated with** $\boldsymbol{w}$ *is the aggregation operator* $F^{\boldsymbol{w}}$ *defined by*

$$F^{\boldsymbol{w}}(\boldsymbol{r}) = \sum_{i=1}^{m} w_i \, r_{\sigma(i)},$$

*where* $\sigma$ *is a permutation of* $\{1, \ldots, m\}$ *such that* $r_{\sigma(1)} \geq \cdots \geq r_{\sigma(m)}$.

OWA operators are symmetric, monotonic and unanimous aggregation operators. Self-dual OWA operators have been characterized by Marichal [46, p. 103] and García-Lapresta and Llamazares [30], while characterizations of strictly monotonic OWA operators can be found in Marichal [46, p. 103] and Llamazares [43].

**Proposition 1.** *If* $F^{\mathbf{w}}$ *is an OWA operator, then:*

1. $F^{\mathbf{w}}$ *is self-dual if and only if* $w_{m+1-i} = w_i$ *for every* $i \in \{1, \ldots, [\frac{m}{2}]\}$.
2. $F^{\mathbf{w}}$ *is strictly monotonic if and only if* $w_i > 0$ *for every* $i \in \{1, \ldots, m\}$.

We will denote by $\mathcal{W}$ the set of weighting vectors associated with self-dual OWA operators, i.e.,

$$\mathcal{W} = \left\{ \mathbf{w} \in [0,1]^m \mid \sum_{i=1}^{m} w_i = 1 \text{ and } w_{m+1-i} = w_i \text{ for all } i \in \{1, \ldots, [\tfrac{m}{2}]\} \right\}.$$

One of the most important issues in the field of OWA operators is the determination of the associated weights. In order to solve this problem, several methods have appeared in the literature: quantifier guided aggregation (Yager [28, 29, 77]), exponential smoothing (Filev and Yager [20]), learning approach (Yager and Filev [80], Torra [65]), genetic algorithms (Nettleton and Torra [56]), linear objective-programming model under partial weight information (Xu and Da [71]), parametric geometric approach (Liu and Chen [40]), normal distribution based method (Xu [69]) and argument-dependent approach (Xu [70]; see Beliakov and Calvo [2] as well). Another important class of these methods is based on the **orness measure**. This concept was introduced by Yager [28] for characterizing the degree with which the aggregation is like an *or* operation. The orness of an OWA operator ranges between 0 and 1. Moreover, it takes the value 0 if and only if the OWA operator is the **minimum**, and it only takes the value 1 when the OWA operator is the **maximum**.

**Definition 9.** *The* **orness measure** *of an OWA operator* $F^{\mathbf{w}}$ *is defined by*

$$orness\,(\mathbf{w}) = \frac{1}{m-1} \sum_{i=1}^{m} (m-i)\, w_i .$$

O'Hagan [58] was the first to suggest the use of the vector which maximizes the entropy of the OWA weights for a given level of orness. His approach is based on the solution of the following constrained optimization problem:

$$\begin{aligned}
\max\ & -\sum_{i=1}^{m} w_i \ln(w_i), \\
\text{s.t.}\ & \mathit{orness}\,(\boldsymbol{w}) = \alpha, \quad 0 \le \alpha \le 1, \\
& w_i \ge 0, \quad i = 1, \ldots, m, \\
& \sum_{i=1}^{m} w_i = 1.
\end{aligned} \tag{1}$$

The analytic properties of these maximal entropy OWA operators were studied by Filev and Yager [19]. Fullér and Majlender [26] transferred problem (1) to a polynomial equation, which is solved for determining the optimal weighting vector. To avoid the resolution of a nonlinear optimization problem, Yager [79] introduces a simpler procedure which tries to keep the spirit of maximizing the entropy for a given level of orness. Similar approaches (through a fixed level of orness) have also been proposed and solved analytically by Fullér and Majlender [27], Wang and Parkan [67] and Majlender [45]. The first authors suggest to select the vector which minimizes the variance of the weighting vector in order to obtain the minimal variability OWA weights. Thus, if we denote by $\boldsymbol{W}_\alpha$ the set of constraints of problem (1), they propose and solve the following mathematical programming problem:

$$\begin{aligned}
\min\ & \frac{1}{m} \sum_{i=1}^{m} w_i^2 - \frac{1}{m^2}, \\
\text{s.t.}\ & \boldsymbol{w} \in \boldsymbol{W}_\alpha.
\end{aligned} \tag{2}$$

Wang and Parkan [67] suggest to minimize the disparities between adjacent weights. For this, they bring up the following constrained optimization problem:

$$\begin{aligned}
\min\ & \max_{i \in \{1,\ldots,m-1\}} |w_i - w_{i+1}|, \\
\text{s.t.}\ & \boldsymbol{w} \in \boldsymbol{W}_\alpha.
\end{aligned} \tag{3}$$

For his part, Majlender [45] determines a parametric class of OWA operators having maximal Rényi entropy OWA weights. Given $\theta \in \mathbb{R}$, the parametric model proposed and solved by this author is the following:

$$\begin{aligned}
\max\ & \log_2 \left( \sum_{i=1}^{m} w_i^\theta \right)^{1/(1-\theta)}, \\
\text{s.t.}\ & \boldsymbol{w} \in \boldsymbol{W}_\alpha.
\end{aligned} \tag{4}$$

However, in these models, the aggregation procedure is only taken into account when the level of orness is previously fixed, and the weights are determined through properties that only concern them (for instance, to maximize the entropy, to minimize the variance, etc.). Moreover, in order to guarantee an egalitarian treatment between the alternatives, we consider self-dual OWA operators whose level of orness is 0.5. Therefore, in our case, the methodologies based on fixing a given level of orness lack sense because the arithmetic mean is always the solution.

Because of the previous reasons, we propose the choice of OWA operator weights to take into account the class of majority rule that we want to obtain when individuals do not grade their preferences between the alternatives. For this purpose, the procedure to follow is described in the following subsection.

### 3.2 Obtaining CAOs from Aggregation Operators

When individuals have crisp preferences, every aggregation operator $F$ can be restricted to crisp profiles:

$$F|_{\{0,\frac{1}{2},1\}^m} : \{0, \tfrac{1}{2}, 1\}^m \longrightarrow [0, 1].$$

If we wish to obtain a CAO from $F$, then it will be necessary to obtain collective intensities of preference within $\{0, \frac{1}{2}, 1\}$ instead of $[0, 1]$. It is possible to get these values by means of a procedure based on the $\alpha$-cuts of $F$.

**Definition 10.** *Let $F$ be an aggregation operator and $\alpha \in [\frac{1}{2}, 1)$. The* $\alpha$**–CAO associated with** *$F$ is the CAO $F_\alpha$ defined by*

$$F_\alpha(\boldsymbol{r}) = \begin{cases} 1, & \text{if } F(\boldsymbol{r}) > \alpha, \\ \frac{1}{2}, & \text{if } 1 - \alpha \le F(\boldsymbol{r}) \le \alpha, \\ 0, & \text{if } F(\boldsymbol{r}) < 1 - \alpha. \end{cases}$$

Thus, when individuals have crisp preferences, we can generate different CAOs from an aggregation operator $F$, by considering appropriate values of the parameter $\alpha \in [\frac{1}{2}, 1)$. Moreover, these CAOs, $F_\alpha$, are symmetric, self-dual, monotonic and unanimous whenever the original aggregation operator $F$ satisfies these properties.

## 4 Generating Voting Systems from OWA Operators

In Definition 10 we provide a procedure which generates CAOs from an aggregation operator by means of a parameter $\alpha \in [\frac{1}{2}, 1)$. Now, we will use this procedure for obtaining the voting systems appearing in Definition 3. In this way, the following subsection is devoted to characterizing the OWA operators that allow us to generalize the mentioned voting systems. However, although we can obtain a specific CAO by means of a wide variety of OWA operators, not all of them are suitable, as we show in Subsection 4.2. For this reason, we also propose a procedure to determine the best-suited OWA operators.

Since the considered CAOs are self-dual, we only take into account self-dual OWA operators in order to guarantee that the obtained $\alpha$-CAOs be self-dual too. The results contained in this section can be found in Llamazares [41, 43].

## 4.1 Characterization Results

In the following theorems we characterize the OWA operators for which we can generate simple, absolute, Pareto, unanimous and absolute special majorities.

**Theorem 1.** *Let $F^{\mathbf{w}}$ be a self-dual OWA operator and $\alpha \in [\frac{1}{2}, 1)$. Then the following statements are equivalent:*

1. $F_\alpha^{\mathbf{w}} = H_S$.
2. *$F^{\mathbf{w}}$ is strictly monotonic and* $\alpha < \dfrac{1 + \min\{w_1, \ldots, w_m\}}{2}$.

**Theorem 2.** *Let $F^{\mathbf{w}}$ be a self-dual OWA operator and $\alpha \in [\frac{1}{2}, 1)$. Then the following statements are equivalent:*

1. $F_\alpha^{\mathbf{w}} = H_A$.
2. a. *If $m$ is odd:* $w_{\frac{m+1}{2}} > \frac{1}{3}$ *and* $\dfrac{3 - w_{\frac{m+1}{2}}}{4} \leq \alpha < \dfrac{1 + w_{\frac{m+1}{2}}}{2}$.

   b. *If $m$ is even:* $w_{\frac{m}{2}} > \frac{1}{4}$ *and* $\frac{3}{4} \leq \alpha < \frac{1}{2} + w_{\frac{m}{2}}$.

**Theorem 3.** *Let $F^{\mathbf{w}}$ be a self-dual OWA operator and $\alpha \in [\frac{1}{2}, 1)$. Then the following statements are equivalent:*

1. $F_\alpha^{\mathbf{w}} = H_P$.
2. $w_1 > \frac{1}{3}$ *and* $1 - w_1 \leq \alpha < \dfrac{1 + w_1}{2}$.

**Theorem 4.** *Let $F^{\mathbf{w}}$ be a self-dual OWA operator and $\alpha, \beta \in [\frac{1}{2}, 1)$, with $[\beta m] > \frac{m}{2}$. Then the following statements are equivalent:*

1. $F_\alpha^{\mathbf{w}} = Q_\beta$.
2. $w_{m-[\beta m]} > \displaystyle\sum_{i=1}^{m-[\beta m]-1} w_i$ *and* $1 - \frac{1}{2} \displaystyle\sum_{i=1}^{m-[\beta m]} w_i \leq \alpha < 1 - \displaystyle\sum_{i=1}^{m-[\beta m]-1} w_i$.

As a particular case of this result (for $[\beta\, m] = m - 1$), we can obtain a characterization of self-dual OWA operators which generalize unanimous majority.

**Corollary 1.** *Let $F^{\mathbf{w}}$ be a self-dual OWA operator and $\alpha \in [\frac{1}{2}, 1)$. Then the following statements are equivalent:*

1. $F_\alpha^{\boldsymbol{w}} = H_U$.
2. $w_1 > 0$ *and* $\alpha \geq 1 - \dfrac{w_1}{2}$.

## 4.2 Choosing the Best-suited OWA Operators

As we have seen in Theorems 1, 2, 3, 4 and Corollary 1, there exist a lot of self-dual OWA operators which generate the same CAO. However, such as we will see in the following example, not all of them are suitable for a specific voting system.

*Example 1.* Consider $m = 5$, $\alpha = 0.504$ and $\boldsymbol{w} = (0.01,\ 0.01,\ 0.96,\ 0.01,\ 0.01)$. By Theorem 1, $F_\alpha^{\boldsymbol{w}} = H_S$. However, these weights are close to that of the vector $\boldsymbol{w}' = (0, 0, 1, 0, 0)$, and in this case, by Theorem 2, $F_\alpha^{\boldsymbol{w}'} = H_A$ for every $\alpha \in [\frac{1}{2}, 1)$ (in fact, with a suitable value of $\alpha$, we would be able to choose the weights of $\boldsymbol{w}$ as close as we want to that of $\boldsymbol{w}'$). Therefore, although $F_\alpha^{\boldsymbol{w}} = H_S$, the choice of $\boldsymbol{w}$ and $\alpha$ does not seem the best for representing simple majority.

In order to avoid the previous situation, we take into account that for each self-dual OWA operator $F^{\boldsymbol{w}}$ and for each symmetric, monotonic and self-dual CAO $H$, the set $\{\alpha \in [\frac{1}{2}, 1) \mid F_\alpha^{\boldsymbol{w}} = H\}$ is an interval with endpoints $\underline{\alpha}(\boldsymbol{w}, H)$ and $\overline{\alpha}(\boldsymbol{w}, H)$ (it can be empty). In this way, we propose to choose the self-dual OWA operators that maximize the measure of interval, i.e., the value $\overline{\alpha}(\boldsymbol{w}, H) - \underline{\alpha}(\boldsymbol{w}, H)$. In the following theorem we show the weighting vectors that satisfy this condition for simple, absolute, Pareto and absolute special majorities.

**Theorem 5.** *Let $H$ be a symmetric, monotonic and self-dual CAO and $\boldsymbol{w}^*$ be the weighting vector solution of the problem*

$$\max_{\boldsymbol{w} \in \mathcal{W}} \overline{\alpha}(\boldsymbol{w}, H) - \underline{\alpha}(\boldsymbol{w}, H).$$

*Then:*

1. *If $H = H_S$, then $F^{\boldsymbol{w}^*}$ is the arithmetic mean and*

$$F_\alpha^{\boldsymbol{w}^*} = H_S \ \Leftrightarrow\ \frac{1}{2} \leq \alpha < \frac{m+1}{2m}.$$

2. *If $H = H_A$, then $F^{\boldsymbol{w}^*}$ is the median. Moreover:*
   *(1) If $m$ is odd, then $F_\alpha^{\boldsymbol{w}^*} = H_A$ for every $\alpha \in [\frac{1}{2}, 1)$.*
   *(2) If $m$ is even, then $F_\alpha^{\boldsymbol{w}^*} = H_A \ \Leftrightarrow\ \frac{3}{4} \leq \alpha < 1$.*
3. *If $H = Q_\beta$, with $[\beta m] > \frac{m}{2}$, then*

$$w_i^* = \begin{cases} \frac{1}{2}, & \text{if } i = m - [\beta m], [\beta m] + 1, \\ 0, & \text{otherwise,} \end{cases}$$

*and $F_\alpha^{\boldsymbol{w}^*} = Q_\beta \ \Leftrightarrow\ \frac{3}{4} \leq \alpha < 1$.*

*4. If* $H = H_P$*, then*

$$w_i^* = \begin{cases} \frac{1}{2}, & \text{if } i = 1, m, \\ 0, & \text{otherwise}, \end{cases}$$

*and* $F_\alpha^{\boldsymbol{w}^*} = H_P \;\Leftrightarrow\; \frac{1}{2} \le \alpha < \frac{3}{4}$.

The obtained OWA operators are the arithmetic mean, the median and the average of the $j$-th and the $(m+1-j)$-th order statistics. Because they are the most suitable for generating some of the most important classes of voting systems, it would be interesting to know the $\alpha$–CAOs associated with these OWA operators.

Since the arithmetic mean is a specific case of quasiarithmetic means, we can consider the result given by García-Lapresta and Llamazares [30, Proposition 4] for **majorities based on difference of votes** in the framework of quasiarithmetic means.

**Theorem 6.** *If* $F^{\boldsymbol{w}}$ *is the arithmetic mean,* $\alpha \in [\frac{1}{2}, 1)$ *and* $k = [m(2\alpha - 1)]$*, then* $F_\alpha^{\boldsymbol{w}}$ *coincides with* $M_k$*, the self-dual CAO defined by*

$$M_k(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) > n_y(\boldsymbol{r}) + k.$$

The resultant CAOs are based on difference of votes: an alternative wins when the difference between the number of votes obtained by this alternative and that obtained by the other is greater than the quantity $[m(2\alpha - 1)]$. These voting systems were introduced in García-Lapresta and Llamazares [30] and they have been recently analyzed by Llamazares [42] within the Social Choice approach. As well as the good properties appearing in the previous mentioned papers, it is worth emphasizing that they are the best voting systems that we look for at the end of Sect. 2 in order to achieve a balance between decisiveness and stability.

**Theorem 7.** *Given* $k \in \{0, 1, \dots, m-1\}$ *and* $k' \in \{k, \dots, m-1\}$*, the* $M_k$ *majority is the most decisive symmetric, self-dual, monotonic and* $k'$*-stable CAO.*

In relation to the median, by 2.a) of Theorem 5, we have that $F_\alpha^{\boldsymbol{w}} = H_A$ for all $\alpha \in [\frac{1}{2}, 1)$ when $m$ is odd. When $m$ is even, the median is the average of the $m/2$-th and the $(m/2+1)$-th order statistics. Therefore, we can obtain the $\alpha$–CAOs associated with it as a particular case of the following result.

**Theorem 8.** *Let* $j \in \{1, \dots, [\frac{m}{2}]\}$ *and* $F^{\boldsymbol{w}}$ *be the OWA operator defined by*

$$w_i = \begin{cases} \frac{1}{2}, & \text{if } i = j, m+1-j, \\ 0, & \text{otherwise}. \end{cases}$$

*Then the following statements are satisfied:*

*1. If* $\frac{1}{2} \le \alpha < \frac{3}{4}$*, then* $F_\alpha^{\boldsymbol{w}}$ *coincides with the self-dual CAO* $H$ *defined by*

$$H(\boldsymbol{r}) = 1 \;\Leftrightarrow\; n_x(\boldsymbol{r}) \ge j \text{ and } n_y(\boldsymbol{r}) < j.$$

2. *If* $\frac{3}{4} \leq \alpha < 1$, *then* $F_\alpha^{\mathbf{w}} = Q_\beta$ *for* $1 - \frac{j}{m} \leq \beta < 1 - \frac{j-1}{m}$.

It is worth emphasizing that when $j = 1$, i.e., when the OWA operator is the average of the maximum and the minimum, the obtained $\alpha$–CAOs are Pareto and unanimous majorities:

$$F_\alpha^{\mathbf{w}} = \begin{cases} H_P, & \text{if } \frac{1}{2} \leq \alpha < \frac{3}{4}, \\ H_U, & \text{if } \frac{3}{4} \leq \alpha < 1. \end{cases}$$

## 5 Concluding Remarks

In this chapter we have considered voting situations where individuals only compare two alternatives. Although individuals usually feel different modalities of preference when they compare the feasible alternatives, classic voting systems require that voters show crisp preferences. Therefore, individuals are forced to identify very different circumstances. Consequently, the outcomes provided by the classic voting procedures could be no faithful with individual opinions and they could lead to inappropriate decisions.

We assume that individuals can show their intensities of preference by means of numerical values within the unit interval. In order to aggregate these intensities, we have considered aggregation operators. More concretely, we have focused our attention on OWA operators and we have presented some characterizations that determine which weighting vectors of self-dual OWA operators and which $\alpha$-cuts allow to generate the genuine considered voting systems (simple, absolute, Pareto, unanimous and absolute special majorities), when individual preferences are crisp. In this sense, by satisfying the obtained conditions we can extend the mentioned classic voting systems to the case of individuals show their intensities of preference through numerical values within the unit interval.

Among the variety of self-dual OWA operators generating a specific classic voting system, we have found those weighting vectors which can be considered as the best-suited for extending the aforementioned classic voting systems. It is worth emphasizing that the founded OWA operators allow us to obtain new voting systems, some of them satisfying very interesting properties. For instance, through the arithmetic mean we can obtain majorities based on difference of votes which achieve a balance between decisiveness and stability. Therefore, the extension of classical voting systems to the field of gradual preferences allows us to find new voting systems which are the solution of some posed problems within the classic voting theory.

**Acknowledgment** This work is partially supported by the Junta de Castilla y León (Consejería de Educación y Cultura, Project VA040A05), the Spanish Ministerio de Ciencia y Tecnología (Project MTM2005-08982-C04-02) and ERDF.

## References


1. Aşan G, Sanver MR (2002) Another characterization of the majority rule. Economics Letters 75: 409–413.
2. Beliakov G, Calvo T (2006) Identification of weights in aggregation operators. In this book.
3. Beliakov G, Calvo T (2006) Interpolarity type construction of general aggregation operators. In this book.
4. Bezdek JC, Spillman B, Spillman R (1978) A fuzzy relation space for group decision theory. Fuzzy Sets and Systems 1: 255–268.
5. Bordogna G, Fedrizzi M, Pasi G (1997) A linguistic modelling of consensus in group decision making based on OWA operators. IEEE Transactions on Systems, Man and Cybernetics 27: 126–132.
6. Calvo T, Kolesárová A, Komorníková M, Mesiar R (2002) Aggregation operators: Properties, classes and construction methods. In Calvo et al. [7], 3–104.
7. Calvo T, Mayor G, Mesiar R (eds) (2002) Aggregation Operators: New Trends and Applications. Physica-Verlag, Heidelberg.
8. Campbell DE (1982) On the derivation of majority rule. Theory and Decision 14: 133–140.
9. Campbell DE (1988) A characterization of simple majority rule for restricted domains. Economics Letters 28: 307–310.
10. Campbell DE, Kelly JS (2000) A simple characterization of majority rule. Economic Theory 15: 689–700.
11. Campbell DE, Kelly JS (2003) A strategy-proofness characterization of majority rule. Economic Theory 22: 557–568.
12. De Baets B, Fodor J (1997) Twenty years of fuzzy preference structures. Rivista di Matematica per le Scienze Economiche e Sociali 20: 45–66.
13. De Baets B, De Meyer H, De Schuymer B, Jenei S (2006) Cyclic evaluation of transitivity of reciprocal relations. Social Choice and Welfare 26: 217–238.
14. Delgado M, Verdegay JL, Vila MA (1993) Linguistic decision making models. International Journal of Intelligent Systems 7: 479–492.
15. Delgado M, Verdegay JL, Vila MA (1993) On aggregation operations of linguistic labels. International Journal of Intelligent Systems 8: 352–370.
16. Dubois D, Koning JL (1991) Social choice axioms for fuzzy set agregation. Fuzzy Sets and Systems 43: 257–274.
17. Dubois D, Prade H (1985) A review of fuzzy set aggregation connectives. Information Sciences 36: 85–121.
18. Ferejohn JA, Grether DM (1974) On a class of rational social decisions procedures. Journal of Economic Theory 8: 471–482.
19. Filev D, Yager RR (1995) Analytic properties of maximal entropy OWA operators. Information Sciences 85: 11–27.
20. Filev D, Yager RR (1998) On the issue of obtaining OWA operator weights. Fuzzy Sets and Systems 94: 157–169.
21. Fishburn PC (1973) The Theory of Social Choice. Princeton University Press, Princeton.
22. Fishburn PC (1983) A new characterization of simple majority. Economics Letters 13: 31–35.
23. Fodor J, De Baets, B (2006) Fuzzy preference modelling: Fundamentals and recent advances. In this book.
24. Fodor J, Marichal JL, Roubens M (1995) Characterization of the ordered weighted averaging operators. IEEE Transactions on Fuzzy Systems 3: 236–240.
25. Fodor J, Roubens M (1994) Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic Publishers, Dordrecht.
26. Fullér R, Majlender P (2001) An analytic approach for obtaining maximal entropy OWA operator weights. Fuzzy Sets and Systems 124: 53–57.
27. Fullér R, Majlender P (2003) On obtaining minimal variability OWA operator weights. Fuzzy Sets and Systems 136: 203–215.
28. García-Lapresta JL (2006) A general class of simple majority decision rules based on linguistic opinions. Information Sciences 176: 352–365.

29. García-Lapresta JL, Llamazares B (2000) Aggregation of fuzzy preferences: Some rules of the mean. Social Choice and Welfare 17: 673–690.
30. García-Lapresta JL, Llamazares B (2001) Majority decisions based on difference of votes. Journal of Mathematical Economics 35: 463–481.
31. García-Lapresta JL, Martínez-Panero M, Meneses LC (To appear) Defining the Borda count in a linguistic decision making context. Information Sciences.
32. Grabisch M (1995) Fuzzy integral in multicriteria decision making. Fuzzy Sets and Systems 69: 279–298.
33. Grabisch M (1996) The application of fuzzy integrals in multicriteria decision making. European Journal of Operational Research 89: 445–456.
34. Grabisch M, Orlovski SA, Yager RR (1998) Fuzzy aggregation of numerical preferences. In: Słowiński R (ed) Fuzzy Sets in Decision Analysis, Operations Research and Statistics. Kluwer Academic Publishers, Dordrecht, 31–68.
35. Herrera F, Herrera-Viedma E (2000) Linguistic decision analysis: Steps for solving decision problems under linguistic information. Fuzzy Sets and Systems 115: 67–82.
36. Herrera F, Herrera-Viedma E, Verdegay JL (1995) A sequential selection process in group decision making with linguistic assessment. Information Sciences 85: 223–239.
37. Herrera F, Herrera-Viedma E, Verdegay JL (1996) A linguistic decision process in group decision making. Group Decision and Negotiation 5: 165–176
38. Herrera F, Herrera-Viedma E, Verdegay JL (1996) Direct approach processes in group decision making using linguistic OWA operators. Fuzzy Sets and Systems 79:175–190.
39. Kandel A, Byatt WJ (1978) Fuzzy sets, fuzzy algebra and fuzzy statistics. Proceedings of IEEE 66: 1619–1639.
40. Liu X, Chen L (2004) On the properties of parametric geometric OWA operator. International Journal of Approximate Reasoning 35: 163–178.
41. Llamazares B (2004) Simple and absolute special majorities generated by OWA operators. European Journal of Operational Research 158: 707–720.
42. Llamazares B (2006) The forgotten decision rules: Majority rules based on difference of votes. Mathematical Social Sciences 51: 311–326.
43. Llamazares B (2006) Choosing OWA operator weights in the field of Social Choice. Mimeo.
44. Llamazares B, García-Lapresta JL (2003) Voting Systems generated by quasiarithmetic means and OWA operators. In: De Baets B, Fodor J (eds) Principles of Fuzzy Preference Modelling and Decision Making. Academia Press, Gent, 195–213.
45. Majlender P (2005) OWA operators with maximal Rényi entropy. Fuzzy Sets and Systems 155: 340–360.
46. Marichal JL (1998) Aggregation Operators for Multicriteria Decision Aid. MA Thesis, Liège University, Liège.
47. Marichal JL, Mathonet P (1999) A characterization of the ordered weighted averaging functions based on the ordered bisymmetry property. IEEE Transactions on Fuzzy Systems 7: 93–96.
48. Maskin ES (1995) Majority rule, social welfare functions and game forms. In: Basu K, Pattanaik PK, Suzumura K (eds) Choice, Welfare and Development. The Clarendon Press, Oxford, 100–109.
49. May KO (1952) A set of independent necessary and sufficient conditions for simple majority decision. Econometrica 20: 680–684.
50. Mesiar R, Komorníková M, Kolesárová A, Calvo T, (2006) Aggregation functions: A revision. In this book.
51. Miroiu A (2004) Characterizing majority rule: from profiles to societies. Economics Letters 85: 359–363.
52. Mizumoto M (1989) Pictorial representations of fuzzy connectives, part I: cases of t-norms, t-conorms and averaging operators. Fuzzy Sets and Systems 31: 217–242.
53. Mizumoto M (1989) Pictorial representations of fuzzy connectives, part II: cases of compensatory operators and self-dual operators. Fuzzy Sets and Systems 32: 45–79.
54. Murofushi T, Sugeno M (1993) Some quantities represented by the Choquet integral. Fuzzy Sets and Systems 56: 229–235.

55. Nakamura K (1992) On the nature of intransitivity in human preferential judgements. In: Novák V, Ramík J et al. (eds) Fuzzy Approach to Reasoning and Decision-Making. Kluwer Academic Publishers, Dordrecht, 147–162.
56. Nettleton D, Torra V (2001) A comparison of active set method and genetic algorithm approaches for learning weighting vectors in some aggregation operators. International Journal of Intelligent Systems 16: 1069–1083.
57. Nurmi H (1981) Approaches to collective decision making with fuzzy preference relations. Fuzzy Sets and Systems 6: 249–259.
58. O'Hagan M (1988) Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. Proc. 22nd Annual IEEE Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, California, 681–689.
59. Ovchinnikov SV (1998) An analytic characterization of some aggregation operators. International Journal of Intelligent Systems 13: 3–10.
60. Sen AK (1970) Collective Choice and Social Welfare. Holden-Day, San Francisco.
61. Świtalski Z (1999) Rationality of fuzzy reciprocal preference relations. Fuzzy Sets and Systems 107: 187–190.
62. Świtalski Z (2003) General transitivity conditions for fuzzy reciprocal preference matrices. Fuzzy Sets and Systems 137: 85–100.
63. Tanino T (1984) Fuzzy preference orderings in group decision making. Fuzzy Sets and Systems 12: 117–131.
64. Torra V (1997) The weighted OWA operator. International Journal of Intelligent Systems 12: 153–166.
65. Torra V (1999) On the learning of weights in some aggregation operators: The weighted mean and OWA operators. Mathware and Soft Computing 6: 249–265.
66. Torra V (2006) On the construction of models for decision using aggregation operators. In this book.
67. Wang YM, Parkan C (2005) A minimax disparity approach for obtaining OWA operator weights. Information Sciences 175: 20–29.
68. Woeginger GJ (2003) A new characterization of the majority rule. Economics Letters 81: 89–94.
69. Xu ZS (2005) An overview of methods for determining OWA weights. International Journal of Intelligent Systems 20: 843–865.
70. Xu ZS (2006) Dependent OWA operators. In: Torra V, Narukawa Y, Valls A, Domingo-Ferrer J (eds) Modeling Decisions for Artificial Intelligence: Third International Conference MDAI 2006, Lecture Notes in Artificial Intelligence 3885. Springer, Berlin, 172–178.
71. Xu ZS, Da QL (2002) The uncertain OWA operator. International Journal of Intelligent Systems 17: 569–575.
72. Xu ZS, Da QL (2003) An overview of operators for aggregating information. International Journal of Intelligent Systems 18: 953–969.
73. Yager RR (1988) On ordered weighted averaging operators in multicriteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18: 183–190.
74. Yager RR (1991) Connectives and quantifiers in fuzzy sets. Fuzzy Sets and Systems 40: 39–75.
75. Yager RR (1993) Families of OWA operators. Fuzzy Sets and Systems 59: 125–148.
76. Yager RR (1995) An approach to ordinal decision making. International Journal of Approximate Reasoning 12: 237–261.
77. Yager RR (1996) Quantifier guided aggregation using OWA operators. International Journal of Intelligent Systems 11: 49–73.
78. Yager RR (2002) Heavy OWA operators. Fuzzy Optimization and Decision Making 1:379–397.
79. Yager RR (2003) E-Z OWA weights. Proc. 10th International Fuzzy Systems Association World Congress, Istanbul, 39–42.
80. Yager RR, Filev D (1999) Induced ordered weighted averaging operators. IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics 29: 141–150.

81. Yager RR, Kacprzyk J (eds) (1997) The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer Academic Publishers, Dordrecht.
82. Yi J (2005) A complete characterization of majority rules. Economics Letters 87: 109–112.
83. Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning. Information Sciences. Part I 8: 199–249. Part II 8: 301–357. Part III 9: 43–80.

# A Linguistic Decision Based Model Applied to Olive Oil Sensory Evaluation

**Luis Martínez, Luis G. Pérez and Jun Liu**

**Abstract** The evaluation is a process that analyzes elements to achieve different objectives such as quality inspection, design, marketing exploitation and other fields in industrial companies. In many of these fields the items, products, designs, etc., are evaluated according to the knowledge acquired via human senses (sight, taste, touch, smell and hearing), in such cases, the process is called *Sensory Evaluation*. In this type of evaluation process, an important problem arises as it is the modelling and management of uncertain knowledge, because the information acquired by our senses throughout human perceptions involves uncertainty, vagueness and imprecision. The Fuzzy Linguistic Approach [34] has showed its ability to deal with uncertainty, ambiguity, imprecision and vagueness, so it seems logic and suitable the use of the Fuzzy Linguistic Approach to model the information provided by the experts in sensory evaluation processes.

The decision analysis has been usually used in evaluation processes because it is a formal methodology that can help to achieve the evaluation objectives. In this chapter we present a linguistic evaluation model for sensory evaluation based on the decision analysis scheme that will use the Fuzzy Linguistic Approach and the 2-tuple fuzzy linguistic representation to model and manage the uncertainty and vagueness of the information acquired through the human perceptions in the sensory evaluation process. This model will be applied to some sensory evaluation processes of the Olive Oil.

## 1 Introduction

The evaluation is a complex cognitive process that involves different mechanisms in which it is necessary to define the elements to evaluate, fix the evaluation framework, gather the information and obtain an evaluation assessment by means of an evaluation process. The aim of any evaluation process is to obtain information about the worth of an item (product, service, material, etc.), a complete description about different aspects, indicators, criteria in order to improve it or to compare with other items to know which ones are the best. The information gathered in this kind of processes is usually provided by a group of individuals, called panel of experts, where

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

each expert expresses their opinions about the item according to their knowledge and their own perceptions.

In this chapter our interest is focused on *Sensory Evaluation* processes[11, 29, 30] that is an evaluation discipline whose information, provided by a panel of experts, is perceived by the human senses of *sight, smell, taste, touch and hearing*. The sensory evaluation is widely used in:

- *Quality inspection* of food and textile products [1, 12, 13, 37] to determine systematically their characteristics by means of a group of experts.
- *Marketing studies* [22, 27] for understanding consumers behaviors and exploiting new markets.
- *Engineering processes* [7, 32] to integrate the data provided by the individuals in their design.
- Etc.

The sensory evaluation is based on the knowledge acquired in a sensory way by a panel of experts that take part in the evaluation process. A suitable mathematical formulation is not easy in this type of problems because human perceptions are subjective and not objective, therefore the assessments provided by the individuals are vague and uncertain. Initially classical computational techniques used in sensory evaluation were based on statistics and factorial analysis, but these methods are not efficient for solving sensory evaluation problems because uncertainties in this type of problems have a non-probabilistic character since they are related to imprecision and vagueness of meanings. In such a case, linguistic descriptors are direct provided by the experts to express their knowledge about the evaluated element. The Fuzzy Linguistic Approach [34] provides a systematic way to represent linguistic variables in an evaluation procedure. The use of linguistic variables implies processes of computing with words [20, 21, 33, 36] such as their fusion, aggregation, comparison, etc.

The evaluation process follows a methodology in order to achieve its objectives. The use of decision analysis approach has been successfully applied to evaluation problems in the literature [2, 8, 19, 25]. In decision theory before making a decision is carried out a decision analysis approach that allows people to make decisions more consistently, i.e., it helps people to deal with difficult decisions. The decision analysis is a suitable approach for evaluation processes because it helps to analyze the alternatives, aspects, indicators of the element/s under study that it is the objective of the evaluation processes. In the literature different linguistic decision making models can be found [6, 14, 24, 15].

The aim of this paper is to propose a linguistic sensory evaluation model based on a decision analysis scheme that uses the Fuzzy Linguistic Approach to represent the experts' assessments, and the 2-tuple fuzzy representation model [16] to provide a computational model to manage the processes of computing with words. And eventually to apply it to some sensory evaluation processes of the olive oil.

This paper is structured as follows, in Sect. 2 we present and review in short the necessary concepts and processes to develop the linguistic sensory evaluation. In Sect. 3 we present our proposal of linguistic sensory evaluation model, and in Sect. 4 we expound an application of this evaluation model. Finally, this paper is concluded in Sect. 5.

## 2 Preliminaries

Our evaluation model is based on the scheme of the Decision Analysis we shall present in this section. Moreover, we shall make a brief review of the Fuzzy Linguistic Approach and the Linguistic 2-tuple Representation Model that will be used to facilitate the computation of the lingusitic information in the evaluation process.

### *2.1 Decision Analysis Steps*

The Decision Analysis is a discipline, which belongs to Decision Making Theory, whose purpose is to help the decision makers to reach a consistent decision in a decision making problem. The evaluation process can be modelled as different types of decision making problems.

In this chapter we model the evaluation process as a Multi-Expert Decision Making (MEDM) problem. In this type of decision problem, decision makers express their opinions about a set alternatives, in order to facilitate the selection of the best one(s). A classical decision analysis scheme is composed by the following phases (see Fig. 1):

- *Identify decision and objectives.*
- *Identify alternatives.*
- *Model:* For example, a decision problem is modelled as a MEDM [18] model that deals with a type of information.
- *Gathering information:* decision makers provide their information.
- *Rating alternatives:* This phase is also known as "aggregation phase" [28] due to the fact in this phase, the individual preferences are aggregated in order to obtain a collective value for each alternative.
- *Choosing best alternatives:* or "exploitation phase" [28] selects the solution from the set of alternatives applying a choice degree [3, 26] to the collective values computed in the previous phase.
- *Sensitive analysis:* in this step the information obtained is analyzed in order to know if it is good enough to make a decision, or otherwise, to go back to initial phases to improve the quantity or/and the quality of the information obtained.
- *Make a decision.*

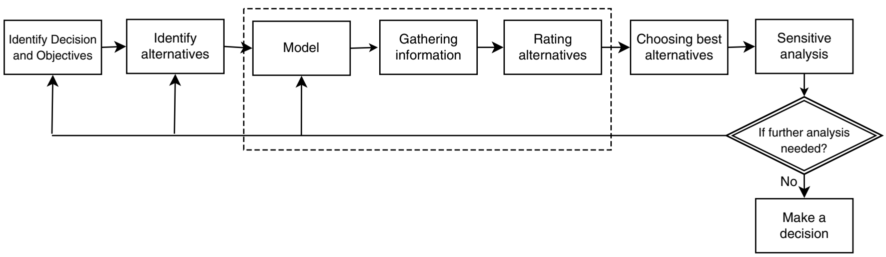



**Fig. 1** Decision analysis scheme

The application of the decision analysis to an evaluation process does not imply the eight phases. The essential phases regarding an evaluation problem are dashed in a rectangle of the Fig. 1.

Additionally the use of *Linguistic* information adds two processes in the **model** and **rating** phases, such as:

1. *The choice of the syntax and semantics of the linguistic terms* that the experts will use to express their assessments about an evaluated element.
2. *To select a linguistic computational technique for rating alternatives* in order to deal with the assessments provided by the experts.

These processes are fixed regarding our proposal in the next subsections.

## *2.2 Fuzzy Linguistic Approach*

Although we usually work in quantitative settings where the information is expressed by numerical values, sometimes we shall need to describe activities of the real world that cannot be assessed in a quantitative form, but rather in a qualitative one, i.e., with vague or imprecise knowledge. In that case, a better approach may be to use linguistic assessments instead of numerical values. The variables which participate in these problems are assessed by means of linguistic terms [34]. This approach is adequate in situations where the information may be unquantifiable due to its nature, and thus, it may be stated only in linguistic terms (e.g., when evaluating the "comfort" or "design" of a car, terms like "bad", "poor", "tolerable", "average", "good" can be used [23]. For instance, when attempting to qualify phenomena related to human perception, such as in sensory evaluation, we are often led to use words in natural language.

Even though, the linguistic approach is less precise than the numerical one, it provides some advantages as, the linguistic assessments are better understood by human beings than numerical ones or that with this approach we also diminished the effects of noise since, as it is known the more refined assessment scale is, the more sensitive to noise and consequently the more error facedown it becomes.

In short, the linguistic approach is appropriated for many problems, since it allows a more direct and adequate representation when we are unable to express it with precision. Hence, the burden of qualifying a qualitative concept is eliminated.

The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables:

**Definition 1.** [34]. *A linguistic variable is characterized by a quintuple (H,T(H),U, G,M) in which H is the name of the variable; T(H) (or simply T) denotes the term set of H, i.e., the set of names of linguistic values of H, with each value being a fuzzy variable denoted generically by X and ranging across a universe of discourse U which is associated with the base variable u; G is a syntactic rule (which usually takes the form of a grammar) for generating the names of values of H; and M is a semantic rule for associating its meaning with each H, $M(X)$, which is a fuzzy subset of U.*

We have to choose the appropriate linguistic descriptors for the term set and their semantics. In order to accomplish this objective, an important aspect to analyze is the "*granularity of uncertainty*", i.e., the level of discrimination among different counts of uncertainty. The universe of the discourse over which the term set is defined can be arbitrary, in this paper we shall use linguistic term sets in the interval [0, 1]. In [4] the use of term sets with an odd cardinal was studied, representing the mid term by an assessment of "approximately 0.5", with the rest of the terms being placed symmetrically around it and with typical values of cardinality , such as 7 or 9.

One possibility of generating the linguistic term set consists of directly supplying the term set by considering all terms distributed on scale on which total order is defined [34]. For example, a set of seven terms $S$, could be given as follows:

$$S = \{s_0 : none, s_1 : verylow, s_2 : low, s_3 : medium, s_4 : high, s_5 : veryhigh, s_6 : perfect\}$$

Usually, in these cases, it is required that in the linguistic term set there exist:

1. A negation operator Neg$(s_i) = s_j$ such that $j = g$-i (g+1 is the cardinality).
2. A max operator: $\max(s_i, s_j) = s_i \; if \; s_i \geq s_j$.
3. A min operator: $\min(s_i, s_j) = s_i \; if \; s_i \leq s_j$

The semantics of the terms is given by fuzzy numbers. A computationally efficient way to characterize a fuzzy number is to use a representation based on parameters of its membership function [4]. The linguistic assessments given by the users are just approximate ones, some authors consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments. The parametric representation is achieved by the 4-tuple $(a, b, d, c)$, where $b$ and $d$ indicate the interval in which the membership value is 1, with $a$ and $c$ indicating the left and right limits of the definition domain of the trapezoidal membership function [4]. A particular case of this type of representation are the linguistic assessments whose membership functions are triangular, i.e., $b = d$, then we represent this type of membership functions by a 3-tuple $(a, b, c)$. An example may be the following:

$$P = (.83, 1, 1) \quad VH = (.67, .83, 1) \quad H = (.5, .67, .83) \quad M = (.33, .5, .67)$$
$$L = (.17, .33, .5) \quad VL = (0, .17, .33) \quad N = (0, 0, .17),$$

which is graphically showed in Fig. 2.

The use of linguistic variables implies processes of computing with words such as their fusion, aggregation, comparison, etc. To perform these computations there are different models in the literature:

- *The linguistic computational model based on the Extension Principle*, which allow us to aggregate and compare linguistic terms through computations on the associated membership functions [8].
- *The symbolic method* [10]. This symbolic model makes direct computations on labels, using the ordinal structure of the linguistic term sets.
- *The 2-tuple fuzzy linguistic computational model* [16]. It uses the 2-tuple fuzzy linguistic representation model and its characteristics to make linguistic computations, obtaining as results linguistic 2-tuples. A linguistic 2-tuple is defined by

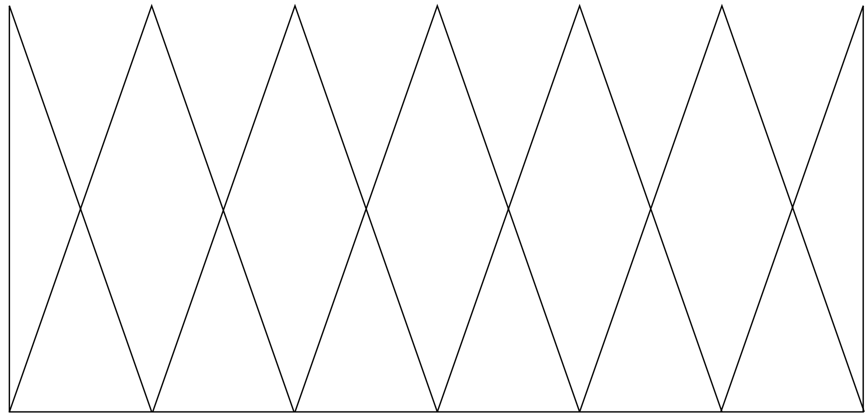



**Fig. 2** A set of seven terms with its semantic

a pair of values, where the first one is a linguistic label and the second one is a real number that represents the value of the symbolic translation.

In the following subsection we shall review the 2-tuple model due to the fact, that it will be the computational model used in our evaluation process.

## *2.3 The 2-Tuple Fuzzy Linguistic Representation Model*

This model has been presented in [16] and has showed itself as useful to deal with evaluation problems similar to the one we are facing in this paper [18, 18].

This linguistic model takes as basis the symbolic aggregation model [10] and in addition defines the concept of Symbolic Translation and uses it to represent the linguistic information by means of a pair of values called linguistic 2-tuple, $(s, \alpha)$, where $s$ is a linguistic term and $\alpha$ is a numeric value representing the symbolic translation.

**Definition 2.** *Let $\beta$ be the result of an aggregation of the indexes of a set of labels assessed in a linguistic term set $S = \{s_0, ..., s_g\}$, i.e., the result of a symbolic aggregation operation. $\beta \in [0, g]$, being $g + 1$ the cardinality of S. Let $i = round(\beta)$ and $\alpha = \beta - i$ be two values, such that, $i \in [0, g]$ and $\alpha \in [-.5, .5)$ then $\alpha$ is called a Symbolic Translation.*

**Definition 3.** [16] *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value supporting the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to $\beta$ is obtained with the following function:*

$$\Delta : [0, g] \longrightarrow S \times [-0.5, 0.5)$$

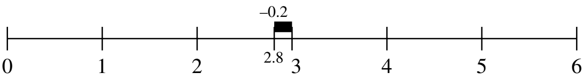



**Fig. 3** Example of symbolic translation

$$\Delta(\beta) = \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-.5, .5) \end{cases} \quad (1)$$

*where $round$ is the usual* round *operation, $s_i$ has the closest index label to "$\beta$" and "$\alpha$" is the value of the symbolic translation.*

*Example 1.* Let's suppose a symbolic aggregation operation over labels assessed in $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ that obtains a result of $\beta = 2.8$, then the representation of this information by means of a 2-tuple will be:

$$\Delta\,(2.8) = (s_3, -0.2)$$

Graphically, it is represented in Fig. 3.

**Proposition 1. [16]** *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and $(s_i, \alpha)$ be a 2-tuple. There is a $\Delta^{-1}$ function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g] \subset \mathcal{R}$.*

**Proof.**

It is trivial, we consider the following function:

$$\begin{aligned} \Delta^{-1} &: S \times [-.5, .5) \longrightarrow [0, g] \\ \Delta^{-1}(s_i, \alpha) &= i + \alpha = \beta \end{aligned} \quad (2)$$

**Remark 1:** From definitions 2 and 3 and from proposition 1, it is obvious that the conversion of a linguistic term into a linguistic 2-tuple consist of adding a value 0 as symbolic translation:

$$s_i \in S \Longrightarrow (s_i, 0)$$

This representation model has associated a computational model that was presented in [16]:

1. **Aggregation of 2-tuples:** The aggregation of linguistic 2-tuples consists of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a linguistic 2-tuple. In [16] we can find several 2-tuple aggregation operators based on classical ones. Here we review the 2-tuple arithmetic mean and the 2-tuple weighted average operators, because we shall use them in our evaluation model:

   **Definition 4.** *Let $x = \{(r_1, \alpha_1), \cdots, (r_n, \alpha_n)\}$ be a set of 2-tuples, the extended Arithmetic Mean $AM^*$ using the linguistic 2-tuples is computed as,*

$$AM^{*}\left(\left(r_{1},\alpha_{1}\right),\ldots,\left(r_{n},\alpha_{n}\right)\right)=\Delta\left(\sum_{i=1}^{n}\frac{1}{n}\Delta^{-1}\left(r_{i},\alpha_{i}\right)\right)=\Delta\left(\frac{1}{n}\sum_{i=1}^{n}\beta_{i}\right) \quad (3)$$

*Example 2.* Let $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ a linguistic term set and $x = \{(s_2, 0.3), (s_5, -0.2), (s_3, 0)\}$ the set of 2-tuples we shall aggregate. The 2-tuple obtained after applying $AM^*$ is:

$$AM^{*}\left(\left(s_{2},0.3\right),\left(s_{5},-0.2\right),\left(s_{3},0\right)\right)=\Delta\left(\frac{1}{3}\sum_{i=1}^{3}\Delta^{-1}\left(r_{i},\alpha_{i}\right)\right)=$$

$$=\Delta\left(\frac{1}{3}\left(2.3+4.8+3\right)\right)=\Delta\left(\frac{1}{3}\cdot 10.1\right)=\Delta\left(3.36\right)=\left(s_{3},0.36\right)$$

**Definition 5.***Let $x = \{(r_1, \alpha_1), \cdots, (r_n, \alpha_n)\}$ be a set of 2-tuples and $W = \{w_1, \cdots, w_n\}$ its associated weights. The 2-tuples weighted mean, $W\_AM^*$, is computed as:*

$$W\_AM^{*}\left(\left(r_{1},\alpha_{1}\right),\ldots,\left(r_{n},\alpha_{n}\right)\right)=\Delta\left(\frac{\sum_{i=1}^{n}\Delta^{-1}\left(r_{i},\alpha_{i}\right)\cdot w_{i}}{\sum_{i=1}^{n}w_{i}}\right)= \quad (4)$$

$$=\Delta\left(\frac{\sum_{i=1}^{n}\beta_{i}\cdot w_{i}}{\sum_{i=1}^{n}w_{i}}\right)$$

*Example 3.* Let $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ a linguistic term set, $x = \{(s_2, 0.3), (s_5, -0.2), (s_3, 0)\}$ the set of 2-tuples we shall aggregate and $w = \{.0.2, 0.3, 0.5\}$ the associated weights. The 2-tuple obtained after applying $W\_AM^*$ is:

$$W\_AM^{*}\left(\left(s_{2},0.3\right),\left(s_{5},-0.2\right),\left(s_{3},0\right)\right)=\Delta\left(\frac{\sum_{i=1}^{3}\Delta^{-1}\left(r_{i},\alpha_{i}\right)\cdot w_{i}}{\sum_{i=1}^{3}w_{i}}\right)=$$

$$\Delta\left(\frac{2.3\cdot 0.2+4.8\cdot 0.3+3\cdot 0.5}{0.2+0.3+0.5}\right)=\Delta\left(3.4\right)=\left(s_{3},0.4\right)$$

More linguistic 2-tuple aggregation operators were defined in [16].

2. **Comparison of 2-tuples:** The comparison of information represented by 2-tuples is carried out according to an ordinary lexico-graphic order.

- if $k < l$ then $(s_k, \alpha_1)$ is smaller than $(s_l, \alpha_2)$
- if $k = l$ then

  a) if $\alpha_1 = \alpha_2$ then $(s_k, \alpha_1)$, $(s_l, \alpha_2)$ represents the same information
  b) if $\alpha_1 < \alpha_2$ then $(s_k, \alpha_1)$ is smaller than $(s_l, \alpha_2)$
  c) if $\alpha_1 > \alpha_2$ then $(s_k, \alpha_1)$ is bigger than $(s_l, \alpha_2)$

3. **Negation Operator of a 2-tuple:** The negation operator over 2-tuples is defined as:

$$Neg\,(s_i, \alpha) = \Delta\left(g - \Delta^{-1}(s_i, \alpha)\right) \tag{5}$$

here $g + 1$ is the cardinality of $S$, $s_i \in S = \{s_0, \ldots, s_g\}$.

## 3 Linguistic Sensory Evaluation Model Based on Decision Analysis

We must keep in mind that the evaluation is used to measure, analyze and interpret the characteristics of the evaluated items according to the knowledge provided by a panel of experts. Classical evaluation methods need to define and know these requirements in an accurate way. However, in sensory evaluation problems the information provided by the experts has been perceived by the senses of sight, touch, smell, taste and hearing, and therefore, those requirements are subjective and involves uncertainty, vagueness and imprecision.

Our aim is to propose a Sensory Evaluation model based on the linguistic decision analysis whose mathematical formalism will be the linguistic 2-tuple model that improves the modelling of the uncertain information provided by the experts and improves the mathematical formalism to operate with this type of information in order to obtain accurate and reliable evaluation results. This proposal consists of the following evaluation phases that are graphically showed in Fig. 4.

- *Identify Evaluated Objects*. This phase is not formalized in this chapter because it is problem-dependent and each problem identifies its objects of interest.
- *Model:* this phase defines the evaluation framework that establishes the evaluation context in which the information is assessed and the problem solved.
- *Gathering information:* the experts express their sensory knowledge about the objects by means of linguistic assessments.
- *Rating objects:* we propose to use the 2-tuple computational model to obtain a rate for every object. In order to accomplish this step, suitable aggregation operators must be chosen.
- *Evaluation results:* it consists of analyzing the results obtained in the previous phase with the purpose of achieving the evaluation process. These results can be used in different ways, such as:

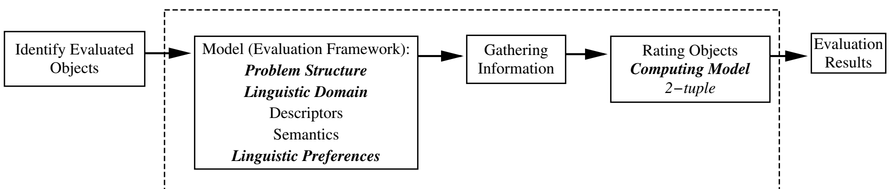



**Fig. 4** A linguistic sensory evaluation scheme based on decision analysis

- To learn which element is better considered by the experts.
- To obtain a global value of an object that can be rated in a product scale to know its quality within its area.
- To know which features are better in the evaluated element.
- To compare several elements to study which aspects make better one element than another.
- To identify which aspects of an element should be improve in order to enhance its quality.
- Etc.

In the following subsections we shall present in further detail the main phases of our sensory evaluation model.

## *3.1 Model*

This phase models the evaluation problem defining its evaluation framework, such that, the problem structure is defined and the linguistic descriptors and semantics that will be used by the experts to provide the information about the sensory features of the evaluated objects are chosen.

First of all, we must analyze which sensory features will be evaluated that depend on the evaluated object and which linguistic term set will be used to assess those features. The linguistic term set will be chosen according to:

1. *The accuracy of the evaluations:* since our senses could recognize and assess some features better than others, the granularity of the linguistic term set that describe those features must be chosen according to the accuracy of our perceptions.
2. *The experience of the experts:* Some of the senses need to be trained and, besides, they usually evolve as much as we used them. Therefore,the granularity of the linguistic term set used by an expert should be also chosen according to the expert's experience.

In this chapter we deal with an evaluation framework such that the different experts provide their sensory perceptions about item features by means of a linguistic label assessed in a fixed term set according to the above conditions. In such a case all the experts provide their sensory subjective preferences using one linguistic term set.

$E = \{e_1, ..., e_n\}$, group of experts

$S = \{s_0, ...., s_g\}$, linguistic term set

$e_i$ expresses his/her preferences in $S$

over a group of sensory features $F = \{f_1, ..., f_h\}$

for a set of items, $X = \{x_1, ..., x_m\}$.

This context facilitates the computational processes of the evaluation problem because it is easy to define for the experts.

### 3.2 Gathering information

Due to the fact that the linguistic decision analysis used in this paper is based on the MEDM problems the experts provide their knowledge by means of utility vectors that contain a linguistic assessment for each evaluated feature.

$\{e_1, ..., e_n\}$, group of experts

$O = \{o_1, ...., o_m\}$, set of evaluated objects

$F = \{f_1, ...., f_h\}$, set of evaluated features for each object

$S = \{s_0, ...., s_g\}$, Linguistic term set

$e_i$ provides his/her preferences in $S$ by means of a utility vector:

$$U_i = \{u^i_{11}, ...., u^i_{1h}, u^i_{21}, ..., u^i_{2h}, ..., u^i_{m1}, ..., u^i_{mh}\}$$

where $u^i_{jk} \in S$ is the assessment provided to the feature $f_k$ of the object $o_j$ by the expert $e_i$.

Consequently in the gathering process every expert $e_i$ will provide his/her utility vector $U_i$ expressed by linguistic labels in the linguistic term set $S$ fixed in the evaluation framework. Due to the fact that the evaluation model will use the linguistic 2-tuple computational model, the linguistic preferences provided by the experts will be transformed into linguistic 2-tuples according to the *Remark 1*.

### 3.3 Rating objects

In this phase the linguistic utility vectors provided by the experts and transformed into linguistic 2-tuples will be used in processes of Computing with Words in order to rate each evaluated object. To do so, the information gathered will be aggregated. Depending on the evaluation problem can be used different types of aggregation operators:

1. *Linguistic Aggregation operators of Non-Weighted information*. These operators aggregate the linguistic information provided by different sources with equal importance, i.e., all sources are equally important in the aggregation process. Examples of linguistic aggregation operators of non-weighted information can be found in [10, 31].
2. *Linguistic Aggregation operators of Weighted information*. These operators aggregate the information provided by different sources which are not equally important. Different proposals of this type of operators have been proposed in [5, 34].

Keeping in mind that the aim of this proposal is the use of a consistent mathematical formalism, as it is the linguistic 2-tuple computational model, to operate with the uncertain information provided by the experts it must be remarked that several aggregation operators of both types have been introduced for this linguistic computational model [16].

The rating process of this proposal consists of two steps:

1. *Computing collective evaluations for each feature*: in the gathering process each expert, $e_i$ provides his/her preferences for every feature $f_k$ of the object $o_j$ by means of a utility assessment, $u^i_{jk}$. Then, the rating process in first place will compute a collective value for each feature, $u_{jk}$, using an aggregation operator, $AG$, on the assessments provided by the experts:

$$u_{jk} = AG(u^1_{jk}, ...., u^n_{jk}) \tag{6}$$

2. *Computing a collective evaluation for each object*: the final aim of the rating process is to obtain a global evaluation, $u_j$, of each evaluated object according to all the experts and features that take part in the evaluation process. To do so, this process will aggregate the collective features values $u_{jk}$ for each object, $o_j$:

$$u_j = AG(u_{j1}, ...., u_{jh}) \tag{7}$$

The aggregation operators will depend on each evaluation problem taking into account if all experts or features are equally important or there are experts or features more important than the others.

The collective evaluation obtained will be the score obtained by the evaluated object in the sensory evaluation problem.

## 4 Evaluating Different Samples of Olive Oil to Obtain a Particular Flavor

Nowadays, the quality of the olive oil plays a key role in its production and final price. This quality depends on several aspects such as the condition of olives when enter the factory, the extraction processes and their sedimentation, or their storage.

The evaluation of the quality of the olive oil is not an easy task and is usually accomplished by Olive Oil Tasting Panel, in which there are between 8 and 12 connoisseurs, which will evaluate, by means of their perceptions acquired via their senses, the features that describe the samples of Olive Oil.

The combination of smell and taste is known as flavor and defines the organoleptic properties of the olive oil. So, we could talk about an olive oil with apple scent and sweet taste or an olive oil slightly pungent with almond scent.

These organoleptic properties, with acidity grade of the olive oil, are essential to obtain their quality. The acidity grade measures the level of free fatty acid, and therefore, an olive oil with a high acidity grade has more free fatty acid and is less healthy than an olive oil with a low acidity grade. Both aspects, the organoleptic properties and the acidity grade, establish the quality of the olive oil.

While it is easy to obtain the acidity grade of a sample of olive oil by means of chemical processes, the organoleptic properties need to be evaluated by the Tasting Panel that will use their perceptions to catch different aspects of its flavor such as fruity, bitter, pungent, etc. Besides, we must realize that although the most usual way utilized to express these perceptions is by means of numerical values, it is not the most suitable because this information has been acquired by means of perceptions, which usually involves uncertainty, vagueness and imprecision. In *http://www.oliveoilsource.com/tasteform1.pdf* we can find an example of a tasting sheet used by the panels of experts.

The companies in the olive oil market usually need to keep the flavor of their olive oil brands through time because its flavor is an essential characteristic of the brands. However, because it is impossible to obtain the quantity of the same kind of olives for the total production of an olive oil brand, they have to mix batches of olive oil in order to reproduce the same flavor. In these processes, the sensory evaluation plays a critical role because before starting any mixing process they need to know which batch of olive oil is suitable for being mixed, which organoleptic properties need to be improved or which ones need to be diminished. In these example, we shall show an example of how to evaluate four samples of olive oil, in order to find out the values of the organoleptic properties of sweetness and pungency. These values will be used in order to decide which batches should be mixed to obtain the flavor that the company is looking for.

## *Evaluation Framework*

An Olive Oil Tasting Panel of eight connoisseurs $E = \{e_1, ..., e_8\}$ will evaluate the sensory feature *sweetness* of four samples of Olive Oil $O = \{o_1, ...., o_4\}$ and two sensory features $F = \{sweetness, pungency\}$. The panel will evaluate these sensory features independently in order to know the value of these features. To do so, two linguistic term set $S$ and $S'$ of nine terms and seven terms respectively were chosen according to conditions presented in subsection 3.1 to assess the sweetness and pungency respectively. Their syntax and semantics are the following ones (see Figs. 5 and 6).

$s_8 = Very\ sweet = (.88, 1, 1)$ $s_7 = Rather\ sweet = (.75, .88, 1)$
$s_6 = Sweet = (.62, .75, .88)$ $s_5 = A\ bit\ sweet = (.5, .62, .75)$
$s_4 = Average = (.38, .5, .62)$ $s_3 = A\ bit\ bitter = (.25, .38, .5)$
$s_2 = Bitter = (.12, .25, .38)$ $s_1 = Rather\ bitter = (0, .12, .25)$
$s_0 = Very\ bitter = (0, 0, .12)$

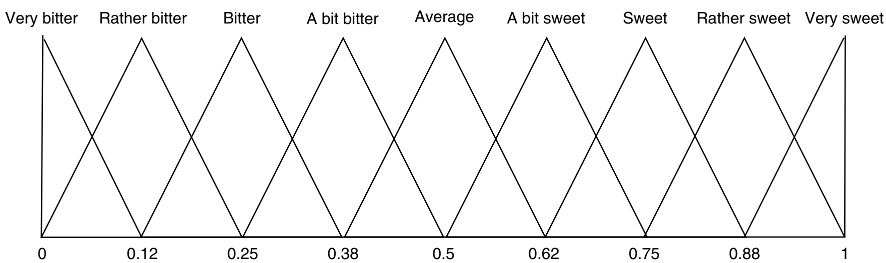



**Fig. 5** A set of nine terms with its semantic chosen to evaluate the *sweetness*

$s'_6 = Very\ pungent = (.83, 1, 1)$ $s'_5 = Pungent = (.66, .83, .1)$
$s'_4 = A\ bit\ pungent = (.5, .66, .83)$ $s'_3 = Average = (.33, .5, .66)$
$s'_2 = A\ bit\ bland = (.17, .33, .5)$ $s'_1 = Bland = (.0, .17, .33)$
$s'_0 = Very\ bland = (0, 0, .17)$

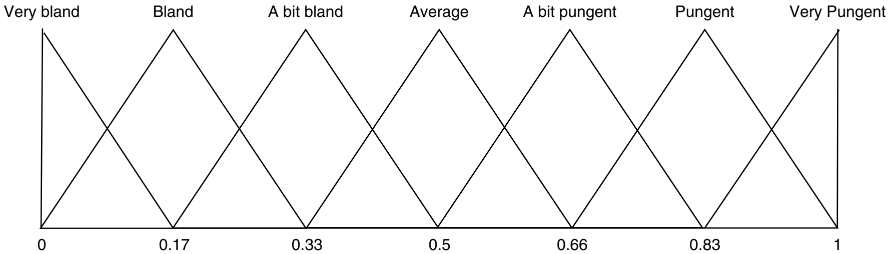



**Fig. 6** A set of nine terms with its semantic chosen to evaluate the *pungency*

## *Gathering Process*

The preferences of our Tasting Panel for sweetness and pungency are showed in Table 1 and Table 2 respectively.

Now, we shall transform their preferences into 2-tuple representation model (Table 3 and Table 4) to manage easily this information.

## *Rating Objects*

In this phase we shall carry out the following steps:

**Table 1** Olive Oil Tasting Panel's utility vectors for the feature *sweetness*

| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ | $e_7$ | $e_8$ |
|---|---|---|---|---|---|---|---|---|
| $o_1$ | $s_4$ | $s_2$ | $s_5$ | $s_3$ | $s_4$ | $s_5$ | $s_2$ | $s_7$ |
| $o_2$ | $s_4$ | $s_3$ | $s_4$ | $s_2$ | $s_2$ | $s_4$ | $s_5$ | $s_3$ |
| $o_3$ | $s_3$ | $s_3$ | $s_5$ | $s_4$ | $s_3$ | $s_2$ | $s_4$ | $s_2$ |
| $o_4$ | $s_5$ | $s_4$ | $s_4$ | $s_5$ | $s_6$ | $s_3$ | $s_7$ | $s_3$ |

**Table 2** Olive Oil Tasting Panel's utility vectors for the feature *pungency*

| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ | $e_7$ | $e_8$ |
|---|---|---|---|---|---|---|---|---|
| $o_1$ | $s'_4$ | $s'_3$ | $s'_4$ | $s'_5$ | $s'_6$ | $s'_4$ | $s'_4$ | $s'_7$ |
| $o_2$ | $s'_5$ | $s'_5$ | $s'_6$ | $s'_3$ | $s'_1$ | $s'_2$ | $s'_5$ | $s'_2$ |
| $o_3$ | $s'_3$ | $s'_4$ | $s'_5$ | $s'_3$ | $s'_2$ | $s'_3$ | $s'_3$ | $s'_3$ |
| $o_4$ | $s'_4$ | $s'_3$ | $s'_5$ | $s'_4$ | $s'_5$ | $s'_4$ | $s'_7$ | $s'_2$ |

**Table 3** Olive Oil Tasting Panel's utility vectors for the feature *sweetness* over the 2-tuple representation model

| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ | $e_7$ | $e_8$ |
|---|---|---|---|---|---|---|---|---|
| $o_1$ | $(s_3, 0)$ | $(s_3, 0)$ | $(s_6, 0)$ | $(s_4, 0)$ | $(s_6, 0)$ | $(s_6, 0)$ | $(s_4, 0)$ | $(s_7, 0)$ |
| $o_2$ | $(s_4, 0)$ | $(s_3, 0)$ | $(s_4, 0)$ | $(s_2, 0)$ | $(s_2, 0)$ | $(s_4, 0)$ | $(s_5, 0)$ | $(s_3, 0)$ |
| $o_3$ | $(s_3, 0)$ | $(s_3, 0)$ | $(s_5, 0)$ | $(s_4, 0)$ | $(s_3, 0)$ | $(s_3, 0)$ | $(s_4, 0)$ | $(s_2, 0)$ |
| $o_4$ | $(s_4, 0)$ | $(s_3, 0)$ | $(s_4, 0)$ | $(s_4, 0)$ | $(s_5, 0)$ | $(s_3, 0)$ | $(s_7, 0)$ | $(s_3, 0)$ |

**Table 4** Olive Oil Tasting Panel's utility vectors for the feature *pungency* over the 2-tuple representation model

| | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ | $e_7$ | $e_8$ |
|---|---|---|---|---|---|---|---|---|
| $o_1$ | $(s'_4, 0)$ | $(s'_3, 0)$ | $(s'_4, 0)$ | $(s'_5, 0)$ | $(s'_6, 0)$ | $(s'_4, 0)$ | $(s'_4, 0)$ | $(s'_7, 0)$ |
| $o_2$ | $(s'_5, 0)$ | $(s'_5, 0)$ | $(s'_6, 0)$ | $(s'_3, 0)$ | $(s'_1, 0)$ | $(s'_2, 0)$ | $(s'_5, 0)$ | $(s'_2, 0)$ |
| $o_3$ | $(s'_3, 0)$ | $(s'_4, 0)$ | $(s'_5, 0)$ | $(s'_3, 0)$ | $(s'_2, 0)$ | $(s'_3, 0)$ | $(s'_3, 0)$ | $(s'_3, 0)$ |
| $o_4$ | $(s'_4, 0)$ | $(s'_3, 0)$ | $(s'_5, 0)$ | $(s'_4, 0)$ | $(s'_5, 0)$ | $(s'_4, 0)$ | $(s'_7, 0)$ | $(s'_2, 0)$ |

1. *Computing collective values for each feature:* In order to simplify the example we have considered that all the experts are equally important. Therefore, we have used the arithmetic mean for 2-tuples for aggregating the information provided by the experts (Tables 5 and 6) obtaining a collective value for sweetness and pungency for each sample according to all the connoisseurs:
2. *Computing a collective evaluation for each object:* In this example the objective is to obtain the evaluation of different organoleptic features independently of each other to classify the different olive oil batches. So it is not necessary to obtain a global evaluation of each olive batch according to the two properties analyzed. However it is important to point out that if it would be necessary to obtain this global evaluation value we should use an aggregation method able to manage linguistic information assessed in different linguistic term sets as the methods showed in [18, 18].

## Evaluation Results

The purpose of this evaluation process was to find out the values of different samples of olive oil regarding their sweetness and pungency properties . If we analyze the aforesaid results (Tables 5 and 6), the sample $o_1$ obtains the highest score for both

**Table 5** Olive Oil Tasting Panel's collective utility vector for the *sweetness*

| $o_1$ | $o_2$ | $o_3$ | $o_4$ |
|---|---|---|---|
| $(s_5 = A\ b\ sw, -.125)$ | $(s_3 = A\ b\ bit, .375)$ | $(s_3 = A\ b\ bit, .375)$ | $(s_4 = Av, .25)$ |

**Table 6** Olive Oil Tasting Panel's collective utility vector for the *pungency*

| $o_1$ | $o_2$ | $o_3$ | $o_4$ |
|---|---|---|---|
| $(s'_5 = Pungent, -.375)$ | $(s'_4 = A\ b\ Pun, -.375)$ | $(s'_3 = Av, .25)$ | $(s'_4 = A\ b\ Pun., .25)$ |

features. The first one, the sweetness, is assessed with *A bit sweet* and therefore it is above the average. The second one, its pungency, is *Pungent* and it is above the average as well.

## 5 Concluding Remarks

When we face a sensory evaluation problem we must realize that we are going to work with knowledge that has been acquired via the human senses sight, taste, touch, smell and hearing. This knowledge is better expressed using words instead of numbers, because humans cannot measure exactly with their senses and words gather accurately the uncertainty related to this way of acquisition of knowledge.

In this paper, we have proposed a sensory evaluation model based on the linguistic decision analysis since it has been applied successfully to similar evaluation problems and we have used the 2-tuple computational model in order to exploit the information because of the advantages that 2-tuple model offers regarding other linguistic computational models.

Finally we have showed an example of how to apply this model to a specific sensory evaluation problem, the evaluation of virgin olive oil, in order to expose the advantages of its use.

**Acknowledgment** This work is supported by Research Project MTM2005-08982-C04-03.

## References

1. A.M. Allison and T. Work. Fiery and frosty foods pose challenges in sensory evaluation. *Food Technology*, 58(5):32–37, 2004.
2. J. Antes, L. Campen, U. Derigs, C Titze, and G.D. Wolle. A model-based decision support system for the evaluation of flight schedules for cargo airlines. *Decision Support Systems*, 22(4):307–323, 1998.
3. K.J. Arrow. *Social Choice and Individual Values*. Yale University Press, New Haven CT, 1963.
4. P.P. Bonissone and K.S. Decker. *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity*. In L.H. Kanal and J.F. Lemmer, Editors., Uncertainty in Artificial Intelligence. North-Holland, 1986.
5. G. Bordogna, M. Fedrizzi, and G. Pasi. A linguistic modeling of consensus in group decision making based on OWA operators. *IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27:126–132, 1997.
6. C.T. Chen. Applying linguistic decision-making method to deal with service quality evaluation problems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(Suppl.):103–114, 2001.
7. G.V. Civille and J. Seltsam. Sensory evaluation methods applied to sound quality. *Noise Control Engineering Journal*, 51(4):262–270, 2003.
8. Bouyssou D., Marchant T., Pirlot M., Perny P., and Tsoukia‘s A. *Evaluation and Decision Models: A critical perspective*. Kluwer Academic Publishers, 2000.

9. R. Degani and G. Bortolan. The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning*, 2:143–162, 1988.
10. M. Delgado, J.L. Verdegay, and M.A Vila. On aggregation operations of linguistic labels. *International Journal of Intelligent Systems*, 8:351–370, 1993.
11. G.B. Dijksterhuis. *Multivariate Data Analysis in Sensory and Consumer Science, Food and Nutrition*. Press Inc. Trumbull, Connecticut, USA, 1997.
12. P. Dillon, W. Moody, R. Bartlett, P. Scully, R. Morgan, and C. James. Sensing the fabric: To simulate sensation through sensory evaluation and in response to standard acceptable properties of specific materials when viewed as a digital image. *Proceedings Lecture Notes In Computer Science*, 2058:205–217, 2001.
13. C.R. Han, C. Lederer, M. McDaniel, and Y.Y. Zhao. Sensory evaluation of fresh strawberries (fragaria ananassa) coated with chitosan-based edible coatings. *Journal Of Food Science*, 70(3):172–178, 2005.
14. F. Herrera and E. Herrera-Viedma. Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 115:67–82, 2000.
15. F. Herrera, E. Herrera-Viedma, L. Martinez, F. Mata, and P.J. Sanchez. *A Multi-Granular Linguistic Decision Model for Evaluating the Quality of Network Services*. Intelligent Sensory Evaluation: Methodologies and Applications. Springer, Ruan Da, Zeng Xianyi (Eds.), 2004.
16. F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.
17. F. Herrera and L. Martínez. A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 31(2):227–234, 2001.
18. F. Herrera, L. Martínez, and P.J. Sánchez. Managing non-homogeneous information in group decision making. *European Journal of Operational Research*, 166(1):115–132, 2005.
19. A. Jiménez, S. Ríos-Insua, and A. Mateos. A decision support system for multiattribute utility evaluation based on imprecise assignments. *Decision Support Systems*, 36(1):65–79, 2003.
20. J. Kacprzyk and S. Zadrozny. Computing with words in decision making: Through individual and collective linguistic choice rules. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(Suppl.):89–102, 2001.
21. J. Lawry. A methodology for computing with words. *Internation Journal of Approximate Reasoning*, 28:51–89, 2001.
22. HS Lee and M. O'Mahony. Sensory evaluation and marketing: measurement of a consumer concept. *Food Quality And Preference*, 16(3):227–235, 2005.
23. E. Levrat, A. Voisin, S. Bombardier, and J. Bremont. Subjective evaluation of car seat comfort with fuzzy set techniques. *International Journal of Intelligent Systems*, 12:891–913, 1997.
24. C.M. Liu, M.J. Wang, and Y.S. Pang. A multiple criteria linguistic decision model (mcldm) for human decision making. *European Journal of Operational Reseach*, (76):466–485, 1994.
25. A.C. Marquez and C. Blanchar. A decision support system for evaluating operations investments in high-technology business. *Decision Support Systems*, 41(2):472–487, 2006.
26. S.A. Orlovsky. Decision-making with a fuzzy preference relation. *Fuzzy Sets Systems*, 1:155–167, 1978.
27. J. Pearce. Sensory evaluation in marketing. *Food Technology*, 34(11):60–62, 1980.
28. M. Roubens. Fuzzy sets and decision analysis. *Fuzzy Sets and Systems*, 90:199–206, 1997.
29. D. Ruan and X. Zeng (Eds.). *Sensory Evaluation: Methodologies and Applications*. Springer, 2004.
30. H. Stone and J.L. Sidel. *Sensory Evaluation Practice*. Academic Press Inc., San Diego, CA, 1993.
31. V. Torra. The weighted OWA operator. *International Journal of Intelligent Syatems*, 12:153–166, 1997.
32. J. Wang, J.B. Yang, and P. Sen. Multi-person and multi-attribute design evaluations using evidential reasoning based on subjective safety and cost analyses. *Reliability Engineering and System Safety*, 52(2):113–129, 1996.
33. P.P. Wang(Ed.). *Computing with Words*. Wiley Series on Intelligent Systems. John Wiley and Sons, 2001.

34. R.R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, 12:237–261, 1995.
35. L.A. Zadeh. The concept of a linguistic variable and its applications to approximate reasoning. *Information Sciences, Part I, II, III*, 8,8,9:199–249,301–357,43–80, 1975.
36. L.A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111, May 1996.
37. X. Zeng, Y. Ding, and L. Koehl. *Intelligent Sensory evaluation: Methodologies and Applications*, chapter A 2-tuple Fuzzy linguistic model for sensory fabric hand evaluation, pages 217–234. Springer, 2005.

# Atanassov's Intuitionistic Fuzzy Sets as a Promising Tool for Extended Fuzzy Decision Making Models

**Eulalia Szmidt and Janusz Kacprzyk**

**Abstract** Since decision making is omnipresent in any human activity, it is quite clear that not much later after the concept of a fuzzy set was introduced as a tool for a description and handling of imprecise concepts, a next rational step was an attempt to devise a general framework for dealing with decision making under fuzziness. Since intuitionistic fuzzy sets (in the sense of Atanassov, to be called A-IFSs, for short) provide a richer apparatus to grasp imprecision than the conventional fuzzy sets, they seem to be a promising tool for extended decision making models. We will present some of the extended models and try to show why A-IFSs make it possible to avoid some more common cognitive biases, the decision makers are prone to do, which call into question the correctness of a decision.

## 1 Introduction

Decision making is one of key "meta-problems" in science, hence it has been a subject of an extensive research effort for decades or even centuries that has resulted in a multitude of models and approaches with roots in diverse areas ranging from social sciences, through cognitive science, mathematical sciences to information technology.

One of main difficulties plaguing the conventional formal models is that they are to be based on exact and certain information. They have been extended to deal with uncertainty, and finally - to capture imprecise descriptions, relations, etc. - fuzzy models have been proposed, based on Zadeh's theory of fuzzy sets.

As many extensions to the conventional theory of fuzzy sets have appeared over the years, traditional fuzzy decision making models have been extended to include those extended fuzzy type descriptions.

Among those extensions, Atanassov's theory of intuitionistic fuzzy sets [5] plays an important role, and has gained popularity in recent [1]. Basically, it introduces, for

[1] Recently there is a debate on the suitability of the name "intuitionistic fuzzy sets" (cf. Dubois, Gottwald, Hájek, Kacprzyk, Prade [13] but this is beyond the scope of this paper and we will not be dealing with this). To maintain the consistency with the convention adopted in this volume we will call the intuitionistic fuzzy sets as introduced by Atanassov, A-IFSs, for short).

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008

each element of a universe of discourse, a degree of membership and a degree of non-membership, both from [0, 1], but which do not sum up to 1 as in the conventional fuzzy sets. Such an extended definition can help more adequately represent situations when, for instance, decision makers abstain from expressing their testimonies, some assessments can not be classified but also can not be discarded, etc. Therefore, A-IFSs provide a richer apparatus to grasp imprecision than the conventional fuzzy sets.

In this paper we will show how one can use elements of the theory of A-IFSs to extend some basic traditional fuzzy decision making models that have been proposed so far.

We start with a review of basic elements of the theory of A-IFSs, in Atanassov's sense but presented in a way limiting our attention to elements to be needed later on. We will emphasize the features of A-IFSs which make it possible to avoid some cognitive biases that can make traditional decision making procedures questionable or make their results of a limited usefulness.

Basically, since the purpose of this article is primarily to set foundations of a wide array of A-IFSs based decision making problems that could be derived as extensions of the traditional fuzzy sets based ones, we will primarily discuss the classic Bellman and Zadeh's [7] general framework of decision making in a fuzzy environment. It is a point of departure for almost all fuzzy decision making, optimization, control, etc. models.

Then, we will consider the use of A-IFSs to represent imprecise preferences by referring to some seminal works by Zadeh, Fodor, De Baets, Roubens, etc. These preference representations will then be used to extend some classic group decision making models under fuzzy preferences and a fuzzy majority proposed by Kacprzyk, Fedrizzi, Nurmi, etc. A related issue of how to measure a degree of consensus in a group of experts will also be dealt with.

In the above analysis we will take as points of departure both the A-IF extensions of various elements of those models (e.g. preferences), and some derived characteristic features that result from the concept of a similarity of (distance between) A-IFSs.

Throughout the paper we will emphasize proper interpretations of A-IF elements to show that they can provide a new quality that cannot be attained by using conventional fuzzy sets.

We will finally indicate some more promising areas for a further research

## 2 A-IFS Theory (Atanassov [3, 5])

Let us start with basic concepts related to fuzzy sets.

**Definition 1.** *A fuzzy set $A'$ in $X = \{x\}$ (where $x$ stands for a generic element of $X$) is given by (cf. Zadeh [63]):*

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

*where* $\mu_A : X \to [0,1]$ *is the membership function of the fuzzy set* $A'$; $\mu_A(x) \in [0,1]$.

The theory of A-IFSs is based both on extensions of corresponding definitions of fuzzy sets objects and definitions of new objects and their properties (Atanassov [3, 4, 5]).

**Definition 2.** *An A-IFS A in X is given by (Atanassov [3, 5]):*

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

*where*

$$\mu_A : X \to [0,1]$$

$$\nu_A : X \to [0,1]$$

*with the condition*

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \quad \forall x \in X$$

*The numbers* $\mu_A(x)$, $\nu_A(x) \in [0,1]$ *denote a degree of membership and non-membership of x to A, respectively.*

Obviously, each fuzzy set $A'$ corresponds to the following A-IFS:

$$A = \{< x, \mu_A(x), 1 - \mu_A(x) > | x \in X\} \tag{3}$$

For each A-IFS in $X$, we will call

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

the intuitionistic fuzzy index (or a hesitation margin) of $x$ in $A$. The intuitionistic fuzzy index expresses a lack of knowledge of whether $x$ belongs to $A$ or not (Atanassov [4, 5]).

It is obvious that

$$0 \leq \pi_A(x) \leq 1 \quad for \quad each \quad x \in X$$

For each fuzzy set $A'$ in $X$, evidently,

$$\pi_A(x) = 1 - \mu_A(x) - [1 - \mu_A(x)] = 0, \quad for \quad each \quad x \in X$$

For definitions of basic concepts, operations, etc. on A-IFSs, cf. Atanassov [5].

The application of A-IFSs instead of fuzzy sets means the introduction of another dimension into a set description. Such a generalization of fuzzy sets gives us an additional possibility to represent imperfect knowledge that may lead to the description of many real problems in a more adequate way.

A-IFSs based models may be adequate mainly in the situations when we face human testimonies, opinions, etc. involving answers of three types:

- yes,
- no,
- abstaining, i.e. which can not be classified (because of different reasons, eg. “I do not know”, “I am not sure”, “I do not want to answer”, “I am not satisfied with any of the options” etc.).

Voting can be a good example as the human voters may be divided into three groups of those who: vote for, vote against and abstain. For some analysis of the very nature of abstention, which is of a key importance for our purposes, see Kang [28].

Applications of A-IFSs to group decision making, negotiations, etc. are presented in (Szmidt and Kacprzyk [42, 43, 45, 48, 52, 53, 54, 59]). Albeit voting, group decision making or negotiations are the representative examples in which inherent features of the A-IF models are especially desirable, also in other situations when a decision is to be made, A-IFSs may show their power that is a result of the two reasons:

- A-IFSs make a decision maker concentrate his/her attention not only on advantages (expressed via memberships of the considered options) but also on disadvantages of the options. This feature of A-IFSs should be stressed because one of cognitive and personal biases in decision making pointed out by psychologists is wishful thinking and optimism – people tend to want to see things in a positive light (neglecting the negatives of the options) what can distort their perception and thinking (Sutherland [39]). A-IF models just at the stage of collecting the data prevent from making the mistake.
- A-IFSs make it possible to additionally (to fuzzy sets) capture a lack of knowledge on an imprecise description itself, e.g. a lack of knowledge concerning the considered options. Clearly, a similar facility can be attributed to the use of interval-valued fuzzy sets (Atanassov and Gargov first noticed in 1989 that the both types of the sets are equipollent [6]) but, from the cognitive bias related point of view assumed here, one can say that interval-valued fuzzy sets, in a sense, just support the cognitive bias mentioned above – a decision maker considers only the membership values (representing advantages) of the considered options. In effect only a lack of knowledge on membership values is taken into account. Non-membership values of the options (representing disadvantages) are not considered by a decision maker. Of course, formally non-membership values can be calculated but for a decision maker they are somewhere at the background and do not need to be considered. In effect non-membership values are not considered – as the psychologists noticed, people tend to be willing to gather facts that support certain conclusions but disregard other facts that support different conclusions (Sutherland [39]). The very essence and structure of A-IF decision models may protect us against the cognitive bias.

Clearly, a question arises how to assign the membership and non-membership values. This is a problem as old as fuzzy sets theory itself, and many procedures have been proposed. For A-IFSs, an approach based on questionnaires and histograms can be found in Szmidt and Baldwin [40, 41].

# 3 Decision Making Under Fuzziness – Bellman and Zadeh's Approach

In this section we briefly present the seminal Bellman and Zadeh general approach to decision making under fuzziness, originally termed *decision making in a fuzzy environment*, a simple yet extremely powerful framework within which virtually all fuzzy models related to decision making, optimization and control have been dealt with. Next, imprecision (fuzziness) of the environment within which the decision making (or control) process proceeds is enlarged by using A-IFSs having additional degree of freedom. In effect A-IF environment (intuitionistic fuzziness), which consists of A-IF goals, A-IF constraints, and A-IF decisions, is introduced and a general approach to decision making under such condition is presented.

## *3.1 Decision Making in a Fuzzy Environment*

In Bellman and Zadeh's [7] setting the imprecision (fuzziness) of the environment within which the decision making (or control) process proceeds is modeled by the introduction of the so called *fuzzy environment* which consists of fuzzy goals, fuzzy constrains, and fuzzy decision.

A formal definition of these elements of the fuzzy environment starts with the assumption of some set of possible *options* (or alternatives, choices decisions) denoted by $X = \{x\}$ where $x$ means a generic element of $X$. The set $X$ contains all the possible (relevant, feasible) values, courses of actions, etc. in the situation considered.

The *fuzzy goal* is now defined as a fuzzy set $G$ in the set of options $X = \{x\}$, characterized by its membership function $\mu_G : X \to [0, 1]$ such that $\mu_G(x) \in [0, 1]$ specifies the grade of membership of a particular option $x \in X$ in $G$.

The *fuzzy constraint* is similarly defined as a fuzzy set $C$ in the set of options $X = \{x\}$, characterized by its membership function $\mu_C : X \to [0, 1]$ such that $\mu_C(x) \in [0, 1]$ specifies the grade of membership of a particular option $x \in X$ in $C$.

The above mentioned identity of handling a fuzzy goal $G$ in $X = \{x\}$ and a fuzzy constraint $C$ in $X = \{x\}$ suggests the following general formulation of the decision making problem in a fuzzy environment:

$$Attain\ G\ \underline{and}\ satisfy\ C \tag{5}$$

which should be meant as to determine a decision (an option or a set of options) which simultaneously fulfills the fuzzy goal and fuzzy constraint; evidently such a decision should belong to those available, or perhaps to those relevant or feasible.

In a formal way, if $G$ is a fuzzy goal and $C$ is a fuzzy constraint, both defined as fuzzy sets in the set of options $X = \{x\}$, the *fuzzy decision* $D$ is a fuzzy set defined also in the set of options $X$ resulting from an aggregation $* : [0, 1] \times [0, 1] \rightarrow [0, 1]$ of $G$ and $C$, that is

$$D = G * C \tag{6}$$

or, in terms of membership functions,

$$\mu_D(x) = \mu_G(x) * \mu_C(x) \quad for\ each \quad x \in X \tag{7}$$

The aggregation "*" is evidently some operation on two fuzzy sets. Therefore, there immediately arises a question as to which one is appropriate. Basically, if we take as a point of departure the general formulation of the decision making problem, i.e. "attain $G$ and satisfy $C$", this operation should correspond to the "and" connective in the definition of the intersection of two fuzzy sets (cf. Kacprzyk [20]).

Let us start with the most important type of the fuzzy decision that is related to Zadeh's standard definition [63] of the intersection of two fuzzy sets.

The *min-type fuzzy decision* is defined as

$$D = G \cap C \tag{8}$$

or, in terms of the membership functions,

$$\mu_D(x) = \mu_G(x) \wedge \mu_C(x) \quad \text{for each } x \in X \tag{9}$$

where "$\wedge$" is the minimum operation, i.e. $a \wedge b = min(a, b)$.

*Example 1.* (cf. Kacprzyk, [20]) Suppose that the fuzzy goal is " $x$ should be much larger than 5", and the fuzzy constraint $C$ is "$x$ should be about 6", as in Fig. 1.

The (membership function of) min-type fuzzy decision is given in bold line, and should be interpreted as follows. The set of possible options is the interval [5, 10]

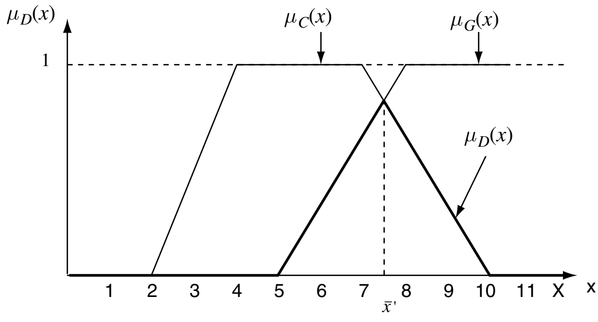



**Fig. 1** Fuzzy goal, fuzzy constraint, fuzzy decision, and the optimal (maximizing) decision

because $\mu_D > 0$, for $5 \leq x \leq 10$. The other options, i.e. $x < 5$ and $x > 10$ are impossible since $\mu_D(x) = 0$. However, not all the options belonging to the interval [5, 10] are equally satisfactory (or preferable). The value of $\mu_D(x) \in [0, 1]$ may be meant as the degree of satisfaction from the choice of a particular $x \in X$, from 0 for full dissatisfaction to 1 for full satisfaction, through all intermediate values; thus, the higher the value of $\mu_D(x)$, the higher the satisfaction from $x$.

We have therefore a concept of a fuzzy decision. The above interpretation of the fuzzy decision's membership function $\mu_D(x)$ as the degree to which a particular option $x$ is satisfactory as a solution to the problem (5) immediately suggests that the best (nonfuzzy) choice in this case would be the one corresponding to the highest value of $\mu_D(x)$.

The *maximizing decision* is defined as an $x^* \in X$ such that

$$\mu_D(x^*) = \max_{x \in X} \mu_D(x) \tag{10}$$

and an example may be found in Fig. 1 where $x^* = 7.5$.

We should, however be aware that the determination of a nonfuzzy (maximizing) decision from a fuzzy decision is basically the problem of defuzzification, and the simple defuzzification of the type (10) is clearly not a perfect solution, and its simplicity is the only advantage.

## 3.2 Decision Making in an A-IF Environment

It is clear that we can expect that A-IFSs, having another degree of freedom in comparison with fuzzy sets, may model decision making processes in a more human consistent and flexible way. So another rational step is an attempt to devise a general framework for dealing with decision making and control under intuitionistic fuzziness. The same way of reasoning as proposed by Bellman and Zadeh in a fuzzy environment is also the core of the approach proposed in this Section but concerns decision making in an A-IF environment.

In the classic Bellman and Zadeh's approach a satisfaction level or any value is given by a membership function $\mu$ whereas when using A-IFSs, a satisfaction level is described by a number from an interval $[\mu, \mu + \pi]$ where $\pi$ is the intuitionistic fuzzy index. It is because our preferences can vary with external conditions. For example, the same queue length in a bank can be assessed as too long to wait in (when the weather is beautiful, we want to go for a walk and there are still several more days ahead to go the bank) or rather short one – when it is the day we must do business in our bank (for more example when A-IFSs are useful for describing real world phenomena, a cf. Szmidt and Kacprzyk [43].

Certainly, the general formulation of the decision making problem in an A-IF environment is the same as in a case of a fuzzy environment, i.e.: "attain $G$ <u>and</u> satisfy $C$". But now both a goal and a constraint are given by two functions, so the formula

$$D = G * C \tag{11}$$

in terms of A-IFSs must be described by two functions.

The counterpart of the min-type fuzzy decision (8) in the case of A-IFSs is defined as

$$D = G \cap C \tag{12}$$

or, in terms of the membership and non-membership functions [3, 5],

$$\mu_D(x) = \mu_G(x) \wedge \mu_C(x) \text{ and } \nu_D(x) = \nu_G(x) \vee \nu_C(x);\ \ for\ each\ \ x \in X \tag{13}$$

The same considerations as for fuzzy sets (cf. Sect. 3) lead to the maximizing decision defined as $x^* \in X$ such that

$$\mu_D(x^*) = \max_{x \in X} \mu_D(x) \text{ and } \nu_D(x^*) = \min_{x \in X} \nu_D(x) \tag{14}$$

Unfortunately, for A-IF cases the (14) have no solution. It can be easily proved that for fuzzy problems only (in a case of linear fuzzy goals and linear fuzzy constraints) an optimal decision $x^*$ exists. For A-IF cases we have

$$\mu_D(\overline{x}) = \max_{x \in X} \mu_D(x) \text{ and } \nu_D(\underline{x}) = \min_{x \in X} \nu_D(x) \tag{15}$$

*Example 2.* Suppose that the fuzzy goal is "$x$ should be much larger than 5", (like in Example 1) but let us also assume that owing to additional (external) conditions, our preferences (goal) vary. The varying preferences are expressed by intuitionistic fuzzy indices in Fig. 2.

Now the goal (A-IF fuzzy goal) is described by two functions: membership function $\mu_G(x)$ and the intuitionistic fuzzy index $\pi_G(x)$. If the value of $x$ attained is at

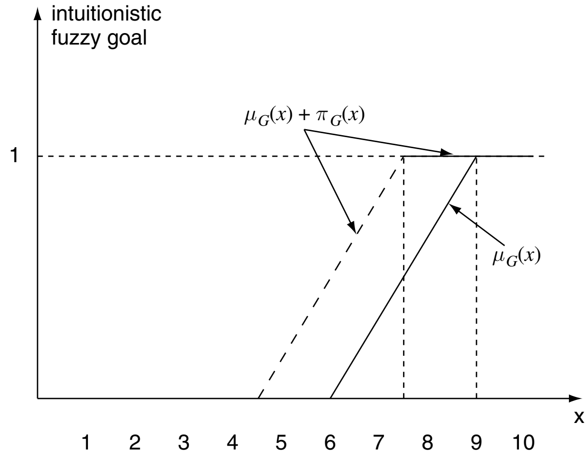



**Fig. 2** A-IF goal $G$, "*x should be much larger than 5*"

least 9, then $\mu_G(x)$ is equal to 1 which means that we are for sure fully satisfied with the $x$ attained. But even for $x$ greater than 7.5 it is possible (but not sure) that our level of satisfaction is equal to 1, as $\mu_G(x) + \pi_G(x)$ is equal to 1 in the interval $[7.5, 9]$. On the other hand, if the $x$ attained does not exceed 4.5, then we are for sure fully dissatisfied with such a value of $x$, or in other words, such a value is impossible. For $x$ belonging to the interval $[4.5, 6]$, our level of satisfaction can be (but not for sure as in the interval $\mu_G(x)$ is equal to 0) greater than zero because the intuitionistic fuzzy index $\pi_G(x)$ is greater than zero here. For $x$ greater than 6, $\mu_G(x)$ is greater than zero, so our level of satisfaction is for sure greater than zero.

Let our A-IF constraint $C$ be *"x should be about 6"*, as in Fig. 3.

A constraint in an A-IF environment, just the same as it was in the case of an A-IF goal, can vary. It is easy to imagine that demand for a product changes. We can foresee it to some extent but when faced with a concrete decision how much to produce, we are able to make a rough decision only. It is illustrated in Fig. 3. We are sure that for $x$ from interval $[5.5, 6.5]$, the level of satisfaction for the constrain is equal to 1. But because of the lack of knowledge expressed by $\pi_C(x)$, also for $x$ belonging to intervals $[5, 5.5)$ and $(6.5, 7]$ the level of satisfaction can be equal to 1. It means that for the above intervals the only sure (safe) information is the lower level of satisfaction given by $\mu_C(x)$ (as our level of satisfaction is described by a number from the interval $[\mu_C(x), \mu_C(x) + \pi_C(x)]$). For $x$ less than 2 or greater than 10 the level of satisfaction is equal to 0 (too low a production level is unprofitable, too big a level requires an expensive storage space). For $x$ belonging to $[2, 2.5)$ or $(9.5, 10]$ our level of satisfaction can be greater than 0 because of intuitionistic fuzzy index $\pi_C(x)$ but it is not sure as $\mu_C(x)$ is equal to 0 here. For the interval $[2.5, 5)$ the higher $x$ the higher the level of satisfaction. For the interval $(7, 9.5]$ the higher $x$ the lower the level of satisfaction. The level of satisfaction is obviously given for every $x$ as a number from the interval $[\mu_C(x), \mu_C(x) + \pi_C(x)]$.

Certainly, both for the A-IF goal $G$ and constraint $C$ it is easy to find for each $x \in X$ non-membership functions. Namely

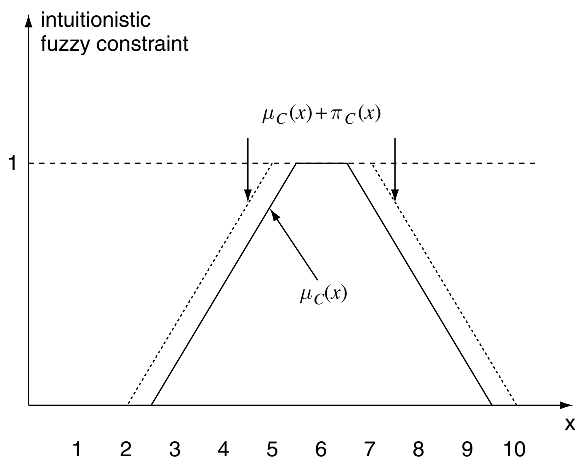
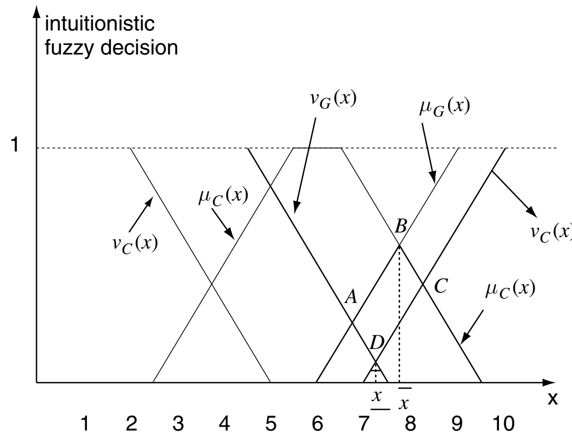



**Fig. 3** A-IF constraint $C$ ,*"x should be about 6"*

**Fig. 4** A-IF decision




$$\nu_G(x) = 1 - (\mu_G(x) + \pi_G(x)) \text{ and } \nu_C(x) = 1 - (\mu_C(x) + \pi_C(x)) \tag{16}$$

and knowing the above values it is possible to use formulas (13).

The min-type fuzzy decision is given (cf. (13)) in bold lines in Fig. 4.

We have therefore a concept of an A-IF decision, that is a solution to the decision making problem considered (attain $G$ and satisfy $C$). In fact we are interested in the area $ABCD$ only because in this region membership functions are greater than non-membership functions. The interested region is for $x$ belonging to the interval [6.5, 8.5]. But in practice, however, if we wish to implement such a solution we need to find a crisp solution for which equations (15) are helpful. For our example $\underline{x}$ is equal to 7.25, $\overline{x}$ is equal to 7.75. Acceptable solutions lie in the interval $[x_{min}, x_{max}] = [6.5, 8.25]$ - here a tradeoff between wanted values of $\mu_D$ and $\nu_D$ occurs.

Note that in Fig. 4, $\mu_D(x) < 1$ and $\nu_D > 0$ which means that there is no option which fully satisfies or could under some conditions satisfy (for $\nu_D(x) = 0$ it would be a chance that values $\pi_D(x) > 0$ could increase $\mu_D(x)$) both the fuzzy goal and fuzzy constraint. In other words, there is a discrepancy or conflict between the goal and the constraint.

In fact determination of a non-fuzzy decision is basically the problem of defuzzification, and the simplest defuzzification of the type $\widetilde{x} = 0.5(\underline{x} + \overline{x})$ is for linear goals and constrains a solution. For the considered example $\widetilde{x} = 7.5$.

## 4 Group Decision Making and Soft Measures of Consensus

We assume that there is a set of individuals who provide their testimonies assumed to be *preferences* over the set of alternatives. The problem is to find a *solution*, i.e. an alternative (or a set of alternatives) which is best acceptable by the group of individuals as a whole. For a different point of departure, involving choice sets or utility functions, we may refer the interested reader to, e.g., Kim [29], Salles [37], etc.

An important research direction, that may help overcome difficulties and negative results, is based on an *individual* and *social fuzzy preference relation*. Suppose that we have a set of $n \geq 2$ alternatives, $S = \{s_1, \ldots, s_n\}$ (we use $S$ for the set of alternatives to be consistent with many papers on this topic, including those cited later on), and a set of $m \geq 2$ individuals, $I = \{1, \ldots, m\}$. Then, an individual's $k \in I$ individual fuzzy preference relation in $S \times S$ assigns a value in the unit interval for the preference of one alternative over another. Though normally some conditions are to be satisfied, as, e.g., reflexivity, connectivity, (max-min) transitivity, etc., it is not clear which of these "natural" properties of preference relations should be assumed (cf. Salles [37]); a deep discussion is in, e.g., Fodor and Roubens' [15], and in De Baets et al.'s paper in this volume.

Here we assume that the individual and social fuzzy preference relations are defined in $S \times S$, i.e. assign to each pair of alternatives a strength of preference of one over another as a value from [0, 1]. The fuzzy, and then A-IF, preferences will be employed only instrumentally, and we will not discuss them and their properties in more detail.

Another basic element is here the concept of a *majority* – notice that a solution is to be an alternative(or alternatives) best acceptable by the group as a whole, that is by (at least!) *most* of its members since in practically no real nontrivial situation it would be accepted by all. Some problems and negative results with group decision making are closely related to too strict a representation of majority (e.g., at least a half, at least 2/3, ...). A natural line of reasoning is to make that strict concept of majority closer to its real human perception by making it vague. A good, often cited example in a biological context may be found in Loewer and Laddaga [32]:

> " ...It can correctly be said that there is a consensus among biologists that Darwinian natural selection is an important cause of evolution though there is currently no consensus concerning Gould's hypothesis of speciation. This means that there is a widespread agreement among biologists concerning the first matter but disagreement concerning the second ..."

and it is clear that a rigid majority as, e.g., more than 75% would evidently not reflect the essence of this statement. However, a strict majority may be necessary as in all political elections. In this section we will mention a fuzzy majority but will not use it in the concept of our A-IF extensions.

We have a set of $n \geq 2$ alternatives, $S = \{s_1, \ldots, s_n\}$, and a set of $m \geq 2$ individuals, $I = \{1, \ldots, m\}$. Each individual $k \in E$ provides his or her testimony as to the alternatives in $S$, assumed to be individual fuzzy preference relations defined over $S$ (i.e. in $S \times S$) (cf., e.g., Kacprzyk, Nurmi and Fedrizzi [27]).

An *individual fuzzy preference relation* of individual $k$, $R_k$, is given by its membership function $\mu_{R_k} : S \times S \longrightarrow [0, 1]$ such that

$$\mu_{R_k}(s_i, s_j) = \begin{cases} 1 & \text{if } s_i \text{ is definitely preferred to } s_j \\ c \in (0.5, 1) & \text{if } s_i \text{ is slightly preferred to } s_j \\ 0.5 & \text{in the case of indifference} \\ d \in (0, 0.5) & \text{if } s_j \text{ is slightly preferred to } s_i \\ 0 & \text{if } s_j \text{ is definitely preferred to } s_i \end{cases} \tag{17}$$

If card $S$ is small enough (as assumed here), an individual fuzzy preference relation of individual $k$, $R_k$, may conveniently be represented by an $n \times n$ matrix $R_k = [r_{ij}^k]$, such that $r_{ij}^k = \mu_{R_k}(s_i, s_j)$; $i, j = 1, \ldots, n$; $k = 1, \ldots, m$. $R_k$ is commonly assumed (also here) to be reciprocal in that $r_{ij}^k + r_{ji}^k = 1$; moreover, it is also normally assumed that $r_{ii}^k = 0$, for all $i, k$; for a different, more justified convention, cf. García-Lapresta and Llamazares [16]. Moreover, we will not use here a more sophisticated concept of a fuzzy preference systems proposed by De Baets et al. which is presented in their paper in this volume. The reasoning is in this case principally the same.

Two lines of reasoning may be followed here (cf. Kacprzyk [17, 18, 19]):

- a direct approach: $\{R_1, \ldots, R_m\} \longrightarrow$ solution, that is, a solution is derived directly (without any intermediate steps) just from the set of individual fuzzy preference relations, and
- an indirect approach: $\{R_1, \ldots, R_m\} \longrightarrow R \longrightarrow$ solution, that is, from the set of individual fuzzy preference relations we form first a social fuzzy preference relation, $R$ (to be defined later), which is then used to find a solution.

A solution is here, unfortunately, not clearly understood – see, e.g., Nurmi [33, 34] for diverse solution concepts. In this paper we will only sketch the derivation of some more popular solution concepts. Then, we will mention some possible approaches in the application of A-IF preference relations.

One of the best solution concepts is that of a core or a set of undominated alternatives. Suppose that the nonfuzzy required majority be $r$ (e.g., at least 50%).

Then, An alternative $s_i \in S$ belongs to the *core* iff there is no other alternative $y \in S$ that defeats $x$ by the required majority $r$.

We can extend the notion of a core to cover fuzzy individual preference relations by defining *the fuzzy $\alpha$-core* as follows (cf. Nurmi [33]): an alternative $s_i \in S$ belongs to the *fuzzy $\alpha$-core* $S_\alpha$ iff there exists no other alternative $s_j \in S$ such that $r_{ji} > \alpha$ for at least $r$ individuals, i.e. iff a sufficient majority of voters does not feel strongly enough against it.

Another nonfuzzy solution concept with much intuitive appeal is a minimax set defined as follows: for each $x, y \in S$ denote the number of individuals preferring $x$ to $y$ by $n(x, y)$. Then define

$$v(x) = \max_y n(y, x) \text{ and } n* = \min_x v(x)$$

and the minimax set is

$$Q(n*) = \{x \mid v(x) = n*\}$$

so that $Q(n*)$ contains those alternatives that in pairwise comparison with any other alternative are defeated by no more than $n*$ votes.

Analogously, we can define a *the minimax degree set $Q(\beta)$* as follows. Given $s_i, s_j \in S$ and let, for individuals $k = 1, \ldots, m$, and $v_D^k(x_j) = \max_i r_{ij}$, and

$v_D(x_j) = \max_k v_D^k(x_j)$, and if $\min_j v_D(x_j) = \beta$, then $Q(\beta) = \{x_j \mid v_D(x_j) = \beta\}$ (cf. Nurmi [33]).

Another concept is a *minimax opposition set*. Let $n_{ij}$ be the number of those individuals for whom $r_{ij} > r_{ji}$ and let $v_f(x_j) = \max_i n_{ij}$. Denote by $\bar{v}_f$ the minimum of $v_f(x_j)$ with respect to $j$, i.e. $\bar{v}_f = min_j v_f(x_j)$, and then $Q(v_f) = \{x_j \mid v_f(x_j) = \bar{v}_f\}$.

A more general solution concept, the *α-minimax set* (cf. Nurmi [33]) denoted $Q^\alpha(v_f^\alpha)$, is defined as follows. Let $n_\alpha(x_i, x_j)$ be the number of individuals for whom $r_{ij} \le \alpha$ for some value of $\alpha \in [0, 0.5)$. We now define $\forall x_i \in S : v_f^\alpha(x_i) = \max_j n_\alpha(x_i, x_j)$ and $\bar{v}_f^\alpha = \min_i v_f^\alpha(x_i)$. Then

$$Q^\alpha(v_f^\alpha) = \{x_i \mid v_f^\alpha(x_i) = \bar{v}_f^\alpha\}$$

For some other solution concepts, mainly those based on fuzzy tournaments, see Nurmi and Kacprzyk [35].

Among solution concepts derived via an indirect approach, i.e. from a social fuzzy preference relation, first, the set $S_\alpha$ of *α-consensus winners* is defined as: $s_i \in S_\alpha$ iff $\forall s_j \neq s_i : r_{ij} \geq \alpha$, with $0.5 < \alpha \le 1$.

Whenever $S_\alpha$ is nonempty, it is a singleton, but it does not always exist.

Thus, it may be useful to find other solution concepts that specify nonempty alternative sets even when $S_\alpha$ is empty. For, instance, Kramer's minimax set is a set of *minimax consensus winners*, $S_M$. Let $\bar{r}_j = \max_i r_{ij}$ and $\bar{r} = \min_j \max_i r_{ij}$. Then $s_i \in S_M$ (the set of minimax consensus winners) iff $\bar{r}_i = \bar{r}$. Clearly $S_M$ is always nonempty, but not necessarily a singleton. It contains those alternatives which, when confronted with their toughest competitors, fare best, i.e. win by the largest score (if $\bar{r} \le 0.5$) or lose by the smallest one (if $\bar{r} > 0.5$).

Consensus is a prerequisite for an effective and efficient group decision making. Degrees of consensus for fuzzy preference have been first considered by Bezdek, Spillman and Spillman [8, 9], and their approach is quite closely related to the A-IF extensions proposed by the authors in [51].

Soft degrees of consensus based on fuzzy linguistic quantifiers as representations of a fuzzy majority proposed in Kacprzyk [19]), and then advanced in Kacprzyk and Fedrizzi [21, 22, 23, 24], and Kacprzyk, Fedrizzi and Nurmi [25, 26], see also Kacprzyk, Nurmi and Fedrizzi [27], and Zadrożny [65], give much more insight into the very essence of consensus but due to the necessity of handling fuzzy quantifiers, have not been extended to A-IFSs and will not be considered in this paper.

Now we will present examples of A-IF extensions using mainly A-IF preference relations. Our analysis will be based on our works that are main contributions in this field. Notably, in Szmidt and Kacprzyk [42, 43, 45, 46, 48, 49, 51, 52] new A-IF models have been proposed to solve some group decision problems, and to determine soft measures of consensus.

The degree of consensus is derived in three steps:

1. for each pair of individuals we derive a degree of agreement as to their preferences between *all* the pairs of alternatives,

2. we aggregate these degrees to obtain a degree of agreement of each pair of individuals as to their preferences between $Q_1$ (a linguistic quantifier as, e.g., "most", "almost all", "much more than 50%", ...) pairs of *relevant* alternatives, and
3. we aggregate these degrees to obtain a degree of agreement of $Q_2$ (a linguistic quantifier similar to $Q_1$) pairs of *important* individuals as to their preferences between $Q_1$ pairs of *relevant* alternatives, and this is meant to be the *degree of consensus* sought.

We assume that both the individuals and alternatives are assigned different degrees of importance and relevance. In the process of the derivation of a degree of consensus a crucial role is played by a fuzzy majority, equated with a fuzzy linguistic quantifier, i.e. $Q_1$ and $Q_2$. For instance, $Q$ = "most" may be given as (cf. Kacprzyk [20]):

$$Q_{\text{“}most\text{”}}(x) = \begin{cases} 1 & \text{for} \quad x > 0.8 \\ 2x - 0.6 & \text{for} \quad 0.3 \leq x \leq 0.8 \\ 0 & \text{for} \quad x < 0.3 \end{cases} \tag{18}$$

and this form of "most" will be used throughout this chapter.

We will employ Zadeh's [64] calculus of linguistically quantified statements.

The relevance of options is given as a fuzzy set defined in the set of alternatives $S$ such that $\mu_B(s_i) \in [0, 1]$ is a *degree of relevance* of option $s_i$, from 0 for fully irrelevant to 1 for fully relevant, through all intermediate values, and the more relevant the option the higher this degree. The relevance of a pair of options, $(s_i, s_j) \in S \times S$, may be defined, say, as

$$b_{ij}^k = \frac{1}{2}[\mu_B(s_i) + \mu_B(s_j)] \tag{19}$$

And analogously, the *importance of individuals*, $I$, is defined as a fuzzy set in the set of individuals such that $\mu_I(k) \in [0, 1]$ is a *degree of importance* of individual $k$, from 0 for fully unimportant to 1 for fully important, through all intermediate values. Then, the importance of a pair of individuals, $(k_1, k_2)$, $b_{k_1,k_2}^I$, may be defined in various ways, e.g., analogously as (19), i.e.

$$b_{k_1,k_2}^I = \frac{1}{2}[\mu_I(k_1) + \mu_I(k_2)] \tag{20}$$

The three step procedure of finding consensus is a starting point for employing A-IFSs instead of fuzzy sets for modelling *individual preference relations*.

In the A-IF approach (cf. Szmidt and Kacprzyk [42, 43, 45, 46, 48, 49], [51, 52], each individual $k$ provides his or her *(individual) A-IF preference relation*, giving not only $R_k$ (given, as previously, by its membership function $\mu_{R_k}$) but also

- $\nu_{R_k}$ – a non-membership function, $\nu_k : S \times S \to [0, 1]$, conveniently represented by a matrix $[\nu_{ij}^k(s_i, s_j)]$; $i, j = 1, \ldots, n$; $k = 1, \ldots, m$, and

**Table 1** The meaning of A-IF parameters when individuals pairwisely compare options

| $\mu_{i,j}$ | $\pi_{i,j}$ | $\nu_{i,j}$ |
|---|---|---|
| advantages<br>(*i*-th option better) | lack of knowledge<br>(?) | drawbacks<br>(*j*-th option better) |

- $\Pi_k$ – a so-called intuitionistic fuzzy index, $\pi_k : S \times S \to [0,1]$, conveniently represented by a matrix $[\pi_{ij}^k(s_i, s_j)]$; $i, j = 1, \ldots, n$; $k = 1, \ldots, m$.

meants as in Table 1.

For each compared pair of options $(i, j)$:

- a degree of membership $\mu_{i,j}$ means the value we assign to all the advantages of $i$-th option or to all the drawbacks of $j$-th option,
- a degree of non-membership $\nu_{i,j}$ means the value we assign to all the drawbacks of $i$-th option or to all the advantages of $j$-th option,
- an intuitionistic fuzzy index $\pi_{i,j}$ means the area concerning both compared options where we have not enough information (or do not wish to express it) to say which option is better.

Using the above A-IF preferences, we have applied the approaches adapted from the fuzzy approach (cf. Kacprzyk [18], Kacprzyk, Fedrizzi and Nurmi [25]) but due to a wide array of the applied methods, we can only refer the reader to [42, 43, 45, 46, 48, 49, 51].

Taking advantage of the power of A-IFSs, we also proposed a different approach to the analysis of extent of the agreement in the group of experts, than in Kacprzyk [18], Kacprzyk, Fedrizzi and Nurmi [25] using the concept of a distance between A-IF preferences to evaluate how far the group is from full agreement (consensus) [52]. An extent of a group agreement was given by a number from interval [0, 1], where 0 means consensus (in a traditional sense), 1 means dissensus. Notice that this is close in spirit to Bezdek, Spillman and Spilman [8, 9].

An advantage of using distances is due to the fact that we can avoid some formal difficulties while aggregating A-IF preferences (cf. Szmidt and Kacprzyk [44, 47, 60] for effective methods of calculating distances). The disadvantage is that in a case of A-IFSs a small distance between options does not automatically guarantee their similarity (in Szmidt and Kacprzyk [57] an illustraiting example is given). The use of the entropy (cf. Szmidt and Kacprzyk [50]) in addition to distances makes it possible to get even more insight (cf. Szmidt and Kacprzyk [55]) as both the distances and entropy should be analysed separately. To avoid the necessity of analyzing the distances and entropy separately, we proposed one measure – similarity (cf. Szmidt and Kacprzyk [56, 57, 59]) that may be viewed to give an additional quality and a more global view.

It is noteworthy that the distances used in the definition of similarity take into account all three functions (membership, non-membership and hesitation margin) characterizing A-IFSs. A motivation for using all three functions and poor effects of omitting one of them (from the point of view of decision making) is given in Szmidt and Kacprzyk [60, 61]).

The extent of similarity for two experts $k_1, k_2$ considering $n$ options (proposed in [57]) can be given as

$$\begin{aligned} Sim^{k_1,k_2} &= \frac{1}{A}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Sim^{k_1,k_2}(i,j) = \\ &= \frac{1}{A}[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (\mid \mu_{ij}(k_1) - \mu_{ij}(k_2) \mid + \mid \nu_{ij}(k_1) - \nu_{ij}(k_2) \mid + \\ &\quad + \mid \pi_{ij}(k_1) - \pi_{ij}(k_2) \mid)]/[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (\mid \mu_{ij}(k_1) - \nu_{ij}(k_2) \mid + \\ &\quad + \mid \nu_{ij}(k_1) - \mu_{ij}(k_2) \mid + \mid \pi_{ij}(k_1) - \pi_{ij}(k_2) \mid)] \end{aligned} \tag{21}$$

where $A = \frac{1}{2C_n^2} = \frac{1}{n(n-1)}$.

Some other measures of similarity are proposed in [56, 57, 59]).

When we have $m$ experts, we examine a similarity of their preferences pairwisely, e.g. given by (21) and next, we find an agreement of all experts

$$Sim = \frac{1}{m(m-1)}\sum_{p=1}^{m-1}\sum_{r=p+1}^{m} Sim^{k_p,k_r} \tag{22}$$

where $Sim^{k_p,k_r}$ is given e.g. by (21).

Using the above concept of similarity measure to analyse the extent of agreement between experts consists in saying if all of the considered pairs of expert's preferences are

- just the same (i.e. full agreement meaning consensus in a traditional sense - the proposed measure of similarity is equal to 0),
- quite opposite (i.e. full disagreement - similarity tends to infinity),
- different to some extent (what means that a distance from consensus is from the open interval (0, 1))
- to the same extent similar as dissimilar - the proposed measure of similarity is equal to 1.

The concept of similarity was also used while supporting medical diagnosis (cf. Szmidt and Kacprzyk [53, 58]) which, as has been already mentioned earlier, makes it possible to avoid drawing the conclusions about a strong similarity between A-IFSs on the basis of small distances between these sets.

A measure of similarity between A-IFSs turned out to be useful as well while proposing a new solution to a multi-criteria decision making problem [62]. More specifically, having a set $M$ of options fulfilling a set of criteria $C$, we wished to rank the options satisfying: $C_j$, and $C_k, \ldots,$ and $C_p$ or $C_s$ whereas each criterion is fulfilled to some extent $\mu$ and is not fulfilled to some extent $\nu$ (when $0 \leq \mu + \nu \leq 1$).

The solution of this problem was proposed by Chen and Tan [12] but it does not always give a proper answer as indicated by Liu [31]. Applying a measures of similarity between A-IFSs made it possible to overcome some deficiencies. We evaluated options comparing them to the positive-ideal solution and negative-ideal solution. The best considered option should be as close as possible to the positive-ideal solution and as far as possible to the negative-ideal solution. In our previous works (cf. e.g, [57]) we have shown that looking for the solution (the best option) taking into account only positive-ideal solution can be misleading.

## 5 Case-Based Reasoning (CBR)

Finally, we will mention in this section a potential use of A-IFSs in case based reasoning. This may be a promising direction in broadly perceived decision making under imprecision, and such an idea was proposed in a recent paper by Szmidt and Kacprzyk [61].

Case-based reasoning (cf. Schank [38], Aamodt and Plaza [1]) is a problem solving paradigm that in many respects is fundamentally different from other major approaches. Instead of making use solely from general knowledge of a problem domain, CBR is able to utilize the specific knowledge of previously experienced, concrete problem situations called cases. A new problem is solved by finding a similar past case, and reusing it in the new problem situation.

Central tasks that all case-based reasoning methods have to deal with are to identify the current problem situation, find a past case similar to the new one, use that case to suggest a solution to the current problem, evaluate the proposed solution, and update the system by learning from this experience.

Reasoning by re-using past cases is a powerful and frequently applied way to solve problems by humans. Several studies have given empirical evidences for the dominant role of specific, previously experienced situations (cases) in human problem solving (e.g Ross[36]). Anderson[2] has shown that people use past cases as models when learning to solve problems, particularly in early learning. Other results indicate (cf. Kolodner[30]) that the use of past cases is a predominant problem solving method among experts as well.

The use of A-IFSs in case-based reasoning seems to be a promising direction. A higher expressing power of A-IFSs can be exemplified by: a financial consultant working on a difficult credit decision task collects material on a finance situation of a company. The collected material consists of both positive and negative indicators, and some indicators are missing. Such a situation, often met in practice, is ideal to be describe via A-IFSs which make use of positive and (independently given) negative information, which at the same time making it possible to express and take into account the amount of lacking information as well. Also the processes of decision making, assessing the soft grades of consensus (which are another central task in CBR) are described and solved in a more adequate way when using

A-IFSs. Additionally, the advantages of A-IFSs can be also noticed due to their possibility of confident eliminating negative cases what can considerably speed up the process of finding the past cases similar to the new one – especially in big case-bases.

When having a current situation (case) described in terms of A-IFSs (the data can be delivered by experts or assigned automatically (cf. Szmidt and Baldwin [40]), and all the aspects of the decision, the case is added to the database. When a new case is to be considered, it should be compared to the already collected cases. At this stage similarity measures are necessary. Although some problems are still open, e.g. aggregation of the data concerning the cases, application of A-IFSs to CBR seems promising.

## 6 Concluding Remarks

The purpose of this paper was to briefly survey some developments in the use of A-IFSs in decision making. We have basically examined some fuzzy decision making models, mainly those based on a general approach due to Bellman and Zadeh [7], and then considered a very important class of group decision making and measuring a degree of consensus where there is a considerable literature on traditional fuzzy approaches.

We have indicated some possible use of A-IFSs in those models. Notably, we have indicated that a new class of models may be more fruitful, mainly those based on the use of distances and similarities between A-IFSs, and also on their entropies. These quantities reflect in a very good way some intrinsic characteristics of aspects of decision making represented by A-IFSs, and seem to be a promising direction.

Finally, we have mentioned – again as a promising direction – the use of A-IFSs in case based reasoning.

We hope that this paper can trigger a further research in the use of A-IFSs in decision making under imprecision, and that A-IFSs will make it possible to express and grasp many further aspects of imprecision that may be of relevance in decision making models. Notably, by employing A-IFSs a decision maker may explicitly express both advantages (membership) and disadvantages (non-membership) of a situation considered independently. An imprecision (hesitation) is taken into account as well. The importance of such an approach lies in the fact that most people concentrate usually on one or two "most visible" aspects of a problem. They do not try to find out the contrary arguments or to consider uncertain (in a wide sense, i.e. not restricted to randomness) aspects of a situation (cf. Sutherland [39]). A-IFSs with their structure make a user consider a situation/problem more properly – from different points of view – including all important aspects which should be taken into account but which, unfortunately, are often omitted by people making decisions. A-IFSs may be certainly viewed as modern and promising means for decision making under imprecision.

## References


1. Aamodt A. and Plaza E. (1994) Case-based reasoning: foundational issues, methodological variations, and system. Artificial Intelligence Communications, IOS Press, 7, 1, 39–59.
2. Anderson J.R. (1983) The architecture of cognition. Harvard University Press, Cambridge.
3. Atanassov K. (1983), Intuitionistic Fuzzy Sets. VII ITKR Session. Sofia (Deposed in Centr. Sci.-Techn. Library of Bulg. Acad. of Sci., 1697/84) (in Bulgarian).
4. Atanassov K. (1986) Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems, 20, 87–96.
5. Atanassov K. (1999), Intuitionistic Fuzzy Sets: Theory and Applications. Springer-Verlag.
6. Atanassov K. and Gargov G. (1989), Interval-valued intuitionistic fuzzy sets. Fuzzy sets and Systems, 31 (3), 343–349.
7. Bellman, R.E. and Zadeh, L.A. (1970). Decision making in a fuzy environment, *Management Science*, 17, 141–164.
8. Bezdek, J.C., Spillman, and B. Spillman, R. (1978) A fuzzy relation space for group decision theory. Fuzzy Sets and Systems, 1, 255–268.
9. Bezdek, J.C., Spillman, B. and Spillman, R. (1979) Fuzzy relation space for group decision theory: An application. Fuzzy Sets and Systems 2, 5–14.
10. Blin, J.M. (1974) Fuzzy relations in group decision theory. J. of Cybernetics, 4, 17–22.
11. Blin, J.M. and Whinston, A.P. (1973) Fuzzy sets and social choice. J. of Cybernetics, 4, 17–22.
12. Chen S.M. and Tan J.M. (1994) Handling multi-criteria fuzzy decision making problems based on vague set theory. Fuzzy Sets and Systems 67 (2), 163–172.
13. Dubois D., Gottwald S., Hájek P., Kacprzyk J., Prade H. (2005) Terminological difficulties in fuzzy set theory – the case of "intuitionistic fuzzy sets". Fuzzy Sets and Systems 156 (3), 485–491.
14. Dubois D. and Prade H. (1994) Fuzzy set modelling in case-based reasoning. International Journal of Intelligent Systems, 13, 345–373.
15. Fodor, J. and Roubens, M. (1994) Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer, Dordrecht.
16. García-Lapresta, J.L. and Llamazares, B. (2000). Aggregation of fuzzy preferences: Some rules of the mean. *Social Choice and Welfare*, 17, 673–690.
17. Kacprzyk, J. (1985) Group decision-making with a fuzzy majority via linguistic quantifiers. Part I: A consensory-like pooling; Part II: A competitive-like pooling. Cybernetics and Systems, 16, 119–129 (Part I), 131–144 (Part II).
18. Kacprzyk, J. (1986) Group decision making with a fuzzy linguistic majority. Fuzzy Sets and Systems, 18, 105–118.
19. Kacprzyk, J. (1987) On some fuzzy cores and 'soft' consensus measures in group decision making. In J.C. Bezdek (Ed.): The Analysis of Fuzzy Information, Vol. 2, CRC Press, Boca Raton, 119–130.
20. Kacprzyk, J. (1997) Multistage Fuzzy Control (Wiley,Chichester ).
21. Kacprzyk, J. and Fedrizzi, M. (1986) 'Soft' consensus measures for monitoring real consensus reaching processes under fuzzy preferences. Control and Cybernetics, 15, 309–323.
22. Kacprzyk, J. and Fedrizzi, M. (1988) A 'Soft' measure of consensus in the setting of partial (fuzzy) preferences. Europ. J. of Operational Research, 34, 315–325.
23. Kacprzyk, J. and Fedrizzi, M. (1989) A 'human-consistent' degree of consensus based on fuzzy logic with linguistic quantifiers. Mathematical Social Sciences, 18, 275–290.
24. Kacprzyk, J. and Fedrizzi, M., Eds. (1990) Multiperson Decision Making Models Using Fuzzy Sets and Possibility Theory, Kluwer, Dordrecht.
25. Kacprzyk J., M. Fedrizzi and H. Nurmi (1992) Group decision making and consensus under fuzzy preferences and fuzzy majority. Fuzzy Sets and Systems, 49, 21–32.
26. Kacprzyk, J. and Nurmi, H. (1998) Group decision making under fuzziness, in R. Słowiński (Ed.): Fuzzy Sets in Decision Analysis, Operations Research and Statistics, Kluwer, Boston, 103–136.
27. Kacprzyk, J., Nurmi, H. and Fedrizzi, M., Eds. (1996) Consensus under Fuzziness, Kluwer, Boston.

28. Kang W.-T. (2004) Protest voting and abstention under pluraluity rule elections, Journal of theoretical Politics, 16, 71–102.
29. Kim, J.B. (1983). Fuzzy rational choice functions. Fuzzy Sets and Systems, 10, 37–43.
30. Kolodner J. (1992) An introduction to case-based reasoning. Artificial Intelligence Review, 6 (1), 3–34.
31. Liu H.W. Multi-criteria decision making methods based on intuitionistic fuzzy sets. In press.
32. Loewer B. and R. Laddaga (1985) Destroying the consensus. In Loewer B. (Guest Ed.): Special Issue on Consensus. Synthese, 62 (1), 79–96.
33. Nurmi, H. (1981) Approaches to collective decision making with fuzzy preference relations, Fuzzy Sets and Systems, 6, 249–259.
34. Nurmi, H. (1987) Comparing Voting Systems, Reidel, Dordrecht.
35. Nurmi, H. and Kacprzyk, J. (1991). On fuzzy tournaments and their solution concepts in group decision making. Europ. J. of Operational Research, 51, 223–232.
36. Ros B.H. (1989) Some psychological results on case-based reasoning. Case-Based Reasoning Workshop, DARPA, 144–147.
37. Salles, M. (1996). Fuzzy utility. In S. Barberà, P.J. Hammond and C. Seidl (Eds.): Handbook of Utility Theory, Vol. I, Kluwer, Boston.
38. Schank R. (1982) Dynamic memory: a theory of reminding and learning in computers and people. Cambridge University Press.
39. Sutherland S. (1994) Irrationality. The Enemy Within. Penguin Books.
40. Szmidt E. and Baldwin J. (2005) Assigning the parameters for Intuitionistic Fuzzy Sets. Notes on IFSs, 11 (6), 1–12.
41. Szmidt E. and Baldwin J. (2006) Intuitionistic Fuzzy Set Functions, Mass Assignment Theory, Possibility Theory and Histograms. 2006 IEEE World Congress on Computational Intelligence, 237–243.
42. Szmidt E. and Kacprzyk J. (1996a) Intuitionistic fuzzy sets in group decision making, Notes on IFS, 2, 15–32.
43. Szmidt E. and Kacprzyk J. (1996c) Remarks on some applications of intuitionistic fuzzy sets in decision making, Notes on IFS, 2 (3), 22–31.
44. Szmidt E. and Kacprzyk J. (1997) On measuring distances between intuitionistic fuzzy sets, Notes on IFS, 3 (4), 1–13.
45. Szmidt E. and Kacprzyk J. (1998a) Group Decision Making under Intuitionistic Fuzzy Preference Relations. IPMU'98, Paris, La Sorbonne, 172–178.
46. Szmidt E. and Kacprzyk J. (1998b) Applications of Intuitionistic Fuzzy Sets in Decision Making. Proc. 8th Congreso Español sobre Tecnologias y Lógica Fuzzy, Univ. de Navarra, 143–149.
47. Szmidt E. and Kacprzyk J. (2000) Distances between intuitionistic fuzzy sets, Fuzzy Sets and Systems, 114 (3), 505–518.
48. Szmidt E. and Kacprzyk J. (2000) On Measures on Consensus Under Intuitionistic Fuzzy Relations. IPMU 2000, 1454–1461.
49. Szmidt E. and Kacprzyk J. (2001) Distance from Consensus Under Intuitionistic Fuzzy Preferences. Proc. EUROFUSE Workshop on Preference Modelling and Applications, Granada, 73–78.
50. Szmidt E., Kacprzyk J. (2001) Entropy for intuitionistic fuzzy sets. Fuzzy Sets and Systems, 118 (3), 467–477.
51. Szmidt E. and Kacprzyk J. (2001) Analysis of Consensus under Intuitionistic Fuzzy Preferences. Proc. Int. Conf. in Fuzzy Logic and Technology. De Montfort Univ. Leicester, UK, 79–82.
52. Szmidt E. and Kacprzyk J. (2002a) Analysis of Agreement in a Group of Experts via Distances Between Intuitionistic Fuzzy Preferences. Proc. IPMU'2002, Annecy, 1859–1865.
53. Szmidt E. and Kacprzyk J. (2002b) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L. Jain, J. Kacprzyk (Eds.): Recent Advances in Intelligent Paradigms and Applications. Springer-Verlag, 57–70.
54. Szmidt E. and Kacprzyk J. (2002c) Evaluation of Agreement in a Group of Experts via Distances Between Intuitionistic Fuzzy Sets. Proc. IEEE-IS'2002 – Int. IEEE Symposium: Intelligent Systems, Varna, 166–170.

55. Szmidt E., Kacprzyk J. (2003) Group Agreement Analysis via Distances and Entropy of Intuitionistic Fuzzy Sets. Int. Conference on Fuzzy Information Processing - Theories and Applications, Beijing, Springer, 631–635.
56. Szmidt E., Kacprzyk J.(2004) A Concept of Similatity for Intuitionistic Fuzzy Sets and its use in Group Decision Making. Proc. 2004 IEEE Int. Conference on Fuzzy Systems, Budapest, 1129-1134.
57. Szmidt E. and Kacprzyk J. (2004) Similarity of intuitionistic fuzzy sets and the Jaccard coefficient. Proc. IPMU 2004, Perugia, 1405–1412.
58. Szmidt E., Kacprzyk J. (2004) A Similarity Measure for Intuitionistic Fuzzy Sets and its Application in Supporting Medical Diagnostic Reasoning. LNAI, 3070, 388–393.
59. Szmidt E. and Kacprzyk J. (2005) A New Concept of a Similarity Measure for Intuitionistic Fuzzy Sets and its Use in Group Decision Making. In V. Torra, Y. Narukawa, S. Miyamoto (Eds.): Modelling Decisions for AI. LNAI 3558, Springer, 272–282.
60. Szmidt E. and Kacprzyk J. (2006) Distances Between Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. 3rd Int. IEEE Conf. “Intelligent Systems”, 716–721.
61. Szmidt E. and Kacprzyk J. (2006) A Model of Case Based Reasoning Using Intuitionistic Fuzzy Sets. 2006 IEEE World Congress on Computational Intelligence, 8428–8453.
62. Szmidt E. and Kacprzyk J. (2006) An Application of Intuitionistic Fuzzy Set Similarity Measures to a Multi-criteria Decision Making Problem. ICAISC 2006, LNAI 4029, Springer-Verlag, 314–323.
63. Zadeh L.A. (1965) Fuzzy sets. Information and Control, 8, 338–353.
64. Zadeh, L.A. (1983) A computational approach to fuzzy quantifiers in natural languages. Comput. Math. Appl., 9 (1), 149–184.
65. Zadrożny, S. (1997) An approach to the consensus reaching support in fuzzy environment. In: J. Kacprzyk, H. Nurmi and M. Fedrizzi (Eds.): Consensus under Fuzziness. Kluwer, Boston, 83–109.

# Fuzzy Methods for Data Mining and Machine Learning: State of the Art and Prospects

**Eyke Hüllermeier**

**Abstract** Methods for the automated induction of models and the extraction of interesting patterns from empirical data have recently attracted considerable attention in the fuzzy set community. This chapter briefly reviews some typical applications and highlights potential contributions that fuzzy set theory can make to machine learning, data mining, and related fields. Finally, a critical consideration of recent developments is given and some suggestions regarding future research are made.

## 1 Introduction

Aspects of knowledge representation and reasoning have dominated research in fuzzy set theory (FST) for a long time, at least in that part of the theory which lends itself to intelligent systems design and applications in artificial intelligence (AI). Yet, problems of automated learning and knowledge acquisition have more and more come to the fore in recent years. This is not very surprising in view of the fact that the "knowledge acquisition bottleneck" seems to remain one of the key problems in the design of intelligent and knowledge-based systems. Indeed, experience has shown that a purely knowledge-driven approach, which aims at formalizing problem-relevant human expert knowledge, is difficult, intricate, and tedious. More often than not, it does not even lead to fully satisfying results. Consequently, a kind of *data-driven* adaptation of fuzzy systems is often worthwhile. In fact, such a "tuning" even suggests itself since, in many applications, data is readily available. Indeed, recent research has shown that the traditional knowledge-driven approach can be complemented by a data-driven one in a reasonable way. In the extreme case, the former is even completely replaced by the latter. For example, several approaches in which fuzzy models (e.g., fuzzy rule bases) are learned from data in a fully automated way have already been developed [1].

In addition to this internal shift within fuzzy systems research, an external development has further amplified the aforementioned trends. This development is the great interest that the field of *knowledge discovery in databases* (KDD) has attracted in diverse research communities in recent years. As a response to the progress in digital data acquisition and storage technology, along with the limited human capabilities in analyzing and exploiting large amounts of data, this field has recently

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

emerged as a new research discipline, lying at the intersection of statistics, machine learning, data management, and other areas. According to a widely accepted definition, KDD refers to the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable structure in data [21]. The central step within the overall KDD process is *data mining*, the application of computational techniques to the task of finding patterns and models in data. Meanwhile, KDD has established itself as a new, independent research field, including its own journals and conferences.

The aim of this chapter is to convey an impression of the current status and prospects of FST in machine learning, data mining, and related fields. After a brief introduction to these fields (Sect. 2), we present a collection of typical applications of FST (Sect. 3). The examples are representative but not complete, and the section is definitely not a comprehensive review of the literature. In Sect. 4, we try to highlight in a more systematic way the potential contributions that FST can make to machine learning and data mining. Finally, we conclude with a critical consideration of recent developments and some suggestions for future research directions in Sect. 5.

## 2 Machine Learning, Data Mining, and Related Fields

The automated learning of models from empirical data is a central theme in several research disciplines, ranging from classical (inferential) statistics to more recent fields such as machine learning. Model induction may serve different purposes, such as accurate *prediction* of future observations or intelligible *description* of dependencies between variables in the domain under investigation, among other things. Typically, a model induction process involves the following steps:

- data acquisition
- data preparation (cleaning, transforming, selecting, scaling, ...)
- model induction
- model interpretation and validation
- model application

A common distinction of performance tasks in empirical[1] machine learning is supervised learning (e.g., classification and regression), unsupervised learning (e.g., clustering) and reinforcement learning. Throughout the chapter, we shall focus on the first two performance tasks that have attracted much more attention in the FST community than the latter one.

In unsupervised learning, the learning algorithm is simply provided with a set of data. The latter typically consists of data points $z \in \mathcal{Z}$, where $\mathcal{Z}$ is the Cartesian

[1] Here, *empirical* learning is used as an antonym to *analytical* learning. Broadly speaking, analytical learning systems do not require external inputs, whereas such inputs are essential for empirical learning systems. An example of analytical learning is speedup learning.

product of the domains of a fixed set of attributes. That is, an observation $z$ is described in terms of a feature vector. However, $\mathcal{Z}$ can also be of more general nature. For example, the analysis of complex objects such as, e.g., sequences, trees, or graphs, which cannot be directly represented as a feature vector, has recently received attention. Broadly speaking, the goal in unsupervised learning is to discover any kind of structure in the data, such as properties of the distribution, relationships between data entities, or dependencies between attributes. This includes, e.g., non-parametric features such as modes, gaps, or clusters in the data (Sect. 3.1), as well as interesting patterns like those discovered in association analysis (Sect. 3.4).

The setting of supervised learning proceeds from a predefined division of the data space into an input space $\mathcal{X}$ and an output space $\mathcal{Y}$. Assuming a dependency between the input attributes and the output, the former is considered as the *predictive* part of an instance description (like the regressor variables in regression analysis), whereas the latter corresponds to the target to be predicted (e.g., the dependent variable in regression). The learning algorithm is provided with a set of labeled examples $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Again, the inputs $x$ are typically feature vectors. A distinction between different types of performance tasks is made according to the structure of the output space $\mathcal{Y}$. Even though problems involving output spaces of a richer structure have been considered recently (e.g., so-called ranking problems [23]), $\mathcal{Y}$ is typically a one-dimensional space. In particular, the output is a categorical attribute (i.e., $\mathcal{Y}$ is a nominal scale) in *classification*. Here, the goal is to generalize beyond the examples given by inducing a model that represents a complete mapping from the input space to the output space (a hypothetical classification function). The model itself can be represented by means of different formalisms such as, e.g., threshold concepts or logical conjunctions. In *regression*, the output is a numerical variable, hence the goal is to induce a real-valued mapping $\mathcal{X} \longrightarrow \mathcal{Y}$ that approximates an underlying (functional or probabilistic) relation between $\mathcal{X}$ and $\mathcal{Y}$ well in a specific sense. So-called *ordinal regression* is in-between regression and classification: the output is measured on an ordinal scale.

As can be seen, supervised machine learning puts special emphasis on induction as a performance task. Moreover, apart from the *efficiency* of the induced model, the *predictive accuracy* of that model is the most important quality criterion. The latter refers to the ability to make accurate predictions of outputs for so far unseen inputs. The predictive accuracy of a model $h : \mathcal{X} \longrightarrow \mathcal{Y}$ is typically measured in terms of the *expected loss*, i.e., the expected value of $\ell(y, h(x))$, where $\ell(\cdot)$ is a loss function $\mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$ (and $(x, y)$ an example drawn at random according to an underlying probability measure over $\mathcal{X} \times \mathcal{Y}$.[2])

Data mining has a somewhat different focus. Here, other aspects such as, e.g., the *understandability*, gain in importance. In fact, the goal in data mining is not necessarily to induce *global* models of the system under consideration (e.g., in the form of a functional relation between input and output variables) or to recover some underlying data generating process, but rather to discover *local* patterns of interest,

[2] Since this measure is normally unknown, the expected loss is approximated by the *empirical* loss in practice, i.e., the average loss on a test data set.

e.g., very frequent (hence typical) or very rare (hence atypical) events. Data mining is of a more explanatory nature, and patterns discovered in a data set are usually of a *descriptive* rather than of a *predictive* nature. Data Mining also puts special emphasis on the analysis of very large data sets and, hence, on aspects of scalability and efficiency.

Despite these slightly different goals, the typical KDD process has much in common with the process of inductive reasoning as outlined above, except for the fact that the former can be (and indeed often is) circular in the sense that the data mining results will retroact on the acquisition, selection, and preparation of the data, possibly initiating a repeated pass with modified data, analysis tools, or queries. A typical KDD process may comprise the following steps:

- data cleaning
- data integration (combination of multiple sources)
- data selection
- data transformation (into a form suitable for the analysis)
- data mining
- evaluation of patterns
- knowledge presentation

Recently, the interest in data mining has shifted from the analysis of large but homogeneous data sets (relational tables) to the analysis of more complex and heterogeneous information sources such as, e.g., texts, images, audio and video data, and the term *information mining* has been coined to describe a KDD process focused on this type of information sources [34].

There are several other fields that are closely related to machine learning and data mining such as, e.g., classical statistics and various forms of data analysis (distinguished by adjectives like multivariate, exploratory, Bayesian, intelligent, ...) Needless to say, it is impossible to set a clear boundary between these fields. Subsequently, we shall simply subsume them under the heading "machine learning and data mining" (ML&DM),[3] understood in a wide sense as the application of computational methods and algorithms for extracting models and patterns from potentially very large data sets.

## 3 Typical Applications of Fuzzy Set Theory

The tools and technologies that have been developed in FST have the potential to support all of the steps that comprise a process of model induction or knowledge discovery. In particular, FST can already be used in the data selection and preparation phase, e.g., for modeling vague data in terms of fuzzy sets [47], to "condense"

[3] Our distinction between machine learning and data mining can roughly be seen as a "modern" or extended distinction between descriptive and inductive statistics. We note, however, that this view is not an *opinio communis*. For example, some people have an even more general view of data mining that includes machine learning as a special case.

several crisp observations into a single fuzzy one, or to create fuzzy summaries of the data [35]. As the data to be analyzed thus becomes fuzzy, one subsequently faces a problem of *fuzzy data* analysis [2].

The problem of analyzing fuzzy data can be approached in at least two principally different ways. First, standard methods of data analysis can be extended in a rather generic way by means of an extension principle. For example, the functional relation between the data points and the coefficients of a linear regression function can be extended to the case of fuzzy data, where the observations are described in terms of fuzzy sets. Thus, the coefficients become fuzzy as well. A second, often more sophisticated approach is based on embedding the data into more complex mathematical spaces, such as fuzzy metric spaces [13], and to carry out data analysis in these spaces [14].

If fuzzy methods are not used in the data preparation phase, they can still be employed in a later stage in order to analyze the original data. Thus, it is not the data to be analyzed that is fuzzy, but rather the methods used for analyzing the data (in the sense of resorting to tools from FST). Subsequently, we shall focus on this type of fuzzy data analysis (where the adjective "fuzzy" refers to the term *analysis*, not to the term *data*), which is predominant in ML&DM. In the following, we focus on fuzzy extensions of some well-known machine learning and data mining methods without repeating the original methods themselves; thus, we assume basic familiarity with these methods.

## *3.1 Fuzzy Cluster Analysis*

In conventional clustering, every object is assigned to one cluster in an unequivocal way. Consequently, the individual clusters are separated by sharp boundaries. In practice, such boundaries are often not very natural or even counterintuitive. Rather, the boundary of single clusters and the transition between different clusters are usually "smooth" rather than abrupt. This is the main motivation underlying fuzzy extensions to clustering algorithms [26]. In fuzzy clustering an object may belong to different clusters at the same time, at least to some extent, and the degree to which it belongs to a particular cluster is expressed in terms of a fuzzy membership. The membership functions of the different clusters (defined on the set of observed points) is usually assumed to form a partition of unity. This version, often called probabilistic clustering, can be generalized further by weakening this constraint: In possibilistic clustering, the sum of membership degrees is constrained to be *at least* one [33]. Fuzzy clustering has proved to be extremely useful in practice and is now routinely applied also outside the fuzzy community (e.g., in recent bioinformatics applications [24]).

## *3.2 Learning Fuzzy Rule Bases*

The most frequent application of FST in machine learning is the induction or the adaptation of rule-based models. This is hardly astonishing, since rule-based models have always been a cornerstone of fuzzy systems and a central aspect of research in the field, not only in ML&DM but also in many other subfields, notably approximate reasoning and fuzzy control. (The terms *fuzzy system* and *fuzzy rule base* are sometimes even used synonymously.)

Fuzzy rule bases can represent both classification and regression functions, and different types of fuzzy models have been used for these purposes. In order to realize a regression function, a fuzzy system is usually wrapped in a "fuzzifier" and a "defuzzifier": The former maps a crisp input to a fuzzy one, which is then processed by the fuzzy system, and the latter maps the (fuzzy) output of the system back to a crisp value. For so-called Takagi-Sugeno models, which are quite popular for modeling regression functions, the defuzzification step is unnecessary, since these models output crisp values directly.

In the case of classification learning, the consequent of single rules is usually a class assignment (i.e. a singleton fuzzy set).[4] Evaluating a rule base (*à la* Mamdani-Assilan) thus becomes trivial and simply amounts to "maximum matching", that is, searching the maximally supporting rule for each class. Thus, much of the appealing interpolation and approximation properties of fuzzy inference gets lost, and fuzziness only means that rules can be activated to a certain degree. There are, however, alternative methods which combine the predictions of several rules into a classification of the query [8].

A plethora of strategies has been developed for inducing a fuzzy rule base from the data given, and we refrain from a detailed exposition here. Especially important in the field of fuzzy rule learning are hybrid methods that combine FST with other methodologies, notably evolutionary algorithms and neural networks. For example, evolutionary algorithms are often used in order to optimize ("tune") a fuzzy rule base or for searching the space of potential rule bases in a (more or less) systematic way [9]. Quite interesting are also *neuro-fuzzy* methods [38]. For example, one idea is to encode a fuzzy system as a neural network and to apply standard methods (like backpropagation) in order to train such a network. This way, neuro-fuzzy systems combine the representational advantages of fuzzy systems with the flexibility and adaptivity of neural networks.

### 3.3 Fuzzy Decision Tree Induction

Fuzzy variants of decision tree induction have been developed for quite a while (e.g. [32, 48]) and seem to remain a topic of interest even today (see [39] for a recent approach and a comprehensive overview of research in this field). In fact, these approaches provide a typical example for the "fuzzification" of standard machine learning methods. In the case of decision trees, it is primarily the "crisp" thresholds used for defining splitting predicates (constraints), such as e.g. `size` $\leq 181$, at inner nodes that have been criticized: Such thresholds lead to hard decision boundaries in

[4] More generally, a rule consequent can suggest different classes with different degrees of certainty.

the input space, which means that a slight variation of an attribute (e.g. `size` = 182 instead of `size` = 181) can entail a completely different classification of an object (e.g., of a person characterized by size, weight, gender, ...) Moreover, the learning process becomes unstable in the sense that a slight variation of the training examples can change the induced decision tree drastically.

In order to make the decision boundaries "soft", an obvious idea is to apply fuzzy predicates at the inner nodes of a decision tree such as, e.g., `size` $\in$ `TALL`, where `TALL` is a fuzzy set (rather than an interval). In other words, a fuzzy partition instead of a crisp one is used for the splitting attribute (here `size`) at an inner node. Since an example can satisfy a fuzzy predicate to a certain degree, the examples are partitioned in a fuzzy manner as well. That is, an object is not assigned to exactly one successor node in a unique way, but perhaps to several successors with a certain degree. For example, a person whose size is 181 cm could be an element of the `TALL`-group to the degree, say, 0.7 and of the complementary group to the degree 0.3.

The above idea of "soft recursive partitioning" has been realized in different ways. Moreover, the problems entailed by corresponding fuzzy extensions have been investigated. For example, how can splitting measures like entropy, originally defined for ordinary sets of examples, be extended to fuzzy sets of examples [10]? Or, how can a new object be classified by a fuzzy decision tree?

### *3.4 Fuzzy Association Analysis*

The use of fuzzy sets in connection with association analysis has been proposed by numerous authors (see [6, 11] for recent overviews), with motivations closely resembling those in the case of rule learning and decision tree induction. Again, by allowing for "soft" rather than crisp boundaries of intervals, fuzzy sets can avoid certain undesirable threshold effects [44], this time concerning the quality measures of association rules (like support and confidence) rather than the classification of objects. Moreover, identifying fuzzy sets with linguistic terms allows for a comprehensible and user-friendly presentation of rules discovered in a database.

Many standard techniques for association rule mining have been transferred to the fuzzy case, sometimes in a rather ad-hoc manner. Indeed, publications on this topic are often more concerned with issues of data preprocessing, e.g., the problem of finding good fuzzy partitions for the quantitative attributes, rather than the rule mining process itself. Still, more theoretically-oriented research has recently been started [18]. For example, the existence of different types of fuzzy rules [19] suggests that fuzzy associations might be interpreted in different ways and, hence, that the evaluation of an association cannot be independent of its interpretation. In particular, one can raise the question which generalized logical operators can reasonably be applied in order to evaluate fuzzy associations, e.g., whether the antecedent part and the condition part should be combined in a conjunctive way (à la Mamdani rules) or by means of a generalized implication (as in implication-based fuzzy rules) [27]. Moreover, since standard evaluation measures for association rules can be generalized in many ways, it is interesting to investigate properties of particular generalizations and to look for an axiomatic basis that supports the choice of specific measures [18].

## *3.5 Fuzzy Methods in Case-Based Learning*

The major assumption underlying case-based learning (CBL) is a commonsense principle suggesting that "similar problems have similar solutions". This "similarity hypothesis" serves as a basic inference paradigm in various domains of application. For example, in a classification context, it translates into the assertion that "similar objects have similar class labels". Similarity-based inference has also been a topic of interest in FST, which is hardly astonishing since similarity is one of the main semantics of fuzzy membership degrees [41, 43]. Along these lines, a close connection between case-based learning and fuzzy rule-based reasoning has been established in [16, 17]. Here, the aforementioned "similarity hypothesis" has been formalized within the framework of fuzzy rules. As a result, case-based inference can be realized as a special type of fuzzy set-based approximate reasoning.

A possibilistic variant of the well-known $k$-nearest neighbor classifier, which constitutes the core of the family of CBL algorithms, has been presented in [29]. Among other things, this paper emphasizes the ability of possibility theory to represent partial ignorance as a special advantage in comparison to probabilistic approaches. In fact, this point seems to be of critical importance in case-based learning, where the reliability of a classification strongly depends on the existence of cases that are similar to the query.

The use of OWA-operators as generalized aggregation operators in case-based learning has been proposed in [49]. In fact, there are several types of aggregation problems that arise in CBL. One of these problems concerns the derivation of a global degree of similarity between cases by aggregating *local* similarity degrees pertaining to individual (one-dimensional) attributes. Usually, this is done by means of a simple linear combination, and this is where OWA-operators provide an interesting, more flexible alternative. A second aggregation problem in CBL concerns the combination of the evidences in favor of different class labels that come from the neighbors of the query case. In [30], it is argued that cases retrieved from a case library must not be considered as independent information sources, as implicitly done by most case-based learning methods. To take interdependencies between the neighbored cases into account, a new inference principle is developed that combines potentially interacting pieces of evidence by means of the (discrete) Choquet-integral. This method can be seen as a generalization of weighted nearest neighbor estimation.

## *3.6 Possibilistic Networks*

So-called graphical models, including Bayesian networks [40] and Markov networks [36], have been studied intensively in recent years. The very idea of such models is to represent a high-dimensional probability distribution (defined on the Cartesian product of the domains of all attributes under consideration) in an efficient way, namely by factorizing it into several low-dimensional conditional or marginal distributions.

By their very nature, graphical models of the above kind provide a suitable means for representing *probabilistic* uncertainty. However, they cannot easily deal with other types of uncertainty such as imprecision or incompleteness. This has motivated the development of *possibilistic networks* as a possibilistic counterpart to probabilistic networks [3]. This approach relies upon possibility theory as an underlying uncertainty calculus, which makes it particularly suitable for dealing with imprecise data (in the form of set-valued specifications of attribute values). In this approach, the interpretation of possibility distributions is based on the so-called context model [25], hence possibility degrees are considered as a kind of upper probability.

## 4 Potential Contributions of Fuzzy Set Theory

In the following, we highlight some potential contributions that FST can make to machine learning and data mining.

### *4.1 Graduality*

The ability to represent gradual concepts and fuzzy properties in a thorough way is one of the key features of fuzzy sets. This aspect is also of primary importance in the context of ML&DM. In machine learning, for example, the formal problem of *concept learning* has received a great deal of attention. A concept is usually identified with its extension, that is a subset $C$ of an underlying set (universe) $U$ of objects. For example, $C$ might be the concept “dog” whose extension is the set of dogs presently alive, a subset of all creatures on earth. The goal of (machine) learning is to induce an *intensional* description of a concept from a set of (positive and negative) examples, that is a characterization of a concept in terms of its properties (a dog has four legs and a tail, it can bark, ...). Now, it is widely recognized that most natural concepts have non-sharp boundaries. To illustrate, consider concepts like woods, river, lake, hill, street, house, or chair. Obviously, these concepts are vague or fuzzy, in that one cannot unequivocally say whether or not a certain collection of trees should be called a wood, whether a certain building is really a house, and so on. Rather, one will usually agree only to a certain extent that an object belongs to a concept. Thus, an obvious idea is to induce *fuzzy concepts*, that are formally identified by a fuzzy rather than a crisp subset of $U$. Fuzzy concepts can be characterized in terms of fuzzy predicates (properties) which are combined by means of generalized logical connectives. In fact, one should recognize that graduality is not only advantageous for expressing the concept itself, but also for modeling the qualifying properties. For example, a “firm ground” is a characteristic property of a street, and this property is obviously of a fuzzy nature (hence it should be formalized accordingly).

Likewise, in data mining, the patterns of interest are often vague and have boundaries that are non-sharp in the sense of FST. To illustrate, consider the concept of a "peak": It is usually not possible to decide in an unequivocal way whether a timely ordered sequence of measurements (e.g., the expression profile of a gene in a microarray experiment, to mention one of the topical application areas of fuzzy data mining) has a "peak" (a particular kind of pattern) or not. Rather, there is a gradual transition between having a peak and not having a peak. Taking graduality into account is also important if one must decide whether a certain property is frequent among a set of objects, e.g., whether a pattern occurs frequently in a data set. In fact, if the pattern is specified in an overly restrictive manner, it might easily happen that none of the objects matches the specification, even though many of them can be seen as approximate matches. In such cases, the pattern might still be considered as "well-supported" by the data.

Unfortunately, the representation of graduality is often foiled in machine learning applications, especially in connection with the learning of predictive models. In such applications, a fuzzy prediction is usually not desired, rather one is forced to come up with a definite final decision. Classification is an obvious example: Usually, a decision in favor of one particular class label has to be made, even if the object under consideration seems to have partial membership in several classes simultaneously. This is the case both in theory and practice: In practice, the bottom line is the course of action (e.g., the choice among a set of applicants) one takes on the basis of a prediction, not the prediction itself. In theory, a problem concerns the performance evaluation of a fuzzy classifier: The standard benchmark data sets (e.g., those from the UCI repository or the StatLib archive[5]) have crisp rather than fuzzy labels. Moreover, a fuzzy classifier cannot be compared with a standard (non-fuzzy) classifier unless it eventually outputs crisp predictions.

Needless to say, if a fuzzy predictor is supplemented with a "defuzzification" mechanism (like a winner-takes-all strategy in classification), many of its merits are lost. In the classification setting, for instance, a defuzzified fuzzy classifier does again produce hard decision boundaries in the input space. Thereby, it is actually reduced to a standard classifier.

Here is an example often encountered in the literature: Suppose the premise of a classification rule to be a conjunction of antecedents of the form $x_i \in A_i$, where $x_i$ is an attribute value and $A_i$ a fuzzy set, and let the rules be combined in a disjunctive way. Moreover, let the consequent of a rule be simply a class assignment. If the standard minimum and maximum operators are used, respectively, as a generalized logical conjunction and disjunction, it is easy to see that the classifier thus obtained induces axis-parallel decision boundaries in the input space, and that the same boundaries can be produced by means of interval-based instead of fuzzy rules.

If a classifier is solely evaluated on the basis of its predictive accuracy, then all that matters is the decision boundaries it produces in the input space. Since a defuzzified fuzzy classifier does not produce a decision boundary that is principally different from the boundaries produced by alternative classifiers (such as decision

[5] `http://www.ics.uci.edu/~mlearn`, `http://stat.cmu.edu/`

trees or neural networks), fuzzy machine learning methods don't have much to offer with regard to generalization performance. Indeed, fuzzy approaches to classification do usually *not* improve predictive accuracy.

Let us finally note that "graduality" is of course not reserved to fuzzy methods. Rather, it is inherently present also in many conventional learning methods. Consider, for example, a concept learner (binary classifier) $c : \mathcal{X} \longrightarrow [0, 1]$ the output of which is a number in the unit interval, expressing a kind of "propensity" of an input $x$ to the concept under consideration. Classifiers of such kind abound, a typical example is a multilayer perceptron. In order to extend such classifiers to multi-class problems (involving more than two classes), one common approach is to apply a one-against-all strategy: For each class $y$, a separate classifier $c_y(\cdot)$ is trained which considers that class as the concept to be learned and, hence, instances of all other classes as negative examples. The prediction for a new input $x$ is then given by the class that maximizes $c_y(x)$. Now, it is of course tempting to consider the $c_y(x)$ as (estimated) membership degrees and, consequently, the collection $\{c_y(x) \mid y \in \mathcal{Y}\}$ of these estimations as a fuzzy classification.

## *4.2 Interpretability*

A primary motivation for the development of fuzzy sets was to provide an interface between a numerical scale and a symbolic scale which is usually composed of linguistic terms. Thus, fuzzy sets have the capability to interface quantitative patterns with qualitative knowledge structures expressed in terms of natural language. This makes the application of fuzzy technology very appealing from a knowledge representational point of view. For example, it allows association rules discovered in a database to be presented in a linguistic and hence comprehensible way. In fact, the user-friendly representation of models and patterns is often emphasized as one of the key features of fuzzy methods.

The use of linguistic modeling techniques does also produce some disadvantages, however. A first problem concerns the interpretation of fuzzy models: Linguistic terms and, hence, models are highly subjective and context-dependent. It is true that the imprecision of natural language is not necessarily harmful and can even be advantageous.[6] A fuzzy controller, for example, can be quite insensitive to the concrete mathematical translation of a linguistic model. One should realize, however, that in fuzzy control the information flows in a reverse direction: The linguistic model is not the end product, as in ML&DM, it rather stands at the beginning.

It is of course possible to disambiguate a model by complementing it with the semantics of the fuzzy concepts it involves (including the specification of membership functions). Then, however, the complete model, consisting of a qualitative (linguistic) and a quantitative part, becomes cumbersome and will not be easily understandable. This can be contrasted with interval-based models, the most obvious

[6] See Zadeh's principle of incompatibility between precision and meaning [50].

alternative to fuzzy models: Even though such models do certainly have their shortcomings, they are at least objective and not prone to context-dependency.

Another possibility to guarantee transparency of a fuzzy model is to let a user of a data mining system specify all fuzzy concepts by hand, including the fuzzy partitions for all of the variables involved in the study under consideration. This is rarely done, however, mainly for two reasons. Firstly, the job is of course tedious and cumbersome if the number of variables is large. Secondly, much flexibility for model adaptation is lost, because it is by no means guaranteed that accurate predictive models or interesting patterns can be found on the basis of the fuzzy partitions as pre-specified by the user. In fact, in most methods the fuzzy partitions are rather *adapted* to the data in an optimal way, so as to maximize the model accuracy or the interestingness of patterns.

A second problem with regard to transparency concerns the complexity of models. A rule-based classifier consisting of, say, 40 rules each of which has a condition part with 5–7 antecedents, will hardly be comprehensible as a whole, even if the various ingredients might be well understandable. Now, since models that are simple, e.g., in the sense of including only a few attributes or a few rules, will often not be accurate at the same time, there is obviously a conflict between accuracy and understandability and, hence, the need to find a tradeoff between these criteria [5].

In fact, this tradeoff concerns not only the size of models, but also other measures that are commonly employed in order to improve model accuracy. In connection with rule-based models, for example, the *weighing* of individual rules can often help to increase the predictive accuracy. On the other hand, the interpretation of a set of weighted rules becomes more difficult.

## 4.3 Robustness

It is often claimed that fuzzy methods are more robust than non-fuzzy methods. Of course, the term "robustness" can refer to many things, e.g., to the sensitivity of an induction method towards violations of the model assumptions.[7] In connection with fuzzy methods, the most relevant type of robustness concerns sensitivity towards variations of the data. Generally, a learning or data mining method is considered robust if a small variation of the observed data does hardly alter the induced model or the evaluation of a pattern.[8]

A common argument supporting the claim that fuzzy models are in this sense more robust than non-fuzzy models refers to the already mentioned "boundary effect", which occurs in various variants and is arguably an obvious drawback of interval-based methods. In fact, it is not difficult to construct convincing demonstrations of this effect: In association analysis (cf. Sect. 3.4), for example, a small

[7] This type of sensitivity is of special interest in robust statistics.

[8] Note that we speak about robustness of the learning algorithm (that takes a set of data as input and outputs a model), not about robustness of the induced model (that takes instances as input and outputs, say, a classification).

shift of the boundary of an interval can have a drastic effect on the support of a fuzzy association rule if many data points are located near the boundary. This effect is alleviated when using fuzzy sets instead of intervals.

Unfortunately, such examples are often purely artificial and, hence, of limited practical relevance. Moreover, there is no clear conception of the concrete meaning of *robustness*. Needless to say, without a formal definition of robustness, i.e., certain types of robustness measures, one cannot argue convincingly that one data mining method is more robust than another one. For example, it makes a great difference whether robustness is understood as a kind of *expected* or a kind of *worst-case* sensitivity: It is true that a shifting of data points can have a stronger effect on, say, the support of an interval-based association rule than on the support of a fuzzy association. However, if the data points are not located at the boundary region of the intervals, it can also happen that the former is not affected at all, whereas a fuzzy rule is almost always affected at least to some extent (since the "boundary" of a fuzzy interval is much wider than that of a standard interval). Consequently, if robustness is defined as a kind of *average* rather than *maximal* sensitivity, the fuzzy approach might not be more robust than the non-fuzzy one.

## 4.4 Representation of Uncertainty

Machine learning is inseparably connected with uncertainty. To begin with, the data presented to learning algorithms is imprecise, incomplete or noisy most of the time, a problem that can badly mislead a learning procedure. But even if observations are perfect, the generalization beyond that data, the process of induction, is still afflicted with uncertainty. For example, observed data can generally be explained by more than one candidate theory, which means that one can never be sure of the truth of a particular model.

Fuzzy sets and possibility theory have made important contributions to the representation and processing of uncertainty. In ML&DM, like in other fields, related uncertainty formalisms can complement probability theory in a reasonable way, because not all types of uncertainty relevant to machine learning are probabilistic and because other formalisms are more expressive than probability.

To illustrate the first point, consider the problem of inductive reasoning as indicated above: In machine learning, a model is often induced from a set of data on the basis of a *heuristic* principle of inductive inference such as, e.g., the well-known Occams's razor. As one can never be sure of the truth of the particular model suggested by the heuristic principle, it seems reasonable to specify a kind of *likelihood* for all potential candidate models. This is done, e.g., in Bayesian approaches, where the likelihood of models is characterized in terms of a posterior probability distribution (probability of models given the data). One can argue, however, that the uncertainty produced by heuristic inference principles such as Occam's razor is not necessarily of a probabilistic nature and, for example, that the derivation of a *possibility distribution* over the model space is a viable alternative. This idea has been suggested in [28] in connection with decision tree induction: Instead of learning a single decision tree, a possibility distribution over the class of all potential trees is derived on the basis of a possibilistic variant of Occam's razor.

The second point, concerning the limited expressivity of probability distributions, was already indicated in Sect. 3.5, where we mentioned that possibility distributions are more suitable for representing partial ignorance in case-based learning. Similarly, possibility theory is used for modeling incomplete and missing data in possibilistic networks (cf. Sect. 3.6).

Finally, we note that apart from possibility theory, other formalisms can be used to model various forms of uncertainty and incomplete information in learning from data. For example, belief functions have been extensively employed in this connection (e.g. [12, 20]).

### *4.5 Incorporation of Background Knowledge*

Roughly speaking, inductive learning can be seen as searching the space of candidate hypotheses for a most suitable model. The corresponding search process, regardless whether it is carried out in an explicit or implicit way, is usually "biased" in various ways, and each bias usually originates from a sort of background knowledge. For example, the *representation bias* restricts the hypothesis space to certain types of input-output relations such as, e.g., linear or polynomial relationships. Incorporating background knowledge is extremely important, because the data by itself would be totally meaningless if considered from an "unbiased" point of view [37].

Fuzzy set-based modeling techniques provide a convenient tool for making expert knowledge accessible to computational methods and, hence, to incorporate background knowledge in the learning process. Here, we briefly outline two possibilities.

One very obvious approach is to combine rule-based modeling and learning. For example, an expert can describe an input-output relation in terms of a fuzzy rule base (as in fuzzy control). Afterwards, the membership functions specifying the linguistic terms that have been employed by the expert can be adapted to the data in an optimal way.[9] In other words, the expert specifies the rough structure of the rule-based model, while the fine-tuning is done in a data-driven way. Let us note that specifying the structure of a model first and adapting that structure to the data afterwards is a general strategy for combining knowledge-based and data-driven modeling, which is not reserved to rule-based models; it is used, for example, in graphical models (cf. Sect. 3.6) as well.

An alternative approach, called constraint-regularized learning, aims at exploiting fuzzy set-based modeling techniques within the context of the regularization (penalization) framework of inductive learning [31]. Here, the idea is to express vague, partial knowledge about an input-output relation in terms of fuzzy constraints

[9] Here, the expert implements a kind of *search bias*, as it determines the starting point of the search process and, hence, the first local optimum to be found.

and to let such constraints play the role of a penalty term within the regularization approach. Thus, an optimal model is one that achieves an optimal tradeoff between fitting the data and satisfying the constraints.

### *4.6 Generalized Aggregation Operators*

Many ML&DM methods make use of logical and arithmetical operators for representing relationships between attributes in models and patterns. In decision tree induction, for example, each inner node represents an equality or an inequality predicate, and these predicates are combined in a conjunctive way along a path of a tree. In nearest neighbor classification, each neighbor provides a certain amount of evidence in favor of the class it belongs to. To make a final decision, this evidence must be aggregated either way, which in the standard approach is done by simply adding them up.

Now, a large repertoire of generalized logical (e.g., t-norms and t-conorms) and arithmetical (e.g., Choquet- and Sugeno-integral) operators have been developed in FST and related fields. Thus, a straightforward way to extend standard learning methods consists of replacing standard operators by their generalized versions. In fact, several examples of this idea have been presented in previous sections.

The general effect of such generalizations is to make models more flexible. For example, while a standard decision tree can only produce axis-parallel decision boundaries, these boundaries can become non-axis-parallel for fuzzy decision trees where predicates are combined by means of a t-norm. Now, it is well-known that learning from empirical data will be most successful if the model class under consideration has just the right flexibility, since both over- and underfitting of a model can best be avoided in that case. Therefore, the question whether a fuzzy generalization will pay off cannot be answered in general: If the original (non-fuzzy) hypothesis space is not flexible enough, the fuzzy version will probably be superior. On the other hand, if the former is already flexible enough, a fuzzification might come along with a danger of overfitting.

## 5 Conclusions

All things considered, it is beyond question that FST has the potential to contribute to machine learning and data mining in various ways. In fact, the previous sections have shown that substantial contributions have already been made. Yet, our remarks also suggest that much scope for further developments is still left. According to our opinion, however, it is very important to focus on the right issues, that is to say, to concentrate more on the strengths and distinctive features of FST.

In particular, we doubt that FST will be very conducive to *generalization performance* and *model accuracy*, albeit the latter is still the dominant quality criterion in machine learning research. This somewhat sceptical view has at least two reasons: Firstly, after several years of intensive research the field of machine learning has reached a somewhat mature state, and a large repertoire of quite sophisticated learning algorithms is now available. Regarding predictive accuracy, a significant improvement of the current quality level can hardly be expected.

Secondly, and perhaps more importantly, FST does not seem to offer fundamentally new concepts or induction principles for the design of learning algorithms, comparable, e.g., to the ideas of *resampling* and *ensemble learning* [15] (like bagging [4] and boosting [22]) or the idea of *margin maximization* underlying discriminative learning methods like SVMs [42], that might raise hope for an improved generalization performance. As mentioned above, even though fuzzifying standard learning methods, e.g., by using fuzzy partitions of numeric attributes or generalized logical and arithmetical operators, can have an effect on the decision boundaries of a classifier or the regression function produced in the case of numeric prediction, the gain in predictive accuracy is mostly insignificant.

In this connection, we also like to question a current research trend in the FST community. It seems that the shift from (knowledge-driven) modeling to (data-driven) learning, as signified in Sect. 1, comes along with a tendency to view fuzzy systems as pure function approximators. In fact, in many recent publications fuzzy sets simply serve as a special kind of basis or kernel function.[10] Thus, there is a high danger of losing sight of the original ideas and intentions of FST, and to produce another type of "black box" approach instead. Truly, renaming a basis function a "fuzzy set" does not mean that a model will suddenly become comprehensible!

Rather than suggesting new solutions to problems for which alternative methods from established fields such as, e.g., approximation theory, neural networks, and machine learning, will probably be more successful, more emphasis should be put on the distinguished features of FST. In this connection, let us highlight the following points:

1. FST has the potential to produce models that are more comprehensible, less complex, and more robust.
2. FST, in conjunction with possibility theory, can contribute considerably to the modeling and processing of various forms of uncertain and incomplete information.
3. Fuzzy methods appear to be particularly useful for data pre- and post-processing.

Concerning the first point, our critical comments in previous sections have shown that, despite of the high potential, many questions are still open. For example, notions like "comprehensibility", "simplicity", or "robustness" still lack an underlying formal theory including a quantification of their intuitive meaning in terms of universally accepted measures. This is probably one of the reasons why model accuracy is still regarded as a more concrete and, hence, more important quality criterion. Anyway, we think that the tradeoff between accuracy on the one side and competitive criteria like interpretability, simplicity, and robustness on the other side is an issue of central importance and a problem to that FST can contribute in a substantial

[10] In [7], for example, a support vector machine is trained and then turned into a fuzzy rule base by identifying each support vector with a fuzzy rule.

way. In fact, fuzzy information granulation appears to be an ideal tool for trading off accuracy against complexity and understandability of models. Of course, a necessary prerequisite for studying this tradeoff in a more rigorous way and, hence, a challenge for future research, is a better understanding and formalization of these alternative criteria.

The second point refers to an aspect that is of primary importance in ML&DM, and that has already been touched on in Sect. 4.4. Meanwhile, the coexistence of various forms of uncertainty, not all of which can be adequately captured by probability theory, has been widely recognized. Still, in machine learning, and more generally in the AI community, fuzzy sets and related uncertainty calculi have not yet obtained a proper acceptance. This situation might be further impaired by the increasing popularity of probabilistic methodology, which in machine learning can mainly be ascribed to the success of *statistical learning theory* [46] as a solid foundation of empirical learning, and in AI to the general acceptance of the Bayesian framework for knowledge representation and reasoning under uncertainty. For the FST community, it is all the more important to show that alternative uncertainty formalisms can complement probability theory in a reasonable way.

Concerning the third point, we feel that this research direction has not received enough attention so far. In fact, even though FST seems to be especially qualified for data pre- and postprocessing, e.g., for data summarization and reduction, approximation of complex and accurate models, or the (linguistic) presentation of data mining results, current research is still more focused on the inductive reasoning or data mining process itself. In this respect, we see a high potential for further developments, especially against the background of the current trend to analyze complex and heterogeneous information sources that are less structured than standard relational data tables.

Finally, there are some other research directions that are worth further exploration. For example, so far most of the work in the FST community has been *methodologically* oriented, focusing on the fuzzy extension of standard learning methods, whereas both the *experimental* validation and the *theoretical* analysis of fuzzy machine learning methods have received much less attention. As mentioned above, validating the predictive performance of a fuzzy method in an empirical way is not as easy, since fuzzy labels for comparison are rarely available in practice. What is a good fuzzy prediction? This question naturally arises if fuzzy predictions are not defuzzified, and it becomes even more intricate if predictions are expressed in terms of still more complex uncertainty formalisms such as, e.g., lower and upper possibility bounds, type-II fuzzy sets, or belief functions. Regarding theoretical analyses of fuzzy learning methods, it would be interesting to investigate whether fuzzy extensions are profitable from a theoretical point of view. For example, is it possible that a class of concepts is, say, PAC-learnable[11] by the fuzzy extension of a learning algorithm but not by the original version? Such results would of course be highly welcome as a formal justification of fuzzy learning methods.

[11] The PAC (probably approximately correct) learning framework is a well-known formal model of inductive learning [45].

## References


1. R. Babuska. *Fuzzy Modeling for Control*. Kluwer Academic Publ., Boston, 1998.
2. H. Bandemer and W. Näther. *Fuzzy Data Analysis*. Kluwer Academic Publishers, Dordrecht, 1992.
3. C. Borgelt and R. Kruse. *Graphical Models – Methods for Data Analysis and Mining*. Wiley, Chichester, 2002.
4. L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
5. J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, editors. *Interpretability Issues in Fuzzy Modeling*. Springer-Verlag, Berlin, 2003.
6. G. Chen, Q. Wei, E. Kerre, and G. Wets. Overview of fuzzy associations mining. In *Proc. ISIS–2003*. Jeju, Korea, September 2003.
7. Y. Chen and JZ. Wang. Support vector learning for fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 11(6):716–728, 2003.
8. O. Cordon, MJ. del Jesus, and F. Herrera. Analyzing the reasoning mechanisms in fuzzy rule based classification systems. *Mathware & Soft Computing*, 5:321–332, 1998.
9. O. Cordon, F. Gomide, F. Herrera, and F. Hoffmann anf L. Magdalena. Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems*, 141(1):5–31, 2004.
10. TH. Dang, B. Bouchon-Meunier, and C. Marsala. Measures of information for inductive learning. In *Proc. IPMU-2004*, Perugia, Italy, 2004.
11. M. Delgado, N. Marin, D. Sanchez, and M. Vila. Fuzzy association rules: general model and applications. *IEEE Trans. on Fuzzy Syst.*, 11(2):214–225, 2003.
12. T. Denoeux. A k-nearest neighbor classification rule based on Dempster-Shafer Theory. *IEEE Transa. on Systems, Man, and Cybernetics*, 25(5):804–813, 1995.
13. P. Diamond and P. Kloeden. *Metric Spaces of Fuzzy Sets: Theory and Applications*. World Scientific, Singapur, 1994.
14. P. Diamond and H. Tanaka. Fuzzy regression analysis. In R. Slowinski, editor, *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, pp. 349–387. Kluwer, 1998.
15. TG. Dietterich. Ensemble methods in machine learning. In J. Kittler and F. Roli, editors, *First International Workshop on Multiple Classifier Systems*, Vol. 1857, pp. 1–15. Springer-Verlag, 2000.
16. D. Dubois, F. Esteva, P. Garcia, L. Godo, R. Lopez de Mantaras, and H. Prade. Fuzzy set modelling in case-based reasoning. *International Journal of Intelligent Systems*, 13:345–373, 1998.
17. D. Dubois, E. Hüllermeier, and H. Prade. Fuzzy set-based methods in instance-based reasoning. *IEEE Transactions on Fuzzy Systems*, 10(3):322–332, 2002.
18. D. Dubois, E. Hüllermeier, and H. Prade. A note on quality measures for fuzzy association rules. *Proc. IFSA–03*, pp. 677–648, Istambul, July 2003.
19. D. Dubois and H. Prade. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84:169–185, 1996.
20. Z. Elouedi, K. Mellouli, and P. Smets. Decision trees using the belief function theory. In *Proc. IPMU-2000*, Vol. I, pp. 141–148, Madrid, 2000.
21. UM. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*. MIT Press, 1996.
22. Y. Freund and RE. Schapire. Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, pp. 148–156, 1996.
23. J. Fürnkranz and E. Hüllermeier. Pairwise preference learning and ranking. In *Proc. ECML–2003*, Cavtat-Dubrovnik, Croatia, 2003. Springer-Verlag.
24. AP. Gasch and MB. Eisen. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biology*, 3(11):1–22, 2002.
25. J. Gebhardt and R. Kruse. A possibilistic interpretation of fuzzy sets by the context model. In *IEEE Intern. Conf. on Fuzzy Systems*, pp. 1089–1096, 1992.
26. F. Höppner, F. Klawonn, F. Kruse, and T. Runkler. *Fuzzy Cluster Analysis*. Wiley, Chichester, 1999.

27. E. Hüllermeier. Implication-based fuzzy association rules. *Proc. PKDD–01*, pp. 241–252, Freiburg, Germany, 2001. Springer-Verlag.
28. E. Hüllermeier. Possibilistic induction in decision tree learning. *Proc. ECML–02*, pp. 173–184, Helsinki, Finland, 2002. Springer-Verlag.
29. E. Hüllermeier. Possibilistic instance-based learning. *Artificial Intelligence*, 148(1–2): 335–383, 2003.
30. E. Hüllermeier. Cho-k-NN: A method for combining interacting pieces of evidence in case-based learning. *Proc. IJCAI–05*, pp. 3–8, Edinburgh, 2005.
31. E. Hüllermeier, I. Renners, and A. Grauel. An evolutionary approach to constraint-regularized learning. *Mathware & Soft Comp.*, 11:109–124, 2004.
32. CZ. Janikow. Fuzzy decision trees: Issues and methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 28(1):1–14, 1998.
33. R. Krishnapuram and J.M. Keller. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110, 1993.
34. R. Kruse and C. Borgelt. Information mining: Editorial. *Int. Journal of Approximate Reasoning*, 32:63–65, 2003.
35. A. Laurent. Generating fuzzy summaries: a new approach based on fuzzy multidimensional databases. *Intelligent Data Analysis Journal*, 7(2):155–177, 2003.
36. S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50(2):157–224, 1988.
37. T.M. Mitchell. The need for biases in learning generalizations. Technical Report TR CBM–TR–117, Rutgers University, 1980.
38. D. Nauck, F. Klawonn, and R. Kruse. *Foundations of Neuro-Fuzzy Systems*. Wiley and Sons, Chichester, UK, 1997.
39. C. Olaru and L. Wehenkel. A complete fuzzy decision tree technique. *Fuzzy Sets and Systems*, 138(2), 2003.
40. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.
41. E.H. Ruspini. Possibility as similarity: The semantics of fuzzy logic. In P.P. Bonissone, H. Henrion, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty In Artificial Intelligence 6*. Elsevier Science Publisher, 1991.
42. B. Schölkopf and AJ. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
43. T. Sudkamp. Similarity as a foundation for possibility. In *Proc. 9th IEEE Int. Conference on Fuzzy Systems*, pp. 735–740, San Antonio, 2000.
44. T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149(1), 2005.
45. LG. Valiant. A theory of the learnable. *CACM*, 17(11):1134–1142, 1984.
46. V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, second edition, 2000.
47. R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, Florida, 1996.
48. R. Weber. Fuzzy-ID3: a class of methods for automatic knowledge acquisition. In *Proc. IIZUKA-92*, Vol. 1, pp. 265–268. 1992.
49. R.R. Yager. Soft aggregation methods in case based reasoning. *Applied Intelligence*, 21:277–288, 2004.
50. L. Zadeh. New Approach to the Analysis of Complex Systems. *IEEE Trans. Syst., Man, Cybern., SMC*, 3(1), 1973.

# Pattern Classification with Linguistic Rules

**Hisao Ishibuchi and Yusuke Nojima**

**Abstract** Linguistic rules are fuzzy rules described by linguistic terms such as *small* and *large*. Here we discuss pattern classification with linguistic rules. The main advantage of using linguistic rules is their high interpretability. We can construct linguistically interpretable fuzzy rule-based classification systems using linguistic rules. First we briefly explain fuzzy rules for function approximation. Next we explain fuzzy rules and fuzzy reasoning for pattern classification. Then we explain linguistic rule extraction from numerical data. Finally we show some future research topics on pattern classification with linguistic rules.

## 1 Introduction

Fuzzy rule-based systems have been successfully implemented in various application areas since the 1970s [38]. In early studies on fuzzy control, fuzzy rules were extracted from human operators as linguistic rules. The extraction of linguistic rules, however, is often a troublesome task for human operators. A number of heuristic rule extraction methods have been proposed to design fuzzy rule-based systems from numerical data with no intervention by human experts (e.g., Nozaki et al. [54] and Wang & Mendel [62]). In the 1990s, learning ability of neural networks was incorporated into fuzzy rule-based systems to improve their accuracy in many studies (e.g., Horikawa et al. [22], Jang [42], Lin & Lee [30], and Nauck & Kruse [51]). Since fuzzy rule-based systems in these studies have network structures similar to neural networks, they are often called fuzzy neural networks or neuro-fuzzy systems. Neuro-fuzzy systems are still an active research area in the fields of neural networks and fuzzy systems [15, 21].

Global optimization ability of evolutionary computation was also incorporated into fuzzy rule-based systems in the 1990s (e.g., Homaifar & McCormick [46], Ishibuchi et al. [9], and Karr & Gentry [27]). Since genetic algorithms were usually used as evolutionary computation, fuzzy rule-based systems in these studies are often called genetic fuzzy systems [40, 12]. Whereas neuro-fuzzy systems usually optimize only continuous parameter values, genetic fuzzy systems can optimize various aspects of fuzzy rule-based systems including combinatorial optimization such as input selection, fuzzy partition and rule selection.

Emphasis was placed on accuracy improvement in neuro-fuzzy systems and genetic fuzzy systems in the 1990s. Good experimental results by neuro-fuzzy systems

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

and genetic fuzzy systems were reported with respect to their accuracy in the above-mentioned studies. The accuracy improvement, however, was achieved at the cost of an inherent advantage of fuzzy rule-based systems over other nonlinear systems such as neural networks. That is, the interpretability of fuzzy rule-based systems was often degraded by improving their accuracy in neuro-fuzzy systems and genetic fuzzy systems. This issue was pointed out in the late 1990s in some studies (e.g., Nauck & Kruse [52] and Setnes et al. [56]). Since there exists a tradeoff between the accuracy and the interpretability in the design of fuzzy rule-based systems, it is impossible to simultaneously optimize these two conflicting design criteria. Recently several approaches have been proposed to find a good compromise between them [2, 55, 57]. In these studies, genetic algorithms were used to optimize a scalarizing fitness function, which was defined by combining accuracy and interpretability measures into a single fitness function. Whereas the definition of an accuracy measure is straightforward in each application area (e.g., the mean squared error and the average error rate), it is not easy to define an interpretability measure. Various aspects are involved in the interpretability of fuzzy rule-based systems such as the number of fuzzy rules, the number of input variables, the granularity of the fuzzy partition for each variable, the number of antecedent conditions in each fuzzy rule, and the separability of adjacent antecedent fuzzy sets [34, 23, 45, 26, 39]. Some of these aspects were taken into account to find a good compromise between the accuracy and the interpretability in [2, 55, 57].

When we use a scalarizing fitness function defined by accuracy and complexity measures, a single fuzzy rule-based system is obtained by a standard single-objective genetic algorithm. The obtained fuzzy rule-based system strongly depends on the scalarizing fitness function. Its definition, however, is problematic. For example, it is difficult to assign relative importance to the average error rate and the number of fuzzy rules in the design of fuzzy rule-based classification systems. Multiobjective genetic algorithms are used in some studies on genetic fuzzy systems to circumvent the difficulty in the definition of the scalarizing fitness function [28, 44, 4, 3, 60, 61]. Multiobjective genetic fuzzy systems try to find a large number of non-dominated (i.e., Pareto-optimal) fuzzy rule-based systems with respect to the two conflicting design criteria: accuracy and interpretability. Obtained fuzzy rule-based systems help us to understand the accuracy-interpretability tradeoff structure in the design of fuzzy rule-based systems.

In this chapter, we discuss pattern classification with linguistic rules. First we briefly explain fuzzy rule-based systems for function approximation. Next we explain fuzzy rule-based classification systems. Then we explain linguistic rule extraction from numerical data for pattern classification. Finally we show some future research topics on pattern classification with linguistic rules.

## 2 Fuzzy Rule-Based Systems for Function Approximation

In this section, we briefly explain fuzzy rule-based systems for function approximation problems. Let us assume that we have $m$ input-output pairs $(\mathbf{x}_p, y_p)$,

$p = 1, 2, ..., m$ where $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$ is an $n$-dimensional input vector and $y_p$ is the corresponding output value. Our task in this section is to approximate an unknown input-output relation in the given numerical data $(\mathbf{x}_p, y_p)$, $p = 1, 2, ..., m$ using a fuzzy rule-based system.

Fuzzy rules of the following type were often used in early studies on fuzzy control since Mamdani [38]:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then } y \text{ is } B_q, \tag{1}$$

where $R_q$ is the label of the $q$th fuzzy rule, $\mathbf{x} = (x_1, x_2, ..., x_n)$ is an $n$-dimensional input vector, $A_{qi}$ is an antecedent fuzzy set $(i = 1, 2, ..., n)$, $y$ is an output variable, and $B_q$ is a consequent fuzzy set. Linguistic terms such as *small* and *large* are often associated with the antecedent and consequent fuzzy sets. An example of fuzzy rules of this type is "If $x_1$ is *small* and $x_2$ is *small* then $y$ is *large*". The main advantage of such a fuzzy rule is its clear linguistic interpretability.

Takagi & Sugeno [59] proposed the use of fuzzy rules of the following type with a linear function in their consequent part in the mid 1980s:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then } y = f_q(\mathbf{x}). \tag{2}$$

The main characteristic is that the consequent part is not defined by the consequent fuzzy set $B_q$ but the linear function $f_q(\mathbf{x})$ of the input vector $\mathbf{x}$:

$$f_q(\mathbf{x}) = b_{q0} + b_{q1}x_1 + ... + b_{qn}x_n, \tag{3}$$

where $b_{qi}$ is a real number $(i = 0, 1, ..., n)$. An example of fuzzy rules of this type is "If $x_1$ is *small* and $x_2$ is *small* then $y = 0.2 + 0.5x_1 - 0.3x_2$". Because Takagi-Sugeno fuzzy rules have high approximation ability of non-linear functions, they have been frequently used in various application areas in the literature. Due to the use of the consequent linear function, the linguistic interpretability of Takagi-Sugeno fuzzy rules in (2) is not high if compared with those in Mamdani model in (1).

The following simplified version of Takagi-Sugeno fuzzy rules has been also frequently used due to its simplicity in the literature:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then } y = b_q, \tag{4}$$

where $b_q$ is a consequent real number. An example of fuzzy rules of this type is "If $x_1$ is *small* and $x_2$ is *small* then $y = 0.236$". The linguistic interpretability of such a fuzzy rule is not high due to the difficulty in the linguistic interpretation of the consequent real number. The learning of fuzzy rules in (4) is straightforward in the framework of neuro-fuzzy systems. This is an advantage of fuzzy rules of this type over the other types.

Let us denote the membership function of the antecedent fuzzy set $A_{qi}$ as $A_{qi}(\cdot)$. The compatibility grade of the input vector $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$ with the antecedent part of the fuzzy rule $R_q$ in (1), (2) and (4) is computed using a t-norm. The following product and minimum operators are often used as a t-norm to calculate the compatibility grade:

$$\mathbf{A}_q(\mathbf{x}_p) = A_{q1}(x_{p1}) \cdot A_{q2}(x_{p2}) \cdot ... \cdot A_{qn}(x_{pn}), \tag{5}$$

$$\mathbf{A}_q(\mathbf{x}_p) = \min\{A_{q1}(x_{p1}), A_{q2}(x_{p2}), ..., A_{qn}(x_{pn})\}. \tag{6}$$

In this chapter, we use the product operator since it is more popular than the minimum operator in recently developed fuzzy rule-based systems.

When we have a fuzzy rule-based system with $N$ Takagi-Sugeno fuzzy rules of the form in (2), the estimated output value $g(\mathbf{x}_p)$ for the input vector $\mathbf{x}_p$ is calculated as the weighted average of the consequent linear function of each fuzzy rule as

$$g(\mathbf{x}_p) = \sum_{q=1}^{N} \mathbf{A}_q^*(\mathbf{x}_p) f_q(\mathbf{x}_p), \tag{7}$$

where

$$\mathbf{A}_q^*(\mathbf{x}_p) = \frac{\mathbf{A}_q(\mathbf{x}_p)}{\sum_{r=1}^{N} \mathbf{A}_r(\mathbf{x}_p)}, \quad q = 1, 2, ..., N. \tag{8}$$

When we use the simplified version of Takagi-Sugeno fuzzy rules in (4), the estimated output value $g(\mathbf{x}_p)$ is calculated in the same manner as

$$g(\mathbf{x}_p) = \sum_{q=1}^{N} \mathbf{A}_q^*(\mathbf{x}_p) b_q. \tag{9}$$

In the case of Mamdani fuzzy rules in (1), an additional defuzzification procedure is required to calculate the estimated output value $g(\mathbf{x}_p)$. If we use a representative real number $b_q$ for the consequent fuzzy set $B_q$ (i.e., if $b_q$ is the defuzzification of $B_q$), the calculation of the estimated output value $g(\mathbf{x}_p)$ using Mamdani fuzzy rules in (1) is exactly the same as (9).

Neuro-fuzzy systems adjust the antecedent and consequent parts of each fuzzy rule using a steepest descent learning algorithm to minimize the squared error:

$$E_p = \frac{1}{2}\{y_p - g(\mathbf{x}_p)\}^2. \tag{10}$$

## 3 Fuzzy Rule-Based Systems for Pattern Classification

In this section, we explain fuzzy rules and fuzzy reasoning for pattern classification. We use a special type of fuzzy rules with a class label in their consequent part instead of a consequent fuzzy set, a real number or a linear function. Of course, we can use fuzzy rules in Sect. 2 for pattern classification as well as function approximation. For example, Roubos & Setnes [55] and Setnes & Roubos [57] used an $n$-input and single-output fuzzy system with simplified Takagi-Sugeno fuzzy rules for an $n$-dimensional three-class problem. The target outputs were specified as 1, 2 and 3 for input patterns from Class 1, Class 2 and Class 3, respectively. It is also possible to use an $n$-input and three-output fuzzy system for the same three-class problem. In this case, the target vectors are specified as (1, 0, 0), (0, 1, 0) and (0, 0, 1) for input patterns from Class 1, Class 2 and Class 3, respectively. This specification of the target vectors has been often used in the application of neural networks to classification problems.

We use, however, a different type of fuzzy rules with a class label in their consequent part in this section. This is because such a fuzzy rule is more intuitive for classification problems and has been used in many studies on fuzzy rule-based classification systems (e.g., Abe & Lan [41], Abe & Thawonmas [50], Ishibuchi et al. [35], and Nauck & Kruse [51, 52]). For various fuzzy rules and fuzzy reasoning schemes for classification problems, see Cordon et al. [24], Ishibuchi et al. [43, 26] and Kuncheva [49].

## *3.1 Fuzzy Rules for Pattern Classification*

Fuzzy rules for an $n$-dimensional classification problem are written as

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then Class } C_q. \tag{11}$$

An example of such a fuzzy rule is "If $x_1$ is *small* and $x_2$ is *small* then Class 2". Fuzzy decision-trees have been used in many studies on fuzzy rule-based classification systems (e.g., [14, 39, 48, 63]). A fuzzy decision tree can be viewed as a set of fuzzy rules in (11).

In some studies (e.g., Gonzalez & Perez [20]), a disjunctive combination of multiple linguistic terms is used as an antecedent fuzzy set $A_{qi}$ such as "If $x_1$ is *small* or *medium* and $x_2$ is *medium* or *large* then Class 1". The disjunctive combination of all linguistic terms can be interpreted as *don't care*. For example, "*small* or *medium* or *large*" can be interpreted as *don't care* when each attribute is divided into three linguistic terms *small*, *medium* and *large*. In this case, the disjunctive combination "*small* or *medium*" can be interpreted as the negation of *large* (i.e., "not *large*"). It should be noted that *don't care* can be also directly used as a special antecedent fuzzy set. We further discuss the handling of *don't care* and disjunctive combinations of multiple linguistic terms in Sect. 4.2 of this chapter.

We also use the following fuzzy rules with a rule weight:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then Class } C_q \text{ with } w_q, \tag{12}$$

where $w_q$ is the rule weight of the fuzzy rule $R_q$. Fuzzy rules of this type have been frequently used in fuzzy rule-based systems from the early 1990s [35]. As shown in the next subsection, the rule weight of each fuzzy rule has a large effect on classification results [6, 37, 53].

Whereas fuzzy rules in (11) and (12) have a single-dimensional fuzzy set $A_{qi}$ for each attribute in their antecedent part, it is also possible to use a multi-dimensional antecedent fuzzy set as follows:

$$\text{Rule } R_q : \text{If } \mathbf{x} \text{ is } \mathbf{A}_q \text{ then Class } C_q, \tag{13}$$

where $\mathbf{x}$ is an $n$-dimensional input vector (i.e., $\mathbf{x} = (x_1, x_2, ..., x_n)$) and $\mathbf{A}_q$ is an $n$-dimensional antecedent fuzzy set in the $n$-dimensional pattern space. Fuzzy rules of this type have been often used in clustering-based rule generation methods (e.g., [41, 50, 51, 52, 55]). Whereas fuzzy rules in (13) have high classification accuracy, their interpretability is low due to the difficulty in the linguistic interpretation of the multi-dimensional antecedent fuzzy set $\mathbf{A}_q$. As we will explain later, linguistically interpretable fuzzy rules with single-dimensional antecedent fuzzy sets in (11) can be derived from fuzzy rules in (13) by projecting their multi-dimensional antecedent fuzzy sets onto each axis of the pattern space.

## *3.2 Fuzzy Reasoning for Pattern Classification*

First we explain fuzzy reasoning for pattern classification using fuzzy rules with no rule weights in (11). As in Sect. 2, let $\mathbf{A}_q(\mathbf{x}_p)$ be the compatibility grade of the input pattern $\mathbf{x}_p$ with the antecedent part of the fuzzy rule $R_q$. When we use a winner-take-all scheme, the maximum compatibility grade for each class is calculated as follows:

$$\alpha_h(\mathbf{x}_p) = \max\{\mathbf{A}_q(\mathbf{x}_p) \mid C_q = h; q = 1, 2, ..., N\},\ h = 1, 2, ..., M, \tag{14}$$

where $h$ is a class index and $M$ is the number of classes. When the input pattern $\mathbf{x}_p$ is to be classified, $\mathbf{x}_p$ is assigned to the class with the maximum value of $\alpha_h(\mathbf{x}_p)$ over the $M$ classes. In this sense, $\alpha_h(\mathbf{x}_p)$ can be viewed as a kind of discriminant function for Class $h$. When we use fuzzy rules in the prevision section for pattern classification as well as function approximation, the estimated output $g(\mathbf{x}_p)$ is used as $\alpha_h(\mathbf{x}_p)$.

Instead of the maximum compatibility in (14), it is also possible to define $\alpha_h(\mathbf{x}_p)$ by the total compatibility grade as

$$\alpha_h(\mathbf{x}_p) = \sum_{\substack{q=1 \\ C_q=h}}^{N} \mathbf{A}_q(\mathbf{x}_p),\ h = 1, 2, ..., M. \tag{15}$$

The input pattern $\mathbf{x}_p$ is classified as the class with the maximum value of $\alpha_h(\mathbf{x}_p)$. Whereas only the winner rule with the maximum compatibility grade is responsible for the classification result in (14), all compatible rules vote for their consequent classes in (15).

For illustration, let us assume that we have the following fuzzy rules:

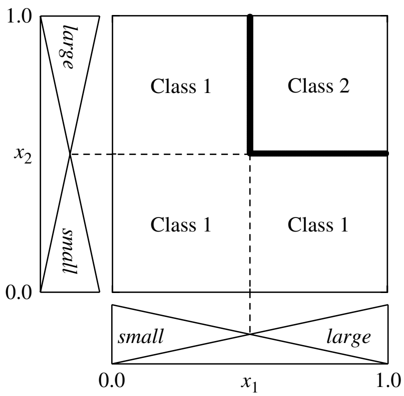
If $x_1$ is $small$ and $x_2$ is $small$ then Class 1,
If $x_1$ is $small$ and $x_2$ is $large$ then Class 1,
If $x_1$ is $large$ and $x_2$ is $small$ then Class 1,
If $x_1$ is $large$ and $x_2$ is $large$ then Class 2,

where $small$ and $large$ are linguistic terms of antecedent fuzzy sets. These fuzzy rules are shown in Fig. 1. The bold line in Fig. 1 (a) shows the classification boundary by the winner-take-all scheme in (14) using the four fuzzy rules.

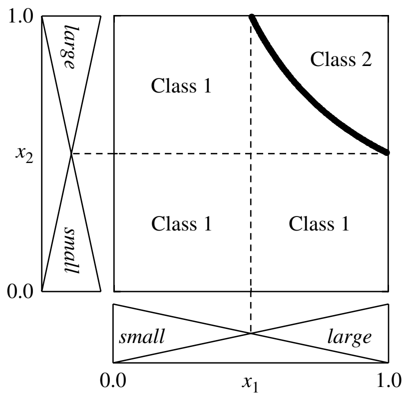
As pointed out in Kuncheva [49, 47], the classification boundary in Fig. 1 (a) by the winner-take-all scheme in (14) is exactly the same as that of the non-fuzzy lookup table with the following four rules:

If $x_1$ is in [0.0, 0.5] and $x_2$ is in [0.0, 0.5] then Class 1,
If $x_1$ is in [0.0, 0.5] and $x_2$ is in [0.5, 1.0] then Class 1,
If $x_1$ is in [0.5, 1.0] and $x_2$ is in [0.0, 0.5] then Class 1,
If $x_1$ is in [0.5, 1.0] and $x_2$ is in [0.5, 1.0] then Class 2.

On the other hand, Fig. 1 (b) by the voting-based scheme in (15) is different from such a non-fuzzy lookup table. As shown in Fig. 1, smoother classification boundaries are obtained by the voting-based scheme. This suggests its possible advantage in terms of the generalization ability of fuzzy rule-based classification systems. The winner-take-all scheme, however, may be preferred when human users want to know why each pattern is classified as a particular class. This is because only a single fuzzy rule is responsible for the classification of each pattern in the winner-take-all scheme. In this chapter, we use the winner-take-all scheme.



(a) Winner-take-all scheme in (14).



(b) Voting-based scheme in (15).

**Fig. 1** Classification results by the four fuzzy rules

When we use fuzzy rules with rule weights in (12), $\alpha_h(\mathbf{x}_p)$ is defined as

$$\alpha_h(\mathbf{x}_p) = \max\{\mathbf{A}_q(\mathbf{x}_p) \cdot w_q \mid C_q = h; q = 1, 2, ..., N\},\ h = 1, 2, ..., M. \quad (16)$$

That is, $\alpha_h(\mathbf{x}_p)$ is defined as the maximum value of the product of the compatibility grade and the rule weight. As demonstrated in [6, 37], various classification boundaries can be obtained by changing only the rule weight of each fuzzy rule with no modification of its antecedent fuzzy sets. In Fig. 2, we show two examples of classification boundaries generated by the above-mentioned four fuzzy rules when they have rule weights.

Classification boundaries can be also adjusted by linguistic hedges such as *very* and *more or less* (e.g., see Casillas et al. [19] for the use of linguistic hedges in fuzzy rule-based systems). We show two examples of classification boundaries in Fig. 3, which are obtained by changing only the fourth fuzzy rule with the Class 2 consequent in Fig. 1 (a). In Fig. 3, we use the winner-take-all scheme for the four fuzzy rules with no rule weights. In Fig. 3 (a), the fourth fuzzy rule is modified as

If $x_1$ is *very large* and $x_2$ is *very large* then Class 2.

On the other hand, we use the following fuzzy rule in Fig. 3 (b):

If $x_1$ is *more or less large* and $x_2$ is *more or less large* then Class 2.

The membership function of *large* is changed as follows by these two linguistic hedges:

$$\mu_{very\ large}(x) = [\mu_{large}(x)]^2, \quad (17)$$

$$\mu_{more\ or\ less\ large}(x) = \sqrt{\mu_{large}(x)}. \quad (18)$$

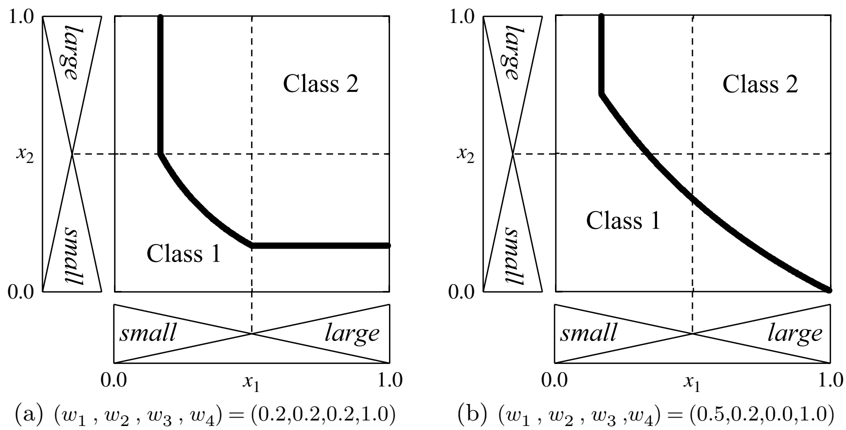
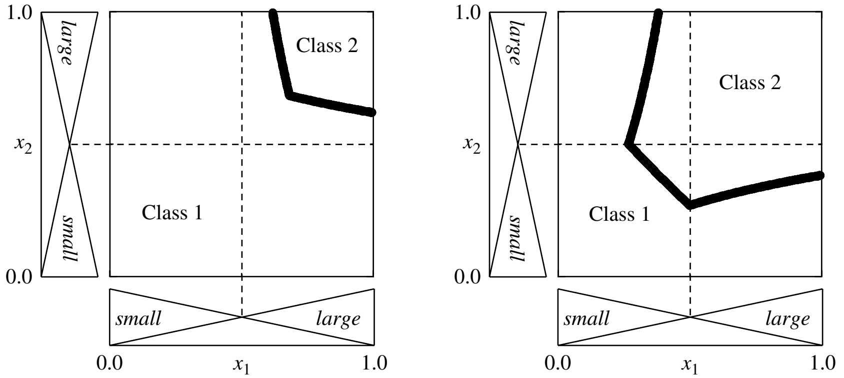



**Fig. 2** Classification results by the four fuzzy rules with rule weights




(a) Use of *very large* instead of *large*.

(b) Use of *more or less large*.

**Fig. 3** Classification results after modifying the fourth fuzzy rule with the Class 2 consequent using the linguistic hedges *very* and *more or less*

## 4 Linguistic Rule Extraction for Pattern Classification

Various approaches to linguistic rule extraction have been proposed in the literature (e.g., [34, 23, 12, 26]). In this section, we explain two approaches to linguistic rule extraction from numerical data for pattern classification. One is based on fuzzy grids and the other is based on fuzzy clustering. These approaches are easy to understand and have been frequently used in the literature for the design of fuzzy rule-based classification systems with linguistic rules. In this section, it is assumed that we have $m$ training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$, $p = 1, 2, ..., m$ from $M$ classes.

### *4.1 Fuzzy Grid-Based Approach*

One of the most popular fuzzy rule extraction methods is Wang & Mendel [62]. Whereas their method was originally proposed for function approximation, it is also applicable to pattern classification with minor modifications. Let us assume that each of the $n$ attributes is divided into $K_i$ antecedent fuzzy sets with linguistic terms $(i = 1, 2, ..., n)$. This means that the $n$-dimensional pattern space is divided into $K_1 \times K_2 \times ... \times K_n$ fuzzy subspaces. In the same manner as Wang & Mendel [62], a single fuzzy rule is generated from each pattern by finding the most compatible fuzzy subspace. The antecedent part of the fuzzy rule is specified by the identified fuzzy subspace while its consequent class is the same as the pattern. In this manner, $m$ fuzzy rules including duplicated ones are generated from the given $m$ training patterns. Among the generated $m$ fuzzy rules, some rules may be conflicting with each other. When multiple fuzzy rules have the same antecedent part and different consequent classes, the conflict is resolved by choosing the class of the pattern with the highest compatibility grade.

Whereas only a single pattern (i.e., the most compatible pattern with the antecedent part) is responsible for the determination of the consequent part of each fuzzy rule in Wang & Mendel [62], all the compatible patterns have an effect on the rule extraction in Ishibuchi et al. [35]. When the consequent class of the fuzzy rule $R_q$ is to be specified, the sum of the compatibility grades with its antecedent part is calculated for each class as

$$\beta_h(\mathbf{A}_q) = \sum_{\mathbf{x}_p \in \text{Class } h} \mathbf{A}_q(\mathbf{x}_p), \;\; h = 1, 2, ..., M, \tag{19}$$

where $\mathbf{A}_q(\mathbf{x}_p)$ is the compatibility grade of the pattern $\mathbf{x}_p$ with the antecedent part of the fuzzy rule $R_q$. The consequent class $C_q$ of the fuzzy rule $R_q$ is specified as the class with the maximum value of (19) over the $M$ classes (i.e., with the maximum total compatibility grade). On the other hand, Wang & Mendel [62] can be viewed as using the following maximum compatibility grade instead of the total compatibility grades in (19):

$$\beta_h(\mathbf{A}_q) = \max\{\mathbf{A}_q(\mathbf{x}_p) \mid \mathbf{x}_p \in \text{Class } h\}, \;\; h = 1, 2, ..., M. \tag{20}$$

It should be noted that all fuzzy subspaces are examined to generate fuzzy rules in Ishibuchi et al. [35] while only a single fuzzy rule is generated from each training pattern in Wang & Mendel [62]. Thus much more fuzzy rules are usually generated by Ishibuchi et al. [35].

Let us explain these two methods using Fig. 4. Wang & Mendel [62] generates conflicting rules with the same antecedent part "If $x_1$ is *medium* and $x_2$ is *medium*" in Fig. 4 (a) and Fig. 4 (b). The conflict is resolved by choosing Class 2 as the consequent class in both plots because a Class 2 pattern (i.e., Pattern B) has the maximum compatibility grade with the antecedent part. On the other hand, the consequent

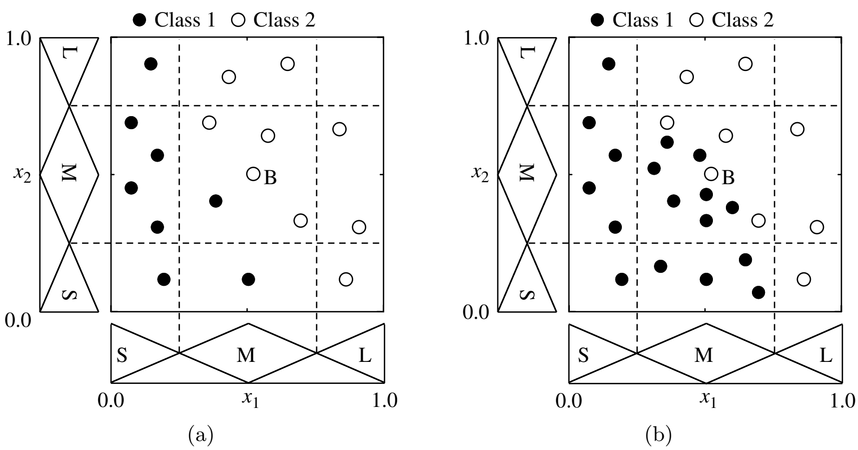



**Fig. 4** Two training data sets in the two-dimensional pattern space. Three antecedent fuzzy sets *small*, *medium* and *large* are denoted by S, M and L, respectively

class by Ishibuchi et al. [35] is Class 2 in Fig. 4 (a) and Class 1 in Fig. 4 (b). This is because Class 1 has the larger sum of the compatibility grades than Class 2 in Fig. 4 (b).

In Fig. 4 (a) and Fig. 4 (b), Wang & Mendel [62] generates no fuzzy rule with the antecedent part "If $x_1$ is $large$ and $x_2$ is $large$" around the upper-right corner. This is because the corresponding fuzzy subspace is not chosen as the most compatible fuzzy subspace for any training patterns. On the other hand, Ishibuchi et al. [35] generates "If $x_1$ is $large$ and $x_2$ is $large$ then Class 2" for the same fuzzy subspace in both plots.

## *4.2 Fuzzy Grid-Based Genetic Fuzzy Systems*

Genetic algorithms have been frequently used as machine learning techniques for generating fuzzy rules from numerical data. Genetic algorithms for fuzzy rule generation are called fuzzy genetics-based machine learning (GBML) algorithms. Fuzzy GBML algorithms are often categorized into three groups [40, 12]: Iterative rule learning approach, Michigan approach, and Pittsburgh approach. These three approaches are classified into two categories according to the coding strategies. One is the "chromosome = set of rules" approach which is usually called the Pittsburgh approach. The other is the "chromosome = rule" approach, which includes the iterative rule learning approach and the Michigan approach. Sometimes the Michigan approach is used in a broad sense to mean the "chromosome = rule" approach in general. Herrera [17] subdivides the "chromosome = rule" approach into three groups: the Michigan approach, the iterative rule learning approach, and the cooperative-competitive approach (for details, see Herrera [17]).

In the iterative rule learning approach, an individual is a single fuzzy rule. Thus each population is a set of fuzzy rules (i.e., each generation consists of a set of fuzzy rules). A heuristic rule evaluation criterion is used as a fitness function to evaluate each fuzzy rule. A genetic algorithm is used to search for the best fuzzy rule with respect to the given criterion. Only a single fuzzy rule is obtained from its single run. A fuzzy rule-based classification system is constructed by iterating the execution of the genetic algorithm. The main characteristic of the iterative rule learning approach is that the training data are adjusted after each run (i.e., when each rule is obtained). Roughly speaking, training patterns covered by the obtained fuzzy rule are removed from the training data after each run. MOGUL [32, 18] and SLAVE [31, 20] are examples of fuzzy GBML algorithms in the category of the iterative rule learning approach. The main advantage of this approach is its efficiency. It does not need long computation time or large memory storage. Its main disadvantage is the lack of the global optimization ability in the level of rule sets (i.e., in the level of fuzzy rule-based systems) because each rule is sequentially obtained without evaluating the performance of the entire fuzzy rule-based system.

As in the iterated rule learning approach, an individual is a single fuzzy rule in the Michigan approach. A heuristic rule evaluation criterion is used as a fitness function to evaluate each fuzzy rule. Whereas a genetic algorithm is used to search for a single fuzzy rule in the iterative rule learning approach, the Michigan approach searches for a set of fuzzy rules (i.e., a fuzzy rule-based system) by its single run. Fuzzy GBML algorithms in the Michigan approach are often called fuzzy classifier systems. It should be noted that fuzzy classifier systems are totally different from fuzzy rule-based classification systems. The former is a framework of fuzzy GBML algorithms whereas the latter is fuzzy rule-based systems for pattern classification. The main characteristic of the Michigan approach is that each population corresponds to a fuzzy rule-based system. Thus the generation update in the Michigan approach can be viewed as the learning of a fuzzy rule-based system. The Michigan approach was used to design linguistically interpretable fuzzy rule-based classification systems in Ishibuchi et al. [8, 5]. The main advantage of the Michigan approach is its efficiency. Its main disadvantage is the lack of the global optimization ability in the level of rule sets. In the Michigan approach (as well as the iterative rule learning approach), the search for a good fuzzy rule-based system is driven by the evaluation of each fuzzy rule.

In the Pittsburgh approach, an individual is the set of fuzzy rules. That is, each individual corresponds to a fuzzy rule-based system. The accuracy of each fuzzy rule-based system is used as (a part of) a fitness function. A genetic algorithm is used to optimize fuzzy rule-based systems. Many genetic fuzzy systems have been developed in the framework of the Pittsburgh approach [40, 12]. The main advantage of the Pittsburgh approach is its global optimization ability in the level of rule sets. This is because the evolution of fuzzy rule-based systems is directly driven by their accuracy. It is also easy in the Pittsburgh approach to incorporate various interpretability measures into the fitness function to find a good compromise between the accuracy and the interpretability. The main disadvantage of the Pittsburgh approach is its inefficiency. It usually needs long computation time and large memory storage. If the available computation time is very small for a large-scale problem, the Pittsburgh approach is not likely to find good rule sets. A hybrid algorithm of the Michigan and Pittsburgh approaches was proposed in Ishibuchi et al. [11] to utilize the advantage of each approach in a single fuzzy GBML algorithm.

The above-mentioned fuzzy GBML algorithms usually assume that each attribute has already been divided into several linguistic terms. When each attribute is divided into three linguistic terms $small$, $medium$ and $large$ as in Fig. 4, the antecedent part of a fuzzy rule for an $n$-dimensional classification problem is represented by an integer string of length $n$ using an alphabet with four integers (i.e., 0: $don't\ care$, 1: $small$, 2: $medium$, and 3: $large$). For example, "21" means the antecedent condition "If $x_1$ is $medium$ and $x_2$ is $small$". Since the $don't\ care$ condition can be removed from the antecedent part, "03" means "If $x_2$ is $large$". This coding method was used in Ishibuchi et al. [8, 5, 11]. Since the consequent part of each fuzzy rule can be specified from compatible training patterns by a heuristic procedure, only its antecedent part is usually coded as a string in fuzzy GBML algorithms for pattern classification.

In SLAVE [20], a binary string is used to represent the antecedent part of each fuzzy rule instead of an integer string. For example, "011" means "$medium$ or $large$" when the corresponding attribute value is divided into three linguistic terms $small$, $medium$ and $large$ as in Fig. 4. In this case, the antecedent part of a fuzzy rule for an $n$-dimensional problem is represented by a binary string of length $3n$. In Fig. 4, "010101" means the antecedent part "If $x_1$ is *medium* and $x_2$ is *small* or *large*". As a special case, "111" is handled as *don't care*.

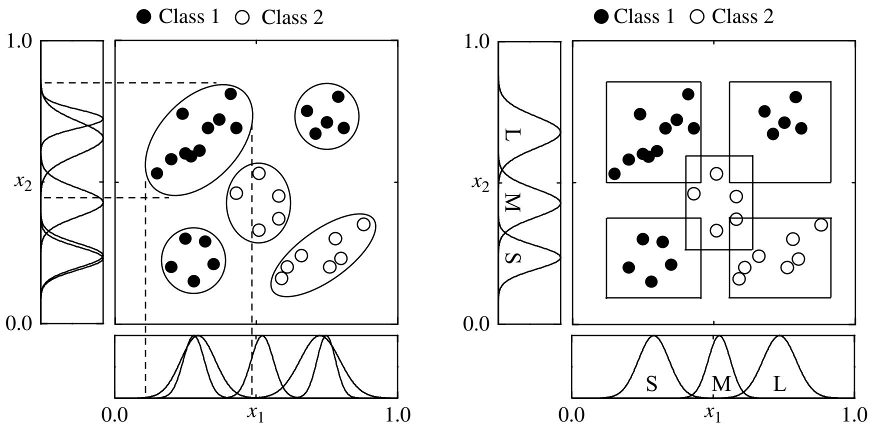
In a modified version of SLAVE by Castillo et al. [31], input selection ability is incorporated into SLAVE to explicitly decrease the complexity of fuzzy rule-based systems (i.e., to improve their interpretability). A real number string "$z_1 z_2 ... z_n \theta$" of length $n + 1$ is used to represent input selection in each fuzzy rule for an $n$-dimensional problem. When $z_i < \theta$, the antecedent condition on the $i$th attribute is not used (i.e., *don't care* is assigned to the $i$th attribute). It should be noted that a different real number string is attached to each fuzzy rule. That is, each fuzzy rule is denoted by a concatenated string of a real number string (for input selection) and a binary string (for antecedent conditions) in the modified SLAVE algorithm in Castillo et al. [31]. Of course, a binary string of length $n$ can be also used to directly represent input selection.

## 4.3 Clustering-Based Approach

Since Sugeno & Yasukawa [58], various clustering-based methods have been proposed to the extraction of linguistically interpretable fuzzy rules from numerical data for modeling and classification (e.g., Nauck & Kruse [51, 52], Roubos & Setnes [55], Setnes et al. [56], and Setnes & Roubos [57]). A general outline of these clustering-based methods can be written for pattern classification as follows:

Step 1: Clustering in the pattern space.
Step 2: Specification of a multi-dimensional fuzzy set for each cluster.
Step 3: Projection of each multi-dimensional fuzzy set onto each attribute to generate a single-dimensional antecedent fuzzy set.
Step 4: Adjustment of each single-dimensional antecedent fuzzy set to improve the accuracy and the interpretability of a fuzzy rule-based classification system.
Step 5: Assignment of a linguistic term to each antecedent fuzzy set.

In Step 1, a prespecified number of clusters are obtained. Fuzzy clustering techniques are used in Step 1. The number of fuzzy rules is the same as the number of clusters. Next a multi-dimensional fuzzy set is specified in the pattern space for each cluster in Step 2. Then each multi-dimensional fuzzy set is projected onto each attribute to generate a single-dimensional antecedent fuzzy set in Step 3. The first three steps are illustrated in Fig. 5 (a). As shown in Fig. 5 (a), the obtained single-dimensional antecedent fuzzy sets do not always have high interpretability due to large overlaps among them. Another difficulty is information loss by the projection (see Fig. 5 (a)). This means that the obtained single-dimensional antecedent fuzzy sets do not always have high classification ability, either. Thus each antecedent fuzzy set is adjusted to improve both the accuracy and the interpretability in Step 4. In this step, the number of the antecedent fuzzy sets is decreased through simplification



(a) The first three steps. (b) The last two steps.

**Fig. 5** Illustration of clustering-based approaches to linguistic rule extraction

procedures. The prespecified number of fuzzy rules in Step 1 can be also decreased in Step 4. Finally an appropriate linguistic term is assigned to each antecedent fuzzy set in Step 5 to generate linguistic rules. The last two steps are illustrated in Fig. 5 (b).

Linguistic rules with higher classification ability are often obtained from clustering-based methods than grid-based ones in the previous subsection. This is because the membership function of each antecedent fuzzy set is adjusted in clustering-based methods while the fixed membership function is usually used with no modifications in fuzzy grid-based methods.

As pointed out in Nauck & Kruse [52], information loss is involved in the projection of multi-dimensional fuzzy sets onto each attribute. This means that the accuracy of fuzzy rule-based systems is degraded by extracting linguistic rules from fuzzy rules with multi-dimensional antecedent fuzzy sets. If our goal is the accuracy maximization, it is recommended to use multi-dimensional antecedent fuzzy sets without projections (e.g., Abe & Thawonmas [50]). On the other hand, if our goal is the interpretability maximization, the use of single-dimensional antecedent fuzzy sets is recommended. Single-dimensional antecedent fuzzy sets obtained by clustering-based methods, however, are not necessarily linguistically interpretable. This is because the consistency with human users' intuition is not included in the adjustment of each antecedent fuzzy set in Step 4.

## 5 Future Research Topics

There are a number of future research topics to be discussed for the design of fuzzy rule-based systems with high accuracy and high linguistic interpretability. Here we briefly explain some future research topics.

### *5.1 Interpretability Measures*

There are many aspects related to the interpretability of fuzzy rule-based systems [34, 23, 45, 39] such as

- Number of fuzzy rules,
- Number of attributes (i.e., inputs) in fuzzy rule-based system,
- Number of antecedent conditions in each fuzzy rule,
- Number of antecedent fuzzy sets for each attribute,
- Separability of adjacent antecedent fuzzy sets,
- Linguistic interpretability of each antecedent fuzzy set.

It is not clear which measures should be taken into account in the design of fuzzy rule-based systems. It is even more unclear how these measures are combined with an accuracy measure. Moreover it is very difficult to quantitatively measure the last aspect: the linguistic interpretability of each antecedent fuzzy set.

### *5.2 Accuracy Improvement of Linguistic Rules*

Several approaches have been proposed to improve the accuracy of linguistic rule-based systems for function approximation. For example, rule selection was used in [33, 19, 16], linguistic hedges such as $very$ and $more\ or\ less$ were attached to antecedent fuzzy sets in [19], antecedent fuzzy sets were subdivided in a hierarchical manner in [33, 13], double consequents with different weights were used in each fuzzy rule in [16, 54], and the combinations of antecedent and consequent fuzzy sets were globally optimized in [36].

For pattern classification, only a few approaches to the accuracy improvement of linguistic rule-based systems have been examined. Among them are rule selection [7, 9], rule weight adjustment [6, 37, 53] and membership function learning [51, 52, 55]. Further studies are needed to design fuzzy rule-based classification systems with high accuracy and high linguistic interpretability.

### *5.3 Multiobjective Optimization of Fuzzy Rule-Based Systems*

As we have already explained, there exists a tradeoff between the accuracy and the interpretability. Thus it is impossible to simultaneously optimize these two design criteria. In many studies on genetic fuzzy systems, these two design criteria were combined into a scalarizing fitness function to find a good compromise between them using a standard single-objective genetic algorithm (e.g., [7, 9, 55, 56, 57]). The main difficulty in these studies is that the specification of an appropriate scalarizing fitness function is not easy. For example, it is usually very difficult to assign relative importance to the average error rate and the number of fuzzy rules in the design of fuzzy rule-based classification systems. In order to circumvent this difficulty, the concept of Pareto-optimality has been introduced for the multiobjective design of fuzzy rule-based systems in the literature. For example, Ishibuchi et al. [28] formulated a two-objective fuzzy rule selection problem to search for Pareto-optimal fuzzy rule-based systems with respect to the classification accuracy and the number of fuzzy rules in the late 1990s. Their idea was extended to the case of three-objective fuzzy rule selection in [44, 3] where the number of antecedent conditions was taken into account as an additional complexity measure. In these studies, a number of Pareto-optimal (or near Pareto-optimal) fuzzy rule-based classification systems were obtained using a multiobjective genetic algorithm. The concept of Pareto-optimality was also incorporated into multiobjective fuzzy GBML algorithms (e.g., [44, 4, 60, 61]). Multiobjective design of classification systems is a very hot issue not only in the fuzzy system community but also in the machine learning community. For example, see Jin [25].

### *5.4 Application to Large-Scale Data Sets*

Fuzzy data mining (i.e., fuzzy rule extraction from large-scale data sets) seems to be a promising research area in the field of fuzzy systems (e.g., [1, 10, 29, 26, 3]). In data mining, emphasis is usually placed on the interpretability of extracted knowledge for human users rather than its accuracy. Thus fuzzy logic has a large potential ability to play an important role in data mining. Despite of its potential ability, fuzzy logic is not widely used in the field of data mining.

## 6 Conclusion

In this chapter, we explained fuzzy rule-based classification systems and linguistic rule extraction from numerical data for pattern classification. When our main goal is to design classification systems with high interpretability, the use of linguistic rules is a good choice. We can construct fuzzy rule-based classification systems with high linguistic interpretability. The use of linguistic rule is also a promising choice in data mining. We can extract linguistically interpretable knowledge from numerical data. On the other hand, the use of fuzzy rules with multi-dimensional antecedent fuzzy sets seems to be a good choice when our main goal is to construct classification systems with high classification accuracy.

Due to the page limitation, our explanations in this chapter did not go into details. For details of linguistic rule extraction, see the book on linguistic data mining by Ishibuchi et al. [26]. For further discussions on the accuracy-interpretability tradeoff in fuzzy rule-based systems, see the edited books by Casillas et al. [34, 23]. Also see the edited book by Jin [25] for recent developments on multiobjective machine learning.

**Acknowledgment** This work was partially supported by Grant-in-Aid for Scientific Research on Priority Areas: KAKENHI (18049065)

## References

1. Abe S, Lan MS (1995) A method for fuzzy rule extraction directly from numerical data and its application to pattern classification. IEEE Trans. on Fuzzy Systems 3: 18–28.
2. Abe S, Thawonmas R (1997) A fuzzy classifier with ellipsoidal regions. IEEE Trans. on Fuzzy Systems 5: 358–368.
3. Abonyi J, Roubos JA, Szeifert F (2003) Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization. International Journal of Approximate Reasoning 32: 1–21.
4. Alcala R, Cano JR, Cordon O, Herrera F, Villar P, Zwir I (2003) Linguistic modeling with hierarchical systems of weighted linguistic rules. International Journal of Approximate Reasoning 32: 187–215.
5. Casillas J, Cordon O, Herrera F (2002) COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules. IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics 32: 526–537.
6. Casillas J, Cordon O, Herrera F, Magdalena L (eds) (2003) Interpretability issues in fuzzy modeling. Springer, Berlin Heidelberg New York.
7. Casillas J, Cordon O, Herrera F, Magdalena L (eds) (2003) Accuracy improvements in linguistic fuzzy modeling. Springer, Berlin Heidelberg New York.
8. Casillas J, Cordon O, del Jesus MJ, Herrera F (2005) Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. IEEE Trans. on Fuzzy Systems 13: 13–29.
9. Castillo L, Gonzalez A, Perez R (2001) Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm. Fuzzy Sets and Systems 120: 309–321.
10. Chakraborty D, Pal NR (2004) A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification. IEEE Trans. on Neural Networks 15: 110–123.
11. Cordon O, Gomide F, Herrera F, Hoffmann F, Magdalena L (2004) Ten years of genetic fuzzy systems: Current framework and new trends. Fuzzy Sets and Systems 141: 5–31.
12. Cordon O, Herrera F (2001) A proposal for improving the accuracy of linguistic modeling. IEEE Trans. on Fuzzy Systems 8: 335–344.
13. Cordon O, Herrera F, Hoffmann F, Magdalena L (2001) Genetic fuzzy systems: Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific, Singapore New Jersey London Hong Kong.
14. Cordon O, Herrera F, Zwir I (2002) Linguistic modeling by hierarchical systems of linguistic rules. IEEE Trans. on Fuzzy Systems 10: 2–20.
15. Cordon O, del Jesus MJ, Herrera F (1999) A proposal on reasoning methods in fuzzy rule-based classification systems. International Journal of Approximate Reasoning 20: 21–45.
16. Cordon O, del Jesus MJ, Herrera F, Lozano M (1999) MOGUL: A methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach. International Journal of Intelligent Systems 14: 1123–1153.
17. Gonzalez A, Perez R (1999) SLAVE: A genetic learning system based on an iterative approach. IEEE Trans. on Fuzzy Systems 7: 176–191.
18. Guillaume S (2001) Designing fuzzy inference systems from data: An interpretability-oriented review. IEEE Trans. on Fuzzy Systems 9: 426–443.
19. Herrera F (2005) Genetic fuzzy systems: Status, critical considerations and future directions. International Journal of Computational Intelligence Research 1: 59–67.
20. Herrera F, Lozano M, Verdegay JL (1998) A learning process for fuzzy control rules using genetic algorithms. Fuzzy Sets and Systems 100: 143–158.

21. Homaifar A, McCormick E (1995) Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. IEEE Trans. on Fuzzy Systems 3: 129–139.
22. Hong TP, Kuo CS, Chi SC (2001) Trade-off between computation time and number of rules for fuzzy mining from quantitative data. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9: 587–604.
23. Horikawa S, Furuhashi T, Uchikawa Y (1992) On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm. IEEE Trans. on Neural Networks 3: 801–806.
24. Hu YC, Chen RS, Tzeng GH (2002) Mining fuzzy association rules for classification problems. Computers & Industrial Engineering 43: 735–750.
25. Hu YC, Chen RS, Tzeng GH (2003) Finding fuzzy classification rules using data mining techniques. Pattern Recognition Letters 24: 509–519.
26. Ishibuchi H, Murata T, Turksen IB (1997) Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. Fuzzy Sets and Systems 89: 135–150.
27. Ishibuchi H, Nakashima T (1999) Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes. IEEE Trans. on Industrial Electronics 46: 157–168.
28. Ishibuchi H, Nakashima T (2001) Effect of rule weights in fuzzy rule-based classification systems. IEEE Trans. on Fuzzy Systems 9: 506–515.
29. Ishibuchi H, Nakashima T, Morisawa T (1999) Voting in fuzzy rule-based systems for pattern classification problems. Fuzzy Sets and Systems 103: 223–238.
30. Ishibuchi H, Nakashima T, Murata T (1999) Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems. IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics 29: 601–618.
31. Ishibuchi H, Nakashima T, Murata T (2001) Three-objective genetics-based machine learning for linguistic rule extraction. Information Sciences 136: 109–133.
32. Ishibuchi H, Nakashima T, Nii M (2004) Classification and modeling with linguistic information granules: Advanced approaches to linguistic data mining. Springer, Berlin Heidelberg New York.
33. Ishibuchi H, Nojima Y (2006) Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. International Journal of Approximate Reasoning (in press).
34. Ishibuchi H, Nozaki K, Tanaka H (1992) Distributed representation of fuzzy rules and its application to pattern classification. Fuzzy Sets and Systems 52: 21–32.
35. Ishibuchi H, Nozaki K, Yamamoto N, Tanaka H (1994) Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. Fuzzy Sets and Systems 65: 237–253.
36. Ishibuchi H, Nozaki K, Yamamoto N, Tanaka H (1995) Selecting fuzzy if-then rules for classification problems using genetic algorithms. IEEE Trans. on Fuzzy Systems 3: 260–270.
37. Ishibuchi H, Yamamoto T (2004) Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets and Systems 141: 59–88.
38. Ishibuchi H, Yamamoto T (2005) Hybridization of fuzzy GBML approaches for pattern classification problems. IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics 35: 359–365.
39. Ishibuchi H, Yamamoto T (2005) Rule weight specification in fuzzy rule-based classification Systems. IEEE Trans. on Fuzzy Systems 13: 428–435.
40. Jang JSR (1993) ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. on Systems, Man, and Cybernetics 23: 665–685.
41. Jin Y (2000) Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement. IEEE Trans. on Fuzzy Systems 8: 212–221.
42. Jin Y (ed) (2006) Multi-objective machine learning. Springer, Berlin Heidelberg New York.
43. Karr CL, Gentry EJ (1993) Fuzzy control of pH using genetic algorithms. IEEE Trans. on Fuzzy Systems 1: 46–53.

44. Kuncheva LI (2000) Fuzzy classifier design. Springer, Berlin Heidelberg New York.
45. Kuncheva LI (2000) How good are fuzzy if-then classifiers? IEEE Trans. on Systems, Man, and Cybernetics—Part B: Cybernetics 30: 501–509.
46. Lin CT, Lee CSG (1991) Neural-network-based fuzzy logic control and decision system. IEEE Trans. on Computers 40: 1320–1336.
47. Lin CT, Yeh CM, Liang SF, Chung JF, Kumar N (2006) Support-vector-based fuzzy neural network for pattern classification. IEEE Trans. on Fuzzy Systems 14: 31–41.
48. Mamdani EH (1974) Applications of fuzzy algorithms for simple dynamic plant. Proceedings of IEE 121: 1585–1588.
49. Mikut R, Jakel J, Groll L (2005) Interpretability issues in data-based learning of fuzzy systems. Fuzzy Sets and Systems 150: 179–197.
50. Mitra S, Konwar KM, Pal SK (2002) Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: Generation and evaluation. IEEE Trans. on Systems, Man, and Cybernetics—Part C: Applications and Reviews 32: 328–339.
51. Nauck D, Kruse R (1997) A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy Sets and Systems 89: 277–288.
52. Nauck D, Kruse R (1999) Obtaining interpretable fuzzy classification rules from medical data. Artificial Intelligence in Medicine 16: 149–169.
53. Nozaki K, Ishibuchi H, Tanaka H (1996) Adaptive fuzzy rule-based classification systems. IEEE Trans. on Fuzzy Systems 4: 238–250.
54. Nozaki K, Ishibuchi H, Tanaka H (1997) A simple but powerful heuristic method for generating fuzzy rules from numerical data. Fuzzy Sets and Systems 86: 251–270.
55. Roubos H, Setnes M (2001) Compact and transparent fuzzy models and classifiers through iterative complexity reduction. IEEE Trans. on Fuzzy Systems 9: 516–524.
56. Setnes M, Babuska R, Verbruggen HB (1998) Rule-based modeling: Precision and transparency. IEEE Trans. on Systems, Man, and Cybernetics—Part C 28: 165–169.
57. Setnes M, Roubos H (2000) GA-fuzzy modeling and classification: Complexity and performance. IEEE Trans. on Fuzzy Systems 8: 509–522.
58. Sugeno M, Yasukawa T (1993) A fuzzy-logic-based approach to qualitative modeling. IEEE Trans. on Fuzzy Systems 1: 7–31.
59. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. on Systems, Man, and Cybernetics 15: 116–132.
60. Wang H, Kwong S, Jin Y, Wei W, Man KF (2005) Agent-based evolutionary approach for interpretable rule-based knowledge extraction. IEEE Trans. on Systems, Man, and Cybernetics—Part C: Applications and Reviews 35: 143–155.
61. Wang H, Kwong S, Jin Y, Wei W, Man KF (2005) Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction. Fuzzy Sets and Systems 149: 149–186.
62. Wang LX, Mendel JM (1992) Generating fuzzy rules by learning from examples. IEEE Trans. on Systems, Man, and Cybernetics 22: 1414–1427.
63. Yuan Y, Shaw MJ (1995) Induction of fuzzy decision trees. Fuzzy Sets and Systems 69: 125–139.

# An Overview of Mining Fuzzy Association Rules

**Tzung-Pei Hong and Yeong-Chyi Lee**

**Abstract** Data mining is the process of extracting desirable knowledge or interesting patterns from existing databases for specific purposes. Many types of knowledge and technology have been proposed for data mining. Among them, finding association rules from transaction data is most commonly seen. Most studies have shown how binary valued transaction data may be handled. Transaction data in real-world applications, however, usually consist of fuzzy and quantitative values, so designing sophisticated data-mining algorithms able to deal with various types of data presents a challenge to workers in this research field. This chapter thus surveys some fuzzy mining concepts and techniques related to association-rule discovery. The motivation from crisp mining to fuzzy mining will be first described. Some crisp mining techniques for handling quantitative data will then be briefly reviewed. Several fuzzy mining techniques, including mining fuzzy association rules, mining fuzzy generalized association rules, mining both membership functions and fuzzy association rules, will then be described. The advantages and the limitations of fuzzy mining will also be discussed.

## 1 Introduction

Knowledge discovery in databases (KDD) has thus become a process of considerable interest in recent years, as the amounts of data in many databases have grown tremendously large. KDD means the application of nontrivial procedures for identifying effective, coherent, potentially useful, and previously unknown patterns in large databases [20, 45]. The KDD process generally consists of three phases: pre-processing, data mining and post-processing [17, 2]. Among them, data mining plays a critical role to KDD.

Depending on the types of databases to be processed, mining approaches may be classified as working on transaction databases, temporal databases, relational databases, and multimedia databases, among others [31]. On the other hand, depending on the classes of knowledge derived, mining approaches may be classified as finding association rules, classification rules, clustering rules, and sequential patterns [14], among others. Among them, finding association rules in transaction databases is most commonly seen in data mining [41, 50, 18, 1, 42, 25, 48, 51].

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008

An association rule is an expression $X \rightarrow Y$, where $X$ is a set of items and $Y$ is usually a single item. It means in the set of transactions, if all the items in $X$ exist in a transaction, then $Y$ is also in the transaction with a high probability. Most studies have shown how binary valued transaction data may be handled. Transaction data in real-world applications, however, usually consist of fuzzy and quantitative values. Designing sophisticated data-mining algorithms able to deal with various types of data presents a challenge to workers in this research field.

Fuzzy set theory was first proposed by Zadeh in 1965 [58]. It has been used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning [35]. It is primarily concerned with quantifying and reasoning using natural language in which words can have ambiguous meanings. This can be thought of as an extension of traditional crisp sets, in which each element must either be in or not in a set. The theory has been applied in many fields such as manufacturing, engineering, diagnosis, economics, among others [35, 37, 59]. Several fuzzy learning algorithms for inducing rules from given sets of data have been designed and used with good effects in specific domains [33, 34, 19, 24, 46, 29, 53]. Strategies based on decision trees [40] were proposed in [16, 47, 21]. Wang et al. also proposed a fuzzy version space learning strategy for managing vague information [53].

Recently, the fuzzy set theory has also been applied to data mining to find interesting association rules or sequential patterns in transaction data with quantitative values [23, 29, 28, 8, 6, 3, 57]. This chapter thus attempts to survey some fuzzy mining concepts and techniques about association rules. The remaining part of this chapter is organized as follows. Some related crisp mining approaches for association rules are first reviewed in Sect. 2. Some fuzzy mining techniques, including mining fuzzy association rules, mining fuzzy generalized association rules, mining both membership functions and fuzzy association rules, integrating clustering with fuzzy association rules, are then described in Sects. 3 to 5. The advantages and the limitations of fuzzy mining are discussed and conclusions are given in Sect. 6.

## 2 Some Crisp Data-mining Approaches for Association Rules

The goal of data mining is to discover important associations among items such that the presence of some items in a transaction will imply the presence of some other items. To achieve this purpose, Agrawal and his co-workers proposed the famous *Apriori* algorithm that based on the concept of large itemsets to find association rules in transaction data [41, 50]. They divided the mining process into two phases. In the first phase, candidate itemsets were generated and counted by scanning the transaction data. If the number of an itemset appearing in the transactions (called the support of an itemset) was larger than a pre-defined threshold value (called minimum support), the itemset was considered a large itemset. Itemsets containing only one item were processed first. Large itemsets containing only single items were then combined to form candidate itemsets containing two items. This process was repeated until all large itemsets had been found. In the second phase, association rules were induced from the large itemsets found in the first phase. All possible association rules from each large itemset were formed, and their confidence values were calculated. The confidence was defined as the support of the itemset in the left hand side of the rule over the support of the itemset in the whole rule. Those rules with calculated confidence values larger than a predefined threshold (called minimum confidence) were output as association rules.

Han et al. proposed the Frequent-Pattern-tree structure (FP-tree) for efficiently mining association rules without generation of candidate itemsets [22]. The approach compressed a database into a tree structure storing only large items. Three steps were involved in FP-tree construction. The database was first scanned to find all items with their frequency. The items with their supports larger than a predefined minimum support were selected as large 1-itemsets (items). Next, the large items were sorted in descending frequency. At last, the database was scanned again to construct the FP-tree according to the sorted order of large items. The construction process was executed tuple by tuple, from the first transaction to the last one. After all transactions were processed, the FP-tree was completely constructed. After the FP-tree was constructed from a database, a mining procedure called FP-Growth [22] was executed to find all large itemsets. FP-Growth did not need to generate candidate itemsets for mining, but derived frequent patterns directly from the FP-tree. It was a recursive process, handling the frequent items one by one. A conditional FP-tree was generated for each frequent item, and from the tree the large itemsets with the processed item could be recursively derived. Many mining methods for finding association rules based on the FP-tree structure have also been proposed [32, 4, 49, 60].

Cai et al. proposed weighted mining to reflect different importance on items [36]. Each item was attached a numerical weight given by users. Weighted supports and weighted confidences were then defined to determine interesting association rules.

In the above approach, the items in transactions are binary. That is, the value of an item is treated as "1" if the item is present in a transaction, and as "0" if it is not present. Items in real applications may, however, be presented with rich data types, such as quantitative and categorical types. Some approaches were then proposed for handling items with quantitative and categorical values while discovering association rules. Piatetsky-Shapiro proposed a mining approach, which partitioned quantitative attributes into intervals [27]. The value of each quantitative attribute could be presented as a range rather than a single value, and was allowed to appear in the left-hand side (antecedent) of a rule.

Srikant and Agrawal then proposed a method for mining association rules from transactions with quantitative attributes [51]. Their proposed method first determined the number of partitions for each quantitative attribute, and then mapped all possible values of each attribute into a set of consecutive integers. The number of intervals was determined by a partial completeness factor, which was used to evaluate the lost information from partition. Adjacent intervals were also allowed to be merged into a large one. Their approach then found large itemsets whose support values were greater than the user-specified minimum-support levels. These large itemsets were then processed to generate association rules. An interest measure was also introduced to get interesting association rules.

Fukuda et al. proposed a mining approach to discover optimized association rules from quantitative data [18]. Two problems were discussed, one for mining optimized support rules and the other for mining optimized conference rules. Mining optimized support rules focused on finding an association rule with a maximum support value on the condition that the confidence value of the rule had to be larger than a given minimum confidence. On the other hand, mining optimized confidence rules emphasized on finding an association rule with a maximum confidence value on the condition that the support value of the rule had to be larger than a given minimum support. The authors first found the ranges of quantitative attributes by sampling, sorting and filling data into buckets before the mining process proceeded. Association rules containing quantitative attributes in the left-hand side were then found by an effective search approach. In addition, Rastogi and Shim extended Fukuda et al.'s approach to find the association rules with disjunction conditions [30]. Lent et al. proposed a geometry-based algorithm to deal with the problem of clustering association rules from quantitative attributes [11]. Each association rule was initially represented by a primitive attribute value (or range). Adjacent association rules were then clustered together to form generalized rules. Many other researches are still in progress.

## 3 Mining Fuzzy Association Rules

As mentioned above, the fuzzy set theory is concerned with quantifying and reasoning using natural language. It is thus very suitable to handle quantitative values by fuzzy sets. Several fuzzy mining approaches have thus been proposed to find interesting association rules or sequential patterns in transaction data with quantitative values. In this section, we focus on the data mining approaches for finding association rules with predefined membership functions.

Chan and Au proposed an F-APACS algorithm to mine fuzzy association rules [23]. They first transformed quantitative attribute values into linguistic terms and then used the adjusted difference analysis to find interesting associations among attributes. It had the advantage that the user-specified thresholds were not needed since the statistical analysis was used. In addition, both positive and negative associations could be found.

Kuok et al. proposed a fuzzy mining approach to handle numerical data in databases with attributes and derived fuzzy association rules [3]. At nearly the same time, Hong et al. proposed a fuzzy mining algorithm to mine fuzzy rules from quantitative transaction data [29]. Basically, the fuzzy mining algorithms first used membership functions to transform each quantitative value into a fuzzy set in linguistic terms. The algorithm then calculated the scalar cardinality of each linguistic term on all the transaction data. The mining process based on fuzzy counts was then performed to find fuzzy association rules. Hong et al. described the fuzzy mining steps in details. Their fuzzy mining algorithm is described below as a good reference [8].

## *The Fuzzy Data Mining Algorithm:*

INPUT: A set of $n$ training data, each with $m$ attribute values, a set of membership functions, a predefined minimum support value $\alpha$, and a predefined confidence value $\lambda$.

OUTPUT: A set of fuzzy association rules.

STEP 1: Transform the quantitative value $v_j^{(i)}$ of each transaction datum $D^{(i)}$ for each attribute $A_j$, $j$=1 to $m$, into a fuzzy set $f_j^{(i)}$ represented as $\left(\frac{f_{j1}^{(i)}}{R_{j1}} + \frac{f_{j2}^{(i)}}{R_{j2}} + \cdots + \frac{f_{jl}^{(i)}}{R_{jl}}\right)$ using the given membership functions, where $i$ presents the current transaction number that is processed, $R_{jk}$ is the $k$-th fuzzy region (linguistic term) of attribute $A_j$, $f_{jk}^{(i)}$ is the value of the membership function in $R_{jk}$ for the value $v_j^{(i)}$, and $l$ (=$|A_j|$) is the number of fuzzy regions for $A_j$.

STEP 2: Calculate the count of each attribute region (linguistic term) $R_{jk}$ in the transaction data:

$$count_{jk} = \sum_{i=1}^{n} f_{jk}^{(i)}.$$

STEP 3: Collect each attribute region (linguistic term) to form the candidate set $C_1$.

STEP 4: Check whether $count_{jk}$ of each $R_{jk}$ ($1 \le j \le m$ and $1 \le k \le |A_j|$) is larger than or equal to the predefined minimum support value $\alpha$. If $R_{jk}$ satisfies the above condition, put it in the set of large 1-itemsets ($L_1$). That is:

$$L_1 = \{R_{jk} | count_{jk} \ge \alpha, 1 \le j \le m \text{ and } 1 \le k \le |A_j|\}.$$

STEP 5: IF $L_1$ is not null, then do the next step; otherwise, exit the algorithm.

STEP 6: Set $r$=1, where $r$ is used to represent the number of items kept in the current large itemsets.

STEP 7: Join the large itemsets $L_r$ to generate the candidate set $C_{r+1}$ in a way similar to that in the apriori algorithm [41] except that two regions (linguistic terms) belonging to the same attribute cannot simultaneously exist in an itemset in $C_{r+1}$. Restated, the algorithm first joins $L_r$ and $L_r$ under the condition that $r$-1 items in the two itemsets are the same and the other one is different. It then keeps in $C_{r+1}$ the itemsets which have all their sub-itemsets of $r$ items existing in $L_r$ and do not have any two items $R_{jp}$ and $R_{jq}$ ($p \ne q$) of the same attribute $R_j$.

STEP 8: Do the following substeps for each newly formed ($r$+1)-itemset $s$ with items $(s_1, s_2, \cdots, s_{r+1})$ in $C_{r+1}$:

(a) Calculate the fuzzy value of each transaction data $D^{(i)}$ in $s$ as $f_s^{(i)} = f_{s_1}^{(i)} \wedge f_{s_2}^{(i)} \wedge \cdots \wedge f_{s_{r+1}}^{(i)}$, where $f_{s_j}^{(i)}$ is the membership value of $D^{(i)}$ in region $s_j$. If the minimum operator is used for the intersection, then:

$$f_s^{(i)} = Min_{j=1}^{r+1} f_{s_j}^{(i)}.$$

(b) Calculate the count of $s$ in the transactions as:

$$count_s = \sum_{i=1}^{n} f_s^{(i)}$$

(c) If $count_s$ is larger than or equal to the predefined minimum support value $\alpha$, put $s$ in $L_{r+1}$.

STEP 9: IF $L_{r+1}$ is null, then do the next step; otherwise, set $r = r+1$ and repeat STEPs 6 to 8.
STEP 10: Collect the large itemsets together.
STEP 11: Construct association rules for each large $q$-itemset $s$ with items ($s_1$, $s_2$, ..., $s_q$), $q \geq 2$, using the following substeps:

(a) Form each possible association rule as follows:

$$s_1 \wedge \cdots \wedge s_{k-1} \wedge s_{k+1} \wedge \cdots \wedge s_q, k = 1\ to\ q.$$

(b) Calculate the confidence values of all association rules using:

$$\frac{\sum_{i=1}^{n} f_s^{(i)}}{\sum_{i=1}^{n} (f_{s_1}^{(i)} \wedge \cdots \wedge f_{s_{k-1}}^{(i)} \wedge f_{s_{k+1}}^{(i)} \wedge \cdots \wedge f_{s_q}^{(i)})}.$$

STEP 12: Output the association rules with confidence values larger than or equal to the predefined confidence threshold $\lambda$.

In the above fuzzy mining approach, all the linguistic terms are used. As an alternative, each item can use only the linguistic term with the maximum cardinality in later mining processes [28]. It can thus keep the same number of items as the original attributes. The alternative therefore focuses on the most important linguistic terms and can reduce its time complexity. Its derived set of association rules is, however, more incomplete than that by considering all the linguistic items. Trade-off thus exists between the rule completeness and the time complexity.

In addition, items may have different importance. Yue et al. thus extended the above concept to find fuzzy association rules with weighted items from transaction data [57]. Each item was given a weight to represent the importance of an item, and each weight was in a range of [0, 1]. They also adopted Kohonen self-organized mapping to derive fuzzy sets for numerical attributes. Weighted supports and weighted conferences were utilized to discover weighted fuzzy association rules.

Hong et al. then proposed a mining approach for extracting interesting weighted fuzzy association rules from transactions, with the parameters (minimum support and minimum confidence) needed in the mining process were given in linguistic terms [6]. Items were evaluated by managers as linguistic terms to reflect their importance, which were then transformed as fuzzy sets of weights. The approach then transformed linguistic weighted items, minimum supports and minimum confidences into fuzzy sets, then filtered weighted large itemsets out by fuzzy operations. Weighted association rules with linguistic supports and confidences were then derived from the weighted large itemsets.

## 4 Mining Fuzzy Generalized Association Rules

Most mining algorithms for association rules focused on single-concept levels. However, mining multiple-concept-level rules may lead to discovery of more specific and important knowledge from data. Relevant item taxonomies are usually predefined in real-world applications and can be represented by hierarchy trees. Terminal nodes on the trees represent actual items appearing in transactions; internal nodes represent classes or concepts formed by lower-level nodes. A simple example is given in Fig. 1.

In this example, food falls into two classes: brink and bread. Brink can be further classified into milk and juice. Similarly, clothes are divided into jackets and T-shirts. Only the terminal items (milk, juice, bread, jacket and T-shirt) can appear in transactions.

Srikant and Agrawal thus proposed a method for finding generalized association rules, which included generalized items (the terminal nodes in taxonomy) [48]. Their mining process could be divided into four phases. In the first phase, ancestors of items in each given transaction were added according to the predefined taxonomy. In the second phase, large itemsets were generated in a way similar to the *Apriori* algorithm. In the third phase, all possible generalized association rules were induced from the large itemsets found in the second phase. The rules with calculated confidence values larger than the minimum confidence were kept. In the fourth phase,

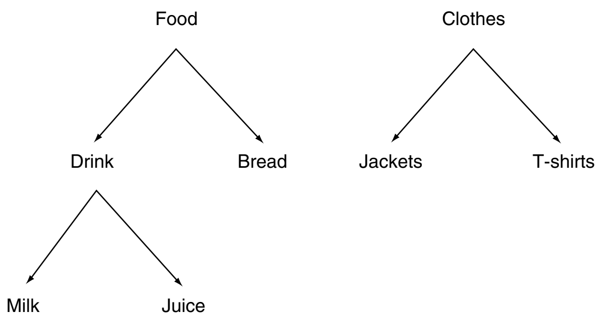



**Fig. 1** An example of predefined taxonomic structures

uninteresting association rules were pruned and interest rules were output according to the following interest requirements:

- A rule that had no ancestor rules (by replacing the items in a rule with their ancestors in the taxonomy) was mined out;
- The support of a rule was *R* times larger than the expected supports of its ancestor rules, where *R* was called a minimum interest;
- The confidence of a rule was *R* times larger than the expected confidences of its ancestor rules.

In addition, Han and Fu also proposed a method for finding level-crossing association rules at multiple levels [1]. Nodes in predefined taxonomies were first encoded using sequences of numbers and the symbol "*" according to their positions in the hierarchy tree. A top-down progressively deepening search approach was then used to explore level-crossing association relationships.

Wei et al. considered the partial relationships possibly existing in taxonomy [56]. An item may partially belong to more than one parent item. For instance, tomato may partially belong to both fruit and vegetable with different degrees. Wei et al. thus defined a fuzzy taxonomic structure and considered the extended degrees of support, conference and interest measures for mining generalized association rules.

Hong et al. proposed a fuzzy multiple-level mining algorithm for extracting implicit knowledge from transactions stored as quantitative values [5]. The proposed generalized fuzzy mining algorithm was based on Srikant and Agrawal's approach to find fuzzy interesting rules from quantitative data. The quantitative items may be from any level of the given taxonomy. They also proposed another fuzzy multiple-level mining algorithm which adopted a top-down progressively deepening approach [28]. Shen et al. devised an algorithm for mining generalized association rules with fuzzy taxonomic structure from quantitative databases based on Wei's algorithm [39]. They handled the quantitative data by Srikant and Agrawal's algorithm [51], which partitioned quantitative attributes into several intervals. Kaya et al. then extended Hong et al's approaches [6, 43] for discovering multi-cross-level fuzzy weighted association rules [9].

In summary, mining fuzzy generalized association rules is a little different from mining fuzzy association rules in the following ways.

1. Non-terminal items must be processed. These items are usually considered expanded items.
2. The candidate set $C_2$ must be processed in a particular way. The two items in an itemset in $C_2$ must not have relationships in the taxonomy. This check needs to be done only for $C_2$. The other candidate sets derived from $C_2$ will not have hierarchical relationships due to the characteristic of sub-itemset checking.
3. Cross-level association rules need to be found.
4. The interestingness of association rules must consider generalization criteria from the taxonomy. Interest requirements are checked to remove uninteresting rules.

## 5 Mining both Membership Functions and Fuzzy Association Rules

The proposed approaches in the above sections mined fuzzy rules under a given set of membership functions. The given membership functions had a critical influence on the final mining results. Although many approaches for learning membership functions were proposed [12, 13, 38, 54], most of them were usually used for classification or control problems. Recently, researches about fuzzy data-mining algorithms for extracting both association rules and membership functions from quantitative transactions have been proposed. For example, Wang et al. tuned membership functions for intrusion detection systems based on similarity of association rules [55]. Kaya et al. proposed a GA-based clustering method to derive a predefined number of membership functions for getting a maximum profit within an interval of user specified minimum support values [7]. In their approach, the membership functions of quantitative attributes were obtained by GAs and were then used to discover fuzzy association rules. Its goal was to output the membership functions which would generate the maximized number of large itemsets. For this purpose, the parameter values of membership functions of the quantitative attributes were encoded into a real-valued string for evolving. Hong et al. proposed several algorithms to dynamically adapt membership functions by genetic algorithms and used them to fuzzify the quantitative transactions [44, 26]. They proposed a GA-based framework for searching membership functions suitable for mining problems and then used the final best set of membership functions to mine association rules. The proposed framework is shown in Fig. 2.

The proposed framework maintained a population of sets of membership functions, and used the genetic algorithm to automatically derive the resulting one. It first transformed each set of membership functions into a fixed-length string. It then chose appropriate strings for "mating", gradually creating good offspring membership function sets. The offspring membership function sets then underwent recursive evolution until a good set of membership functions has been obtained. The fitness was evaluated by the number of large 1-itemsets and the suitability of membership functions. The suitability of membership functions was composed of two terms, overlap and coverage. The overlap ratio of two membership functions was defined as the overlap length divided by half the minimum span of the two functions. If the overlap length was larger than half the span, then these two membership functions were thought of as a little redundant. Appropriate punishment was thus considered in this case. The coverage ratio of a set of membership functions for an item was defined as the coverage range of the functions divided by the maximum quantity of that item in the transactions. The more the coverage ratio was, the better the derived membership functions were. Besides, a larger number of 1-itemsets would usually result in a larger number of all itemsets with a higher probability, which would thus usually imply more interesting association rules. The evaluation by 1-itemsets was, however, faster than that by all itemsets or interesting association rules. Using the number of large 1-itemsets could thus achieve a trade-off between execution time and rule interestingness.

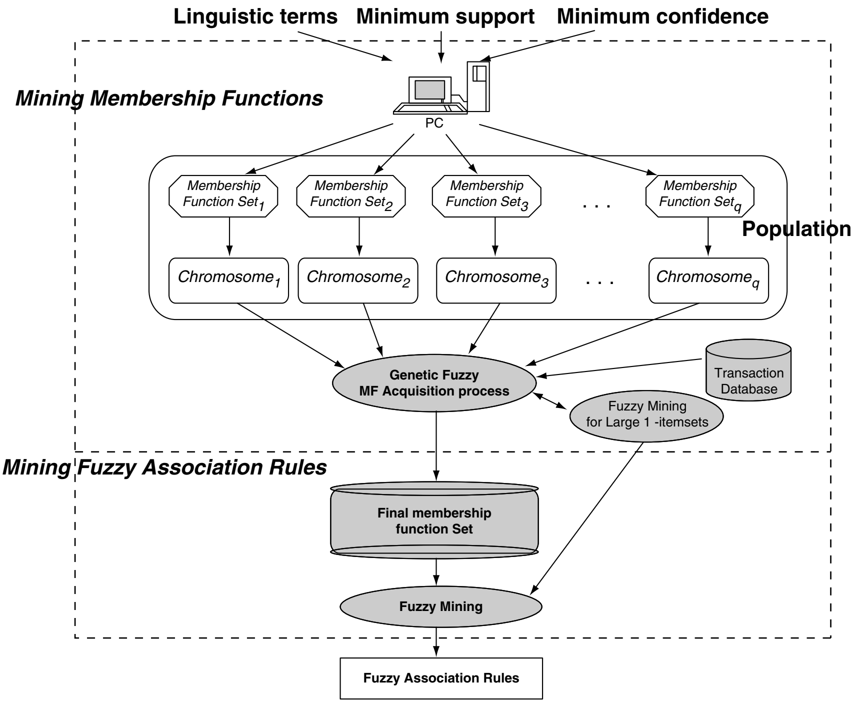



**Fig. 2** GA-based framework for searching membership functions

Hong et al. also proposed an enhanced approach, called cluster-based fuzzy-GA mining algorithm, to speed up the evaluation process and keep nearly the same quality of solutions [15]. The proposed algorithm first divided the chromosomes in a population into $k$ clusters by using the $k$-means clustering approach. All the chromosomes in a cluster then used the number of large 1-itemsets derived from the representative chromosome in the cluster and their own suitability of membership functions to calculate the fitness values. Since the chromosomes with similar coverage and overlap factors would form a cluster, they would have nearly the same shape of membership functions and induce about the same number of large 1-itemsets. For each cluster, the chromosome which was the nearest to the cluster center was thus chosen to derive its number of large 1-itemsets. All chromosomes in the same cluster then used the number of large 1-itemsets derived from the representative chromosome as their own. Finally, each chromosome was evaluated by this number of large 1-itemsets divided by its own suitability value. The evaluation cost could thus be greatly reduced due to the time-saving in finding 1-itemsets.

## 6 Discussion and Conclusion

This chapter has given a wide survey about fuzzy data mining for association rules. Several fuzzy mining techniques, including mining fuzzy association rules, mining fuzzy generalized association rules, mining both membership functions and fuzzy association rules, have been described. Some crisp mining techniques for handling quantitative data have also been briefly reviewed.

Verlinde et al. investigated the difference between the rules by fuzzy mining and crisp mining through experiments and claimed the difference was small from the experimental results [52]. They constrained the rules with only a single item in both the antecedent and the consequence. The experiments were made in the environments where the data were normally distributed and the data boundaries were at the sides of the normal distributions. It is apparent that the rule difference in this situation is small since the data used by fuzzy mining and by crisp mining are not different very much. The boundary effect reduced by the fuzzy mining approaches on data in large databases and with normal distribution is thus not significant. However, as we know, the fuzzy set is usually used for modeling human perception on a concept. The membership functions defined are not necessarily corresponding to the distribution of a data. It is also the reason why the fuzzy set was proposed and not replaced with the probability theory.

In summary, when compared to conventional crisp-set mining methods for quantitative data, fuzzy-mining approaches can get smoother mining results due to the fuzzy membership characteristics. The mined rules are expressed in linguistic terms, which are more natural and understandable for human beings. Besides, nearly all the fuzzy mining approaches can be easily degraded into crisp ones by assigning membership functions with values always equal to 1 or 0. These make fuzzy mining promising in real applications.

**Acknowledgment** The authors would like to thank Mr. Chien-Shing Chen for his help in making the experiments. This research was supported by the National Science Council of the Republic of China under contract NSC93-2213-E-390-001.

## References

1. Agrawal R, Imielinksi T and Swami A (1993) Mining association rules between sets of items in large database. In: Proceedings of the ACM SIGMOD Conference, Washington DC, USA, pp. 207–216.
2. Agrawal R, Srikant R (1994) Fast algorithm for mining association rules. In: Proceedings of the International Conference on Very Large Databases, pp. 487–499.
3. Agrawal R, Srikant R (1995) Mining sequential patterns. In: Proceedings of the 8th IEEE International Conference on Data Engineering, pp. 3–14.
4. Blishun AF (1987) Fuzzy learning models in expert systems. Fuzzy Sets and Systems, 22: 57–70.
5. Cai CH, Fu WC, Cheng CH, Kwong WW (1998) Mining association rules with weighted items. In: Proceedings of the International Database Engineering and Applications Symposium. Cardiff, Wales, UK, pp. 68–77.
6. deCampos LM, Moral S (1993) Learning rules for a fuzzy inference model. Fuzzy Sets and Systems, 59: 247–257.
7. Chan CC, Au WH (1997) Mining fuzzy association rules. Conference on Information and Knowledge Management. Las Vegas, pp. 209–215.

8. Chang RLP, Pavliddis T (1977) Fuzzy decision tree algorithms. IEEE Transactions on Systems, Man and Cybernetics, 7: 28–35.
9. Chen MS, Han J, Yu PS (1996) Data mining: an overview from database perspective. IEEE Transaction on Knowledge and Data Engineering, 8(6): 866–883.
10. Chen CH, Hong TP, Tseng VSM (2006) Cluster-based fuzzy-genetic mining approach for association rules and membership functions. In: Proceedings of the 2006 IEEE International Conference on Fuzzy Systems.
11. Clair C, Liu C, Pissinou N (1998) Attribute weighting: a method of applying domain knowledge in the decision tree process. In: Proceedings of the 7th International Conference on Information and Knowledge Management, pp. 259–266.
12. Clark P, Niblett T (1989) The CN2 induction algorithm. Machine Learning, 3: 261–283.
13. Cordón O, Herrera F, Villar P (2001) Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. In: Proceedings of IEEE Transactions on Fuzzy Systems, 9(4): 667–674.
14. Cordón O, Herrera F, Hoffmann F, Magdalena L (2001) Evolutionary tuning and learning of fuzzy knowledge bases. Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. World Scientific Publishing Company, Singapore.
15. Delgado M, Gonzalez A (1993) An inductive learning procedure to identify fuzzy systems. Fuzzy Sets and Systems, 55: 121–132.
16. Ezeife CI (2002) Mining incremental association rules with generalized FP-tree. In: Proceedings of the 15th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence, pp. 147–160.
17. Famili A, Shen W, Weber R, Simoudis E (1997) Data preprocessing and intelligent data analysis. Intelligent Data Analysis, 1(1): 3–23.
18. Fayyad U, Piatesky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. In: Fayyad U, Piatesky-Shapiro G, Smyth P (eds): Advances in Knowledge Discovery & Data Mining. AAAI/MIT, pp. 1–34.
19. Frawley WJ, Piatetsky-Shapiro G, Matheus CJ (1991) Knowledge discovery in databases: an overview. The AAAI Workshop on Knowledge Discovery in Databases, pp. 1–27.
20. Fukuda T, Morimoto Y, Morishita S, Tokuyama T (1996) Mining optimized association rules for numeric attributes. The ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 182–191.
21. Gonzalez A (1995) A learning methodology in uncertain and imprecise environments. International Journal of Intelligent Systems, 10: 57–371.
22. Han J, Fu Y (1995) Discovery of multiple-level association rules from large database. In: Proceedings of the 21st International Conference on Very Large Data Bases, Zurich, Switzerland, pp. 420–431.
23. Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. The 2000 ACM SIGMOD International Conference on Management of Data, 29(2): 1–12.
24. Hong TP, Lee CY (1996) Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets and Systems, 84: 33–47.
25. Hong TP, Kuo CS, Chi SC (1999) A data mining algorithm for transaction data with quantitative values. In: Proceedings of the 8th International Fuzzy Systems Association World Congress, pp. 874–878.
26. Hong TP, Kuo CS, Chi SC (1999) Mining association rules from quantitative data. Intelligent Data Analysis, 3(5): 363–376.
27. Hong TP, Kuo CS, Chi SC (2001) Trade-off between computation time and number of rules for fuzzy mining from quantitative data. International Journal of Uncertainty. Fuzziness and Knowledge-based Systems, 9(5): 587–604.
28. Hong TP, Chiang MJ, Wang SL (2002) Mining from quantitative data with linguistic minimum supports and confidences. In: Proceedings of the 8th IEEE International Conference on Fuzzy Systems, pp. 494–499.
29. Hong TP, Lin KY, Chien BC (2003) Mining fuzzy multiple-level association rules from quantitative data. Applied Intelligence, 18(1): 79–90.

30. Hong TP, Lin KY, Wang SL (2003) Fuzzy data mining for interesting generalized association rules. Fuzzy Sets and Systems, 138(2): 255–269.
31. Hong TP, Chen CH, Wu YL, Lee YC (2003) Mining membership functions and fuzzy association rules. In: Proceedings of 2003 The Joint Conference on AI, Fuzzy System, and Grey System.
32. Hong TP, Chen CH, Wu YL, Lee YC (2006) A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions. Soft Computing, 10(11): 1091–1101.
33. Hong TP, Lin JW, Wu YL (2006) A Fast Updated Frequent Pattern Tree. In: Proceedings of the 17th IEEE International Conference on Systems, Man, and Cybernetics.
34. Kandel A (1992) Fuzzy Expert Systems. CRC Press, Boca Raton, pp. 8–19.
35. Kaya M, Alhajj R (2003) A clustering algorithm with genetically optimized membership functions for fuzzy association rules mining. In: Proceedings of The IEEE International Conference on Fuzzy Systems, pp. 881–886.
36. Kaya M, Alhajj R (2004) Mining multi-cross-level fuzzy weighted association rules. In: Proceedings of IEEE International Conference Intelligent Systems, Varna, Bulgaria, pp. 225–230.
37. Kuok C, Fu A, Wong M (1998) Mining fuzzy association rules in databases. SIGMOD Record, 27(1):41–46.
38. Lent B, Swami A, and Widom J (1997) Clustering association rules. In: Proceedings of International Conference on Data Engineering (ICDE'97), Birmingham, England, pp. 220–231.
39. Mamdani EH (1974) Applications of fuzzy algorithms for control of simple dynamic plants. Proceedings of IEEE, 121: 1585–1588.
40. Mannila H, Toivonen H, Verkamo AI (1994) Efficient algorithm for discovering association rules. The AAAI Workshop on Knowledge Discovery in Databases, pp. 181–192.
41. Mannila H (1997) Methods and problems in data mining. In: Proceedings of the 6th International Conference on Database Theory (ICDT'97), LNCS, Springer-Verlag, 1186: 41–55.
42. Park JS, Chen MS, Yu PS (1997) Using a hash-based method with transaction trimming for mining association rules. IEEE Transactions on Knowledge and Data Engineering, 9(5): 812–825.
43. Piatetsky-Shapiro G (1991) Discovery, analysis, and presentation of strong rules. Knowledge Discovery in Databases, AAAI/MIT Press, pp. 229–248.
44. Qiu Y, Lan YJ, Xie QS (2004) An improved algorithm of mining from FP-tree. In: Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai, pp. 26–29.
45. Quinlan JR (1987) Decision tree as probabilistic classifier. In: Proceedings of the 4th International Machine Learning Workshop, Morgan Kaufmann, San Mateo, CA., pp. 31–37.
46. Rastogi R, Shim K (1998) Mining optimized association rules with categorical and numeric attributes. In: Proceedings of the 14th IEEE International Conference on Data Engineering, Orlando, pp. 503–512.
47. Rives J (1990) FID3: fuzzy induction decision tree. In: Proceedings of the 1st International symposium on Uncertainty, Modeling and Analysis, pp. 457–462.
48. Setnes M, Roubos H (2000) GA-fuzzy modeling and classification: Complexity and performance. IEEE Transaction on Fuzzy System, 8(5): 509–522.
49. Shen H, Wang S, Yang J (2004) Fuzzy taxonomic, quantitative database and mining generalized association rules. In: Proceedings of the 4th International Conference on Rough Sets and Current Trends in Computing (RSCTC 2004), Uppsala, Sweden, pp. 610–617.
50. Srikant R, Agrawal R (1995) Mining generalized association rules. In: Proceedings of the 21st International Conference on Very Large Data Bases, pp. 407–419.
51. Srikant R, Agrawal R (1996) Mining quantitative association rules in large relational tables. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Monreal, Canada, pp. 1–12.
52. Verlinde H, Cock MD, Boute R (2006) Fuzzy versus quantitative association rules: a fair data driven comparison. IEEE Transactions on Systems, Man and Cybernetics, 36(3): 679–684.
53. Wang CH, Hong TP, Tseng SS (1996) Inductive learning from fuzzy examples. In: Proceedings of the fifth IEEE International Conference on Fuzzy Systems, New Orleans, pp. 13–18.

54. Wang CH, Hong TP, Tseng SS (1998) Integrating fuzzy knowledge by genetic algorithms. IEEE Transactions on Evolutionary Computation, 2(4): 138–149.
55. Wang W, Bridges SM (2000) Genetic algorithm optimization of membership functions for mining fuzzy association rules. In: Proceedings of the International Joint Conference on Information Systems, Fuzzy Theory and Technology, pp. 131–134.
56. Wei Q, Chen G (1999) Mining generalized association rules with fuzzy taxonomic structures. In: Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society (NAFIPS), NY, USA, pp. 477–481.
57. Yue S, Tsand E, Yeung D, Shi D (2000) Mining fuzzy association rules with weighted items. In: Proceedings of the IEEE International Conference on System, Man and Cybernetics, pp. 1906–1911.
58. Zadeh LA (1965) Fuzzy sets. Information and Control, 8(3): 338–353.
59. Zadeh LA (1988) Fuzzy logic. IEEE Computer, pp. 83–93.
60. Zaïane OR, Mohammed EH (2003) COFI-tree mining: A new approach to pattern growth with reduced candidacy generation. In: Proceedings of the ICDM 2003 Workshop on Frequent Itemset Mining Implementations.

# Subgroup Discovery with Linguistic Rules

**María José del Jesus, Pedro González and Francisco Herrera**

**Abstract** Subgroup discovery can be defined as a form of supervised inductive learning in which, given a population of individuals and a specific property of individuals in which we are interested, find population subgroups that have the most unusual distributional characteristics with respect to the property of interest. Subgroup discovery algorithms aim at discovering individual rules, which must be represented in explicit symbolic form and which must be simple and understandable in order to be recognized as actionable by potential users.

A fuzzy approach for a subgroup discovery process, which considers linguistic variables with linguistic terms in descriptive fuzzy rules, lets us obtain knowledge in a similar way of the human thought process. Linguistic rules are naturally inclined towards coping with linguistic knowledge and to produce more interpretable and actionable solutions. This chapter analyzes the use of linguistic rules for modelling this problem, and shows a genetic extraction model for learning this kind of rules.

## 1 Introduction

Rule learning is an important form of *predictive* machine learning, aimed at inducing a set of rules to be used for classification and/or prediction [6, 31]. Developments in *descriptive induction* have recently also attracted much attention from researchers interested in rule learning. The objective of *descriptive machine learning* is to discover individual rules that define interesting patterns in data, and it includes approaches for mining association rules [2], for subgroup discovery [24, 35] and other non-classificatory induction approaches such as clausal discovery [34] or database dependency [15] among others.

Subgroup discovery is a form of descriptive supervised inductive learning. It aims to discover individual rules (or local patterns of interest, very frequent –hence typical– or very rare –hence atypical–) in relation to a specific property of interest, which must be represented in explicit symbolic form and which must be relatively simple in order to be recognized as actionable by potential users. Therefore, the subgroups discovered in data are of a more explanatory nature and the interpretability of the extracted knowledge for the final user is a crucial aspect in this field.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

As it was claimed by Dubois et al. in [14], the use of fuzzy sets to describe associations between data extends the types of relationships that may be represented, facilitates the interpretation of rules in linguistic terms, and avoids unnatural boundaries in the partitioning of the attribute domains. This is especially useful in medical, control or economic fields where the boundaries of a piece of information used may not be clearly defined. In fact, the use of linguistic variables and linguistic terms in a machine learning process has been thoroughly explored by various authors in predictive induction (see for instance Ishibuchi et al.'s book [22] for a complete and understandable up-to-date description of the design of classification and modelling fuzzy systems). There are some proposals using fuzzy logic in descriptive induction, for the extraction of fuzzy association rules [10, 20], and for subgroup discovery fuzzy rules [12, 13].

A fuzzy approach for a subgroup discovery process, which considers linguistic variables with linguistic terms in descriptive fuzzy rules, allows us to obtain knowledge in a similar way to the human thought process. In order to understand this it is enough to remember that much of the logic behind human reasoning is not traditional two-valued or even multivalued logic, but logic with fuzzy truths, fuzzy connectives and fuzzy rules of inference. Fuzzy rules are naturally inclined towards coping with linguistic knowledge, thereby producing more interpretable and actionable solutions in the field of subgroup discovery and in general in the analysis of data to establish relationships and identify patterns [21].

This chapter analyzes the use of linguistic rules in subgroup discovery. A genetic model for the extraction of fuzzy rules in subgroup discovery [12, 13] is described, analyzing its possibilities and limitations. To do so, the chapter is arranged in the following way: In Sect. 2, the subgroup discovery task is introduced. In Sect. 3 is described the use of linguistic rules in the subgroup discovery task. An evolutionary approach to obtain subgroup discovery descriptive fuzzy rules is explained in Sect. 4. Finally, in Sect. 5 the conclusions and further research are outlined.

## 2 Introduction to Subgroup Discovery

Subgroup discovery is a form of supervised inductive learning which is defined as follows [24, 35]: given a set of data and a property of interest to the user (target variable), an attempt is made to locate subgroups which are statistically "most interesting" for the user, e.g., are as large as possible and have the most unusual distributional characteristics with respect to the property of interest. The concept was initially formulated by Klösgen in EXPLORA [24] and by Wrobel in MIDOS [35].

Descriptive machine learning methods for subgroup discovery have the objective of discovering interesting properties of subgroups by obtaining *simple* rules (i.e. with an understandable structure and with few variables), which are *highly significant* and with *high support* (i.e. covering many of the instances of the target class).

An induced subgroup description has the form of an implication,

$$\text{Cond} \rightarrow \text{Class}$$

where the property of interest for subgroup discovery is class value *Class* that appears in the rule consequent, and the rule antecedent *Cond* is a conjunction of features (attribute-value pairs) selected from the features describing the training instances.

The subgroup discovery task relies on the following main properties:

- The description language specifying the subgroups which must be adequate to be applied effectively by the potential users. The subgroup description consists of a set of expressions. In the simplest case, each expression is one-valued; however negation or internal disjunctions are also possible.
- The quality function measuring the interest of the subgroup. A variety of quality functions have been proposed (see for instance [18, 24, 25]). The quality functions used is determined by the type of the target variable, the type of rules and the problem considered. In subsection 2.2 several quality measures used in subgroup discovery algorithms are described.
- The search strategy employed by the algorithm is very important, since the dimension of the search space has an exponential relation with respect to the number of features (or variables) and values considered.

Below related works and the quality measures used in subgroup discovery are shortly revised.

## 2.1 Related Works in Subgroup Discovery

In the specialized bibliography, different methods have been developed which obtain descriptions of subgroups represented in different ways and using different quality measures:

- The first approach developed for subgroup discovery was EXPLORA [24]. It uses decision trees for the extraction of rules. The rules are specified by defining a descriptive schema and implementing a statistical verification method. The interest of the rules is measured using measures such as evidence, generality, redundancy and simplicity.
- MIDOS [35] applies the EXPLORA approach to multirelational databases. It uses optimistic estimation and minimum support pruning. The goal is to discover subgroups of the target relation which have unusual statistical distributions with respect to the complete population. The quality measure is a combination of unusualness and size.
- Subgroup Miner [26] is an extension of EXPLORA and MIDOS. It is an advanced subgroup discovery system which uses decision rules and interactive search in the space of the solutions, allowing the use of large databases, multirelational hypotheses, and the discovery of structures of causal subgroups. This algorithm uses as quality function the classical binomial test to verify if the statistical distribution of the target is significantly different in the extracted subgroup.

- SD [17] is a rule induction system guided by expert knowledge: instead of defining an optimal measure to search and select automatically the subgroups, the objective is to help the expert in performing flexible and effective searches on a wide range of optimal solutions.
- CN2-SD [29] (a modified version of the CN2 algorithm [6]) induces subgroups in the form of rules using the relation between true positives and false positives as a quality measure. It uses a modified weighted relative accuracy as quality measure for the rule selection.
- RSD [30], *Relational Subgroup Discovery*, has the objective of obtaining population subgroups which are as large as possible, with a statistical distribution as unusual as possible with respect to the property of interest, and which are different enough to cover most of the target population. It is a recent upgrade of the CN2-SD algorithm which enables relational subgroup discovery.
- APRIORI-SD [23] is developed by adapting the association rule learning algorithm APRIORI [1] to subgroup discovery, including a new quality measure for the induced rules (the weighted relative accuracy) and using probabilistic classification of the examples. For the evaluation is used the support and significance of each individual rule, and the size, accuracy and area under the ROC curve of the set of rules.
- *Intensive Knowledge* [3] is a subgroup discovery approach which uses several types of application background knowledge to improve the quality of the results of the subgroup discovery task and the efficiency of the search method.
- SDIGA [13] is an evolutionary fuzzy rule induction system which uses as quality measures for the subgroup discovery task adaptations of the measures used in the association rules induction algorithms. Unlike all the other proposals, SDIGA uses linguistic rules as description language to specify the subgroups. This proposal is shown in Sect. 4.

## *2.2 Quality Measures in Subgroup Discovery*

One of the most important aspects of any subgroup discovery algorithm -and a determining factor in the quality of the approach- is the quality measure to be used, both to select the rules and to evaluate the results of the process. *Objective* measures for descriptive induction evaluate each subgroup individually, but can be complemented by their variants to compute the mean of the induced set of descriptions of subgroups, allowing comparison between different subgroup discovery algorithms.

There are different studies about objective quality measures for the descriptive induction process [17, 25, 32] but it is difficult to reach an agreement about their use. Below, the more widely used objective quality measures in the specialized bibliography of subgroup discovery are described.

- *Coverage for a rule* [29]: this measures the percentage of examples covered on average by one rule of the induced set of rules.

$$Cov(R^i) = Cov(Cond^i \rightarrow Class_j) = p(Cond^i) = \frac{n(Cond^i)}{n_s} \tag{1}$$

where $n(Cond^i)$ is the number of examples which verifies the condition $Cond^i$ described in the antecedent (independently of the class to which belongs), and $n_s$ is the number of examples.

The *average coverage for the set of rules* finally obtained is calculated by the following expression:

$$COV = \frac{1}{n_r}\sum_{i=1}^{n_r} Cov(R^i) \tag{2}$$

where $n_r$ is the number of induced rules.

- *Support for a rule*: In descriptive induction processes the support for a rule is a standard measure which considers, by means of an expression that can vary in different proposals, the number of examples satisfying both the antecedent and the consequent parts of the rule. Lavrac et al. compute in [29] the overall support as the percentage of target examples (positive examples) covered by the rules. The support of a rule is so defined as the frequency of correctly classified examples covered.

$$Sup_1(Cond^i \rightarrow Class_j) \ = p(Class_j.Cond^i) = \frac{n(Class_j.Cond^i)}{n_s} \tag{3}$$

where $n(Class_j.Cond^i)$ is the number of examples which satisfy the conditions for the antecedent ($Cond^i$) and also belong to the value for the target variable ($Class_j$) indicated in the consequent part of the rule. In [3], the support of a rule is computed dividing by the total number of examples. It can also be computed in other ways, such as dividing by the number of examples of the class or other variations.

The *support* for a set of rules is computed by:

$$SUP = \frac{1}{n_s}\sum_{j=1}^{n_c} n(Class_j \cdot \bigvee_{Cond^i \rightarrow Class_j} Cond^i) \tag{4}$$

where $n_c$ is the number of values for the target variable considered. It must be noted that in this expression the examples which belong to many rules are considered only once.

- *Size (for a set of rules)*: The size of a set of rules is a complexity measure calculated as the number of induced rules ($n_r$). Complexity can also be measured as the mean number of obtained rules per class, or the mean of variables per rule.
- *Significance for a rule* [24]: indicates the significance of a finding, if measured by the likelihood ratio of a rule.

$$Sig(Cond^i \rightarrow Class_j) = 2 \cdot \sum_{j=1}^{n_c} n(Class_j.Cond^i) \cdot \log \frac{n(Class_j.Cond^i)}{n(Class_j) \cdot p(Cond^i)} \quad (5)$$

where $p(Cond_i)$, computed as $n(Cond_i)/n_s$, is used as a normalized factor.

It must be noted that, although each rule is for a specific class value, the significance measures impartially the novelty in the distribution, for all the class values.

The *significance for a set of rules* is computed as follows:

$$SIG = \frac{1}{n_r} \sum_{i=1}^{n_r} Sig(R^i) \quad (6)$$

- *Unusualness for a rule*: It is defined as the *weighted relative accuracy* of a rule [28].

$$WRAcc(Cond^i \rightarrow Class_j) = \frac{n(Cond^i)}{n_s} \cdot \left( \frac{n(Class_j.Cond^i)}{n(Cond^i)} - \frac{n(Class_j)}{n_s} \right) \quad (7)$$

The weighted relative accuracy of a rule can be described as the balance between the coverage of the rule ($p(Cond^i)$) and its accuracy gain ($p(Class_j.Cond^i)$ - $p(Class_j)$). It must be noted that the higher a rule's unusualness, the more relevant is it.

The unusualness for a set of rules is computed as follows:

$$WRACC = \frac{1}{n_r} \sum_{i=1}^{n_r} WRAcc(R^i) \quad (8)$$

It must be noted that all the measures here described are crisp because in the majority of the proposals the rules used to represent the knowledge in subgroup discovery are not fuzzy.

## 3 Linguistic Rules in Subgroup Discovery

As it has been described in the previous section many approaches have already been proposed for subgroup discovery task, usually based on non linguistic rules. Since human information processing is mainly based on linguistic information, in order to facilitate the human interpretability of the results, the use of linguistic rules must be considered.

In this section, the use of linguistic rules in subgroup discovery will be analyzed, and a kind of linguistic rules, DNF linguistic rules, and some quality measures for them are described.

### 3.1 The Use of Linguistic Rules in Subgroup Discovery

In any Data Mining problem two main objectives are present:

- to obtain knowledge about patterns in data which must be fitted to the nature and reality of the problem, e.g., knowledge must be as precise as possible,
- to extract knowledge which must be simple, compact and understandable by the final user. That is to say, the obtained knowledge must be close to the form in which the expert represents his knowledge on the problem in order to be actionable by him.

The second objective becomes the most important in descriptive data mining, and specifically in the subgroup discovery task.

The way in which the knowledge is represented by a human expert is inherently qualitative and vague. In this sense the use of Fuzzy Logic in Data Mining allows us to model inaccurate and qualitative knowledge, as well as to handle uncertainty and deal naturally to a reasonable extent with human reasoning. Ever since it was proposed in 1965 by Zadeh [36], it has been applied to many areas of research, fundamentally because of its proximity to human reasoning and because it provides an effective way of capturing the approximate and inexact nature of the real world.

In rule induction processes, Fuzzy Logic is included in such a way that the models extracted are fuzzy rules. In the most interpretable type of fuzzy rules, linguistic fuzzy rules, and therefore the most appropriate for Data Mining, the continuous variables are defined as linguistic variables; that is, variables which take as possible values linguistic labels, the semantics of which are represented by an associated fuzzy set [37].

One of the fundamental aspects when working with linguistic rules is the definition of membership functions associated with the fuzzy sets used. There are several alternatives to determine this aspect:

- When the expert knowledge is not available, uniform partitions with triangular membership functions can be used, as it is shown in Fig. 1 for a variable with 5 linguistic labels.
- When expert knowledge about the problem is available or an analysis of the data can be realized, the definition for the fuzzy partition can be done in one of the following ways:

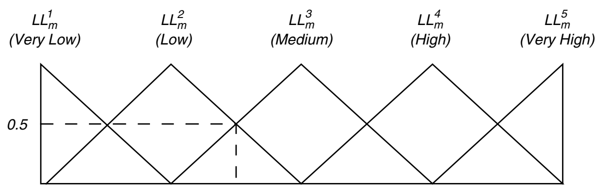



**Fig. 1** Example of fuzzy partition for a continuous variable

- In order to increase the interpretability of the results obtained in some proposals such as [4, 27] for the extraction of fuzzy rules the expert gives the algorithm the continuous variables and their corresponding membership functions. The quality of the results obtained depends on the suitability of the fuzzy sets.
- For many applications it is very difficult to know from the outset which fuzzy sets will be the most appropriate and so different algorithms which learn the fuzzy partitions have been proposed. In [16] the fuzzy sets and the membership functions are generated through clustering techniques. In [19] the definition for the linguistic labels is established by means a genetic algorithm.
- A fuzzy partition can be defined by a heuristic approach which places the fuzzy sets in such a way that each of them will cover approximately the same number of data, if the expert wants to. But it must be considered that, depending on the problem, the interpretation of the resulting fuzzy rules could be decreased.
- Moreover, if it is necessary, a preliminary data analysis which detects outliers in data can be done before the determination of the fuzzy partitions. This way a specific analysis of them can be realized and the fuzzy partition (without these outliers' data) is not biased by them.

## *3.2 DNF Linguistic Rules*

The objective in subgroup discovery is to extract knowledge about a variable of interest for the user, in an easily interpretable way. In order to increase the interpretability of the extracted knowledge, the Disjunctive normal form (DNF) fuzzy rules can be used. A DNF fuzzy rule represents the knowledge in a flexible and compact way, allowing each variable to take more than one value, and facilitating the extraction of more general rules. Linguistic rules allow us to establish flexible limits between the different levels of meaning without ignoring or overemphasizing the elements closest to the edges in the same way as human perception does. In addition, linguistic DNF fuzzy rules allow us to make changes in the initial granularity in each rule in a descriptive way. The following is an example of linguistic DNF fuzzy rule:

IF Number of times pregnant *High* or *Medium* AND Body mass index is *Low* THEN Diabetes is *Tested negative*

Below, the notation used to describe the DNF fuzzy rules is formally described. We consider a problem with:

- a set of features, discrete or continuous

$$\{X_m \ / \ m = 1, .., n_v\}$$

  used to describe the subgroups, where $n_v$ is the number of features;
- a set of values for the target variable

$$\left\{Class_j / j = 1, \dots, n_c\right\}$$

  where $n_c$ is the number of values for the target variable considered;

- a set of examples

$$\left\{ E^k = (e_1^k, e_2^k, \ldots, e_{n_v}^k, class^j)/k = 1, \ldots, n_s \right\}$$

where $class^j$ is the target variable value for the sample $E^k$(i.e., the class for this example) and $n_s$ is the number of examples for the descriptive induction process;
- a set of linguistic labels for the continuous variables. The number of linguistic labels and the definition for the corresponding fuzzy sets depend on each variable

$$X_m : \left\{ LL_m^1, LL_m^2, \ldots, LL_m^{l_m} \right\}$$

In this expression, variable $X_m$, has $l_m$ different linguistic labels to describe its domain in an understandable way.

A fuzzy rule $R^i$ can be described as:

$$R^i : Cond^i \rightarrow Class_j$$

where the antecedent describes the subgroup.

Below is an example of a DNF fuzzy rule:

$$R^1 : \text{If } (X_1 \text{ is } LL_1^1 \text{ or } LL_1^3) \text{ and } (X_7 \text{ is } LL_7^1) \text{ then } Class_j \qquad (9)$$

It must be noted that, in the DNF rule, any subset of the complete set of variables (with any combination of linguistic labels related with the operator OR) can take part in the rule antecedent. In this way a subgroup is a compact and interpretable description of patterns of interest in data.

For these rules, we consider that

- an example $E^k$ verifies the antecedent part of a rule $R^i$ if

$$APC(E^k, R^i) = T(TC(\mu_{LL_1^1}(e_1^k), ..., \mu_{LL_1^{l_1}}(e_1^k)), ..., TC(\mu_{LL_{n_v}^1}(e_{n_v}^k), ..., \mu_{LL_{n_v}^{l_{n_v}}}(e_{n_v}^k))) > 0 \qquad (10)$$

where:
  - APC (Antecedent Part Compatibility) is the degree of compatibility between an example and the antecedent part of a fuzzy rule, i.e., the degree of membership for the example to the fuzzy subspace delimited by the antecedent part of the rule,
  - $LL_1^{l_1}$ is the linguistic label number $l_1$ of the variable *1*,
  - $\mu_{LL_1^{l_1}}(e_1^k)$ is the degree of membership for the value of the feature *1* for the example $E^k$ to the fuzzy set corresponding to the linguistic label $l_1$ for this feature,

  - T is the t-norm selected to represent the meaning of the AND operator –the fuzzy intersection–, in our case the minimum t-norm, and
  - TC is the t-conorm selected to represent the meaning of the OR operator –the fuzzy union–, in our case the maximum t-conorm.
- an example $E^k$ is covered by a rule $R^i$ if

$$APC(E^k, R^i) > 0 \quad AND \quad E^k \in Class_j \tag{11}$$

This means that an example is covered by a rule if the example has a degree of membership higher than 0 to the fuzzy subspace delimited by the antecedent part of the fuzzy rule, and the value indicated in the consequent part of the rule agrees with the value of the target feature for the example. For the categorical variables, the degrees of membership are just 0 or 1.

## 3.3 Quality Measures for DNF Linguistic Rules

When using linguistic rules, it is necessary to define quality measures to manage this type of rules. Some of the quality measures used in the bibliography for the induction of fuzzy rules are next detailed:

- *Confidence of a fuzzy rule* [13]: The confidence of a rule determines the relative frequency of examples satisfying the complete rule among those satisfying only the antecedent. In our proposal the expression used for confidence reflects the degree to which the examples within the zone of the space marked by the antecedent verify the information indicated in the consequent part of the rule. To calculate this factor an adaptation of Quinlan's accuracy expression [33] is used in order to generate fuzzy classification rules [8]: the sum of the degree of membership of the examples of this class and the fuzzy input subspace determined by the antecedent, divided by the sum of the degree of membership of all the examples that verifies the antecedent part of this rule (irrespective of their class) to the same zone:

$$Conf(R^i) = \frac{\sum\limits_{E^k \in E / E^k \in Class_j} APC(E^k, R^i)}{\sum\limits_{E^k \in E} APC(E^k, R^i)} \tag{12}$$

- *Support of a fuzzy rule*, defined in [13] as the degree of coverage that the rule offers to examples of that class:

$$Sup_2(R^i) = \frac{n(Class_j.Cond^i)}{n(Class_j)} \tag{13}$$

where $n(Class_j)$ is the number of examples of the class *j*. A variation of this measure will be detailed in next section.

# 4 A Genetic Algorithm for the Induction of Linguistic Rules in Subgroup Discovery

In this section an evolutionary model for the extraction of linguistic rules for the subgroup discovery task, SDIGA (Subgroup Discovery Iterative Genetic Algorithm), which uses DNF rules is described [13]. The model follows the IRL approach –later explained– and works as follows:

- The core of the model is a genetic algorithm (GA) which uses a post-processing step based on a simple local search, a hill-climbing procedure. The hybrid GA extracts one simple and interpretable fuzzy rule with an adequate level of support and confidence. The post-processing step consists of a local search process increasing the generality of the rule.
- This hybrid GA is included in an iterative process for the extraction of a set of fuzzy rules for the description of subgroups supported by different areas (not necessarily disjuncts) of the instance space. In this way is obtained a set of *different* solutions generated in successive runs of the GA corresponding to the same value of the target feature. The method to guide the GA evolution over different –although may be overlapped– fuzzy rules is explained in detail in the next subsection.

The objective is to obtain a set of rules which describe subgroups for all the values of the target feature, and so the iterative process must be carried out as many times as different values the target feature has.

Once the basis of the proposal is outlined, the GA and the iterative rule extraction model are described in detail. The results of a comparison of the proposal with other subgroup discovery algorithms are also detailed.

## *4.1 Hybrid Genetic Algorithm for the Induction of a Fuzzy Rule*

The hybrid GA extracts a single fuzzy rule in an attempt to optimize the confidence and support. In the following subsections the elements of the hybrid GA are described.

### 4.1.1 Chromosome Representation

The genetic representation of the solutions is the most determining aspect of the characteristics of any genetic learning proposal. The "*Chromosome = Rule*" approach (in which each individual codifies a single rule) is more suited in subgroup discovery because the objective is to find a reduced set of rules in which the quality of each rule is evaluated independently of the rest. This is the encoding approach used in the evolutionary proposal next described.

The GA discovers a single fuzzy rule whose consequent is prefixed to one of the possible values of the target feature. Only the antecedent is represented in the chromosome and all the individuals in the population are associated with the same value of the target feature.

All the information relating to a rule is contained in a fixed-length chromosome with a binary representation in which, for each feature a bit for each one of the possible values of the feature is stored; in this way, if the corresponding bit contains the value 0 it indicates that the value is not used in the rule, and if the bit contains the value 1 it indicates that the corresponding value is included. If a rule contains all the bits corresponding to a feature with the value 1, or all of them contain the value 0, the feature is ignored and does not take part in the rule. In Fig. 2, $V_0$ and $V_1$ have 3 possible values, and $V_2$ and $V_k$ have 2 possible values. In this example, neither $V_2$ nor $V_k$ take part in the rule ($V_2$ does not take any of its values, and $V_k$ takes all, and so both variables are irrelevant for the rule).

### 4.1.2 Fitness Function

The objective of the rule discovery process is to obtain rules with high confidence, and which are understandable and general. It means that the problem has at least two objectives to maximize: the support and the confidence of the rule. To achieve this, the weighted sum method that weights a set of objectives into a single objective is the simplest approach, and lets us introduce the expert criteria related to the importance of the objectives for a specific problem in the rule generation process. So, this proposal uses a weighted lineal combination in the following way:

$$fitness\ (c) = \frac{\omega_1 \times Sup_3(c) + \omega_2 \times Conf(c)}{\omega_1 + \omega_2} \tag{14}$$

where confidence (*Conf*) and support ($Sup_3$) of the rule are defined as:

- *Confidence*. This determines the accuracy of the rule, in that it reflects the degree to which the examples within the zone of the space determined by the antecedent verify the information specified in the consequent of the rule, and it is computed as in [12].
- *Support*. This measures the degree of coverage that the rule offers to examples belonging to the class specified in the rule consequent. It is calculated in a different way than in [5] to promote different fuzzy rules being obtained in different runs of the hybrid GA. To do so, for the computation of the support it is only considered the examples not marked (i.e. the examples not covered by other fuzzy rules previously obtained by means of the past runs of the hybrid GA). Thus, the support is defined as the quotient between the examples of this partial set covered by the rule represented in the chromosome and the total number of examples of this partial set:

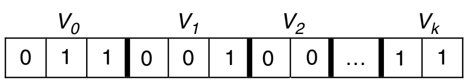



**Fig. 2** Encoding model of a DNF rule

$$Sup_3(R^i) = \frac{Ne^+(R^i)}{Ne_{NC}} \tag{15}$$

where $Ne_{NC}$ is the number of examples of the class specified in the consequent left uncovered by the previous rules, and $Ne^+(R^i)$ is the number of examples covered by the rule which are left uncovered by the previous rules, using (11) to determine when an example is covered by a rule.

This way of measuring support is sensible, when using the GA within an iterative process, in order to obtain different rules each time the GA is run. From the second iteration, rules which cover examples belonging to zones delimited by previously obtained rules are penalized, because the support factor only considers examples which have not been described by rules already obtained. No distance function is used as differences are penalized on a phenotypical level. This penalization does not eliminate the examples covered by previously obtained fuzzy and they take part in the computation of the confidence measure.

The overall objective of the evaluation function is to direct the search towards rules which maximize accuracy, minimizing the number of negative and examples not-covered. Whereas covered examples are used in the calculation of the confidence, they are not used in the calculation of the support, to prevent the obtaining of rules inconsistent with the examples previously penalized.

### 4.1.3 Reproduction Model and Genetic Operators

The GA includes a steady-state reproduction model [5], in which the original population is only modified through the substitution of the worst individuals by individuals resulting from crossover and mutation. The recombination is carried out by means of a two-point crossover operator and a biased random mutation operator.

The crossover is applied over the two best individuals of the population, obtaining two new individuals, which will substitute the two worst individuals in the population. This strategy leads to a high selective pressure with the aim of getting a quick convergence of the algorithm.

Mutation is carried out as follows. First, according to the mutation probability, the chromosome and the gene of the chromosome to be muted are determined. Then, the biased random mutation operator is applied in two different ways, with probability 0.5 in each case. In the first way, the mutation causes the elimination of the variable to which the gene corresponds, setting to 0 all the values of this variable, as is shown in Fig. 3a. The second type of mutation randomly assigns 0 or 1 to all the values of the variable, as can be seen in Fig. 3b. So, half the mutations have the effect of eliminating the corresponding variable, and the rest randomly set the values for the variable to be muted.

The mutation is applied according to the mutation probability not only on the two best individuals in the population but on all the population. In order to obtain diversity by means of the application of this operator, a greater population size than the habitual in steady-state evolutionary models must be selected.

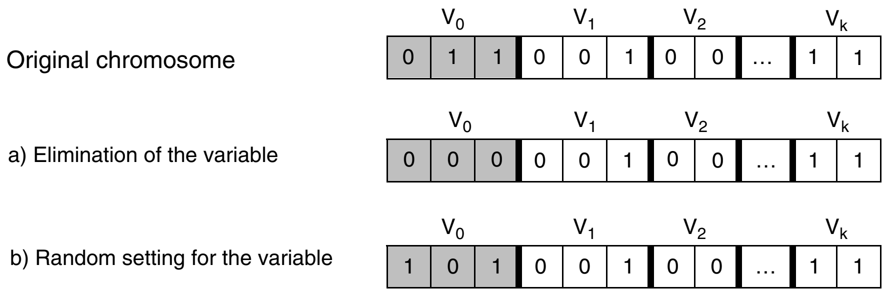



**Fig. 3** Types of mutation for a variable in a DNF rule

### 4.1.4 Hybrid GA Post-processing Phase: Local Search Algorithm

The post-processing phase, which improves the obtained rule by a hill-climbing process, modifies the rule in order to increase the degree of support. To accomplish this, in each iteration a variable is selected such that when it is eliminated, the support of the resulting rule is increased; in this way more general rules are obtained. Finally, the optimized rule will substitute the original only if it overcomes minimum confidence.

The diagram of the post-processing phase is as shown in Fig. 4.

## *4.2 Iterative Rule Extraction Model*

The fuzzy descriptive rule extraction model follows the Iterative Rule Learning (IRL) approach [9], in which each chromosome represents a rule, but the GA solution is the best individual obtained and the global solution is formed by the best individuals obtained when the algorithm is run multiple times. The objective of the model is to obtain a set of rules giving information on the majority of available examples for each value of the target feature.

The data mining process is carried out by means of an iterative algorithm allowing the generation of several rules (one for each GA run) whereas the generated rules reach a minimum level of confidence (previously specified) and give information on areas of search space in which examples which are not described by the rules generated by the previous iterations, remain. The repetition mechanism promotes the generating of different rules (in the sense that they give information on different groups of examples). This is achieved by penalizing –once a rule is obtained– the set of examples represented by the same one in order to generate future rules. It is important to point out that this penalization does not prevent the extraction of overlapped rules because the examples covered by previously obtained fuzzy rules are not eliminated and they take part in the computation of the confidence measure. In subgroup discovery algorithms, the possibility of extracting information on described examples is not eliminated since redundant descriptions of subgroups can show the properties of groups from a different perspective.

```
START
  Best_Rule ← R
  Best_support ← support(R)
  Better ← True
  REPEAT WHILE Better
    Better ← False
    FOR (m=1 to nv)
      R'm ← Best_Rule without considering variable m
      IF (support(R'm)>=support(R) AND
       confidence(R'm)>=confidence(R))
        Better ← True
        IF (support (R'm) > Best_support)
          Best_support ← support (R'm)
          Best_Rule ← R'm
    END FOR
  END WHILE
  IF (confidence(Best_Rule)>=min_conf)
     Return Best_Rule
  ELSE
     Return R
END
```

**Fig. 4** The post-processing phase of the hybrid GA

The confidence of the obtained rule in each iteration must be higher than a previously specified minimum value. In descriptive induction algorithms, one of the fundamental problems, and partially significant to the quality of the obtained results, is the specification of the minimum confidence required for the rules to be extracted. This value depends greatly on the problem to be solved and its solution is a problem which is still not completely resolved. In [38] a method based on fuzzy logic for the setting of the minimum confidence level is described.

## *4.3 Comparison Between the Proposal and Other Subgroup Discovery Algorithms*

To verify the applicability of the proposal, we have compared the results of the model (and of the model with canonical rules, a particular case of DNF rules) with the results of other subgroup discovery algorithms.

For the experimental evaluation and comparison of the approach proposed, the datasets *breast-w* and *diabetes*, both of them containing medical data, and available in the UCI repository have been used. The diabetes dataset contains continuous variables, and is used to show the results of the fuzzy rules extracted by the proposal in comparison with other subgroup discovery algorithms. On the other hand, our proposal can also manage categorical variables, and the breast-w dataset is used to show the behaviour of this proposal with this kind of problems.

The experiments have been carried out in the same way as in [29] to allow the comparison: 10-fold cross validation for the error estimation.

Due to the proposal is a non-deterministic approach, we have carried out 5 runs of each training/test set. The results are the averages of the values obtained by the test partitions. After obtaining the rules with algorithm SDIGA, the measures of Coverage (*Cov*), Support ($Sup_1$), Size, Significance (*Sig*) and Unusualness (*WRAcc*), which are not used in other knowledge extraction processes, were calculated with the expressions indicated in Sect. 2 in order to make the comparison. The parameters used are:

- Population size: 100
- Maximum number of evaluations of individuals in each GA run: 10000
- Mutation probability: 0.01
- Number of linguistic labels for the continuous variables: 3
- Quality measure weights for the fitness function: $w_1$ = 0.4 and $w_2$= 0.3

The specification of the weights for the fitness function depends on the expert knowledge of the characteristics and/or complexity of the problem to be solved. In this chapter, we use these values considering a slight promotion of the extraction of general rules.

Tables 1 and 2 show the results obtained. The tables the results obtained with the two versions of the SDIGA algorithm (SDIGA, using canonical rules [12], and SDIGA-DNF using DNF rules) for 4 minimum confidence values (named "SDIGA CfMin 0.6" for the SDIGA algorithm with a minimum confidence value of 0.6, and so on), the results for the CN2 algorithm modifying the unusualness measure (CN2-WRAcc), and the results of the CN2-SD using different parameters for the weights (CN2-SD ($\gamma = x$) is the CN2-SD algorithm using multiplicative weights with $\gamma = x$, and CN2-SD (add.) is the CN2-SD algorithm using additive weights).

For each measure, the average value and the standard deviation (sd) are detailed. "*COV*" is the average coverage of the set of rules as measured in [2] "*SUP*" is the overall support of a set of rules as computed in [4], "*Siz*" is the number of rules in the induced set of rules, "*SIG*" is the average significance of a set of rules as measured in [6], and "*WRACC*" is the average rule unusualness as computed in [8].

Both models of SDIGA (using canonical and DNF linguistic rules) perform better than the other non fuzzy algorithms for the measures coverage (COV), support (SUP) and size (Siz). This means that our proposal obtains a reduced set of rules with a high percentage of examples covered on average, a high number of examples satisfying both the antecedent and the consequent parts of the rules (i.e., a higher percentage of target positive examples leaving a smaller number of examples unclassified is covered), and with a low number of rules. On the other hand, the results for interest measures show different behaviour in the two problems: significance (SIG) and unusualness (WRACC) of SDIGA are similar to the other algorithms for the breast-w problem, but are worse for the diabetes one.

**Table 1** Comparison of subgroup discovery algorithms for Breast-W dataset

| Algorithm | COV | (sd) | SUP | (sd) | Siz | (sd) | SIG | (sd) | WRACC | (sd) |
|---|---|---|---|---|---|---|---|---|---|---|
| CN2 WRAcc | 0.150 | 0.04 | 0.900 | 0.02 | 8.8 | 0.95 | 13.300 | 1.69 | 0.063 | 0.04 |
| CN2-SD (γ=0.5) | 0.208 | 0.05 | 0.890 | 0.09 | 7.9 | 0.50 | 27.100 | 3.37 | 0.095 | 0.02 |
| CN2-SD (γ=0.7) | 0.174 | 0.04 | 0.840 | 0.04 | 8.5 | 1.75 | 2.100 | 0.02 | 0.079 | 0.01 |
| CN2-SD (γ=0.9) | 0.218 | 0.05 | 0.930 | 0.02 | 9.0 | 0.24 | 20.500 | 2.45 | 0.093 | 0.07 |
| CN2-SD (add.) | 0.260 | 0.04 | 0.860 | 0.05 | 9.2 | 1.24 | 26.600 | 3.43 | 0.111 | 0.04 |
| SDIGA CfMin 0.6 | 0.199 | 0.13 | 0.497 | 0.34 | 5.9 | 3.03 | 6.459 | 2.46 | 0.002 | 0.03 |
| SDIGA CfMin 0.7 | 0.213 | 0.13 | 0.481 | 0.32 | 5.7 | 2.80 | 7.627 | 2.84 | 0.010 | 0.03 |
| SDIGA CfMin 0.8 | 0.238 | 0.19 | 0.439 | 0.30 | 4.0 | 2.06 | 5.782 | 3.08 | 0.006 | 0.03 |
| SDIGA CfMin 0.9 | 0.211 | 0.20 | 0.423 | 0.33 | 3.0 | 1.08 | 6.470 | 3.80 | 0.022 | 0.03 |
| SDIGA-DNF CfMin 0.6 | 0.398 | 0.07 | 0.983 | 0.03 | 5.4 | 0.88 | 16.910 | 3.81 | 0.113 | 0.03 |
| SDIGA-DNF CfMin 0.7 | 0.414 | 0.07 | 0.981 | 0.02 | 5.2 | 0.74 | 17.399 | 4.05 | 0.116 | 0.03 |
| SDIGA-DNF CfMin 0.8 | 0.435 | 0.09 | 0.969 | 0.03 | 4.5 | 1.36 | 18.523 | 5.81 | 0.124 | 0.03 |
| SDIGA-DNF CfMin 0.9 | 0.478 | 0.07 | 0.923 | 0.07 | 2.4 | 0.81 | 24.434 | 6.63 | 0.156 | 0.03 |

Analyzing the results it is observed that the use of different measures in the rule extraction process of CN2-SD with respect to SDIGA implies:

- the increase of the number of rules,
- the decrease of coverage and support, but
- the increase of the interest measurement values

The inclusion of these measures (or adaptation of them to the fuzzy rules) can be considered in the improvement of SDIGA by means of a multiobjective version of it.

The comparison between the results of the two models of linguistic rules extracted by SDIGA shows that the model which uses DNF linguistic rules obtains

**Table 2** Comparison of subgroup discovery algorithms for Diabetes dataset

| Algorithm | COV | (sd) | SUP | (sd) | Siz | (sd) | SIG | (sd) | WRACC | (sd) |
|---|---|---|---|---|---|---|---|---|---|---|
| CN2 WRAcc | 0.275 | 0.04 | 0.820 | 0.03 | 5.2 | 0.79 | 15.800 | 1.07 | 0.065 | 0.06 |
| CN2-SD (γ=0.5) | 0.296 | 0.06 | 0.920 | 0.06 | 6.0 | 0.68 | 14.900 | 1.95 | 0.085 | 0.07 |
| CN2-SD (γ=0.7) | 0.344 | 0.05 | 0.850 | 0.01 | 5.6 | 1.35 | 11.000 | 1.43 | 0.099 | 0.04 |
| CN2-SD (γ=0.9) | 0.299 | 0.05 | 0.950 | 0.01 | 5.4 | 0.30 | 15.200 | 1.85 | 0.086 | 0.07 |
| CN2-SD (add.) | 0.381 | 0.04 | 0.870 | 0.05 | 4.6 | 0.86 | 2.100 | 0.01 | 0.092 | 0.03 |
| SDIGA CfMin 0.6 | 0.462 | 0.06 | 0.939 | 0.04 | 4.3 | 0.68 | 3.286 | 2.25 | 0.028 | 0.02 |
| SDIGA CfMin 0.7 | 0.431 | 0.07 | 0.882 | 0.07 | 3.9 | 0.33 | 3.515 | 2.13 | 0.030 | 0.01 |
| SDIGA CfMin 0.8 | 0.707 | 0.09 | 0.875 | 0.07 | 2.0 | 0.00 | 3.967 | 3.23 | 0.042 | 0.02 |
| SDIGA CfMin 0.9 | 0.707 | 0.09 | 0.875 | 0.07 | 2.0 | 0.00 | 3.967 | 3.23 | 0.042 | 0.02 |
| SDIGA-DNF CfMin 0.6 | 0.849 | 0.09 | 0.992 | 0.01 | 2.8 | 0.38 | 0.788 | 1.01 | 0.024 | 0.01 |
| SDIGA-DNF CfMin 0.7 | 0.854 | 0.09 | 0.992 | 0.01 | 2.9 | 0.35 | 0.633 | 0.54 | 0.023 | 0.01 |
| SDIGA-DNF CfMin 0.8 | 0.931 | 0.04 | 0.978 | 0.02 | 2.0 | 0.00 | 0.437 | 0.34 | 0.024 | 0.01 |
| SDIGA-DNF CfMin 0.9 | 0.935 | 0.03 | 0.976 | 0.02 | 2.0 | 0.00 | 0.418 | 0.29 | 0.023 | 0.01 |

better results than the model which uses canonical linguistic rules. As main conclusions of this short comparison study we can conclude that SDIGA allows us to obtain subgroup discovering linguistic rules:

- with very high values of the measures of coverage and support, and so the linguistic rules can be considered very general and significantly representing the knowledge of the examples of the different values of the target variable;
- highly compact, because both the sizes of the set of rules and also the number of variables involved are small;
- highly descriptive, due to the use of DNF linguistic rules, allowing a representation of the knowledge near to human reasoning, and making the extracted knowledge very actionable, a main objective in any subgroup discovery algorithm;
- with a variable interest measure behaviour.

The use of DNF linguistic rules allows us to describe the extracted knowledge in a more flexible way and moreover, to make changes in the initial granularity in each rule in a descriptive way. In this kind of fuzzy rule, as defined in (9), fuzzy logic contributes to the interpretability of the extracted rules due to the use of a knowledge representation close to the expert, also allowing the use of continuous features without a previous discretization.

## 5 Conclusions

This chapter gives a survey about the use of linguistic rules in the data mining task of subgroup discovery. The subgroup discovery task has been defined, different proposals have been described, and the use of linguistic rules has been analyzed. Then an example of model using linguistic rules for the subgroup discovery task and its advantages has been described.

In summary, for the subgroup discovery task, that searches for unknown and interesting knowledge which can be used for the user, the use of linguistic rules allows the extraction of knowledge in a more natural way and improves it interpretability: Since words play a central role in human information processing, linguistic rules can be used for describe knowledge about subgroups in data which can be actionable by the user.

Finally, we point out some open problems in the development of a fuzzy approach for subgroup discovery:

- To consider the support measure based on /fuzzy set concepts.
- The definition of quality measures for subgroup discovery adapted to the use with linguistic rules.
- The use of multiobjective genetic algorithms [7, 11], analyzing the meaning of pareto-optimal solutions from the subgroup discovery point of view, can provide an interesting tool to get (for getting) set of rules with a trade-off among all the objectives used in the evolutionary model.

**Acknowledgment** This Work was Supported in part by the Spanish Ministry of Education and Science (MEC) under Projects TIN-2005-08386-C05-03 and TIN-2005-08386-C05-01.

## References

1. Agrawal R, Imielinski T, Swami AN (1993) Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of the International Conference on Management of Data (ACM SIGMOD 1995). Washington, DC, pp. 207–216
2. Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo I (1996) Fast Discovery of Association Rules. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in Knowledge Discovery and Data Mining. AAAI Press, California, pp. 307–328
3. Atzmueller M, Puppe F, Buscher H-P (2004) Towards Knowledge-Intensive Subgroup Discovery. In: Proceedings Lernen, Wissensentdeckung und Adaptivität Workshop (LWA'04). Berlin, pp. 117–123
4. Au WH, Chan KCC (1998) An effective algorithm for discovering fuzzy rules in relational databases. In: Proceedings of the IEEE International Conference on Fuzzy Systems (Fuzz IEEE'98). Anchorage (USA), pp. 1314–1319
5. Bäck T, Fogel D, Michalewicz Z (1997) Handbook of Evolutionary Computation. Oxford University Press, Oxford
6. Clark P, Niblett T (1989) The cn2 induction algorithm. Machine Learning 3(4): 261–283
7. Coello CA, Van Veldhuizen DA, Lamont GB (2002) Evolutionary Algorithms for Solving Multi-Objective Problems. Kluwer Academic Publishers, New York
8. Cordón O, del Jesus MJ, Herrera F (1998) Genetic Learning of Fuzzy Rule-based Classification Systems Co-operating with Fuzzy Reasoning Methods. International Journal of Intelligent Systems 13 (10/11): 1025–1053
9. Cordón O, Herrera F, Hoffmann F, Magdalena L (2001) Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases. World Scientific, Singapore
10. Chen G, Wei Q (2002) Fuzzy association rules and the extended mining algorithms. Information Sciences 147: 201–228
11. Deb K (2001) Multi-Objective Optimization using Evolutionary Algorithms. John Wiley & Sons, Chichester
12. Del Jesus MJ, González P, Herrera F, Mesonero M (2005) Evolutionary Induction of Descriptive Rules in a Market Problem. In Ruan D, Chen G, Kerre E, Wets G (eds) Intelligent Data Mining: Techniques and Applications. Springer Verlag, pp. 267–292
13. Del Jesus MJ, González P, Herrera F, Mesonero M (Accepted) Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. IEEE Trans. Fuzzy Systems
14. Dubois D, Prade H, Sudamp T (2005) On the representation, measurement, and discovery of fuzzy associations. IEEE Trans. on Fuzzy Systems 13: 250–262
15. Flach PA, Savnik I (1999) Database dependency discovery: a machine learning approach. AI Communications 12(3): 139–160
16. Fu AW, Wong MH, Sze SC, Wong WC, Wong WL, Yu WK (1998) Finding fuzzy sets for the mining of fuzzy association rules for numerical at-tributes. In: First International Symposium on Intelligent Data Engineering and Learning (IDEAL'98). Hong Kong, pp. 263–268
17. Gamberger D, Lavrac N (2002) Expert-guided subgroup discovery: Methodology and application. Journal of Artificial Intelligence Research 17: 1–27
18. Gamberger D, Lavrac N, Krstacic G (2003) Active subgroup mining: a case study in coronary heart disease risk group detection. Artificial Intelligence in Medicine 28 (1): 27–57
19. Hong TP, Chen CH, Wu YL, Lee YC (2004) Using divide-and-conquer GA strategy in fuzzy data mining. In: Ninth International Symposium on Computers and Communications (ISCC 2004). Alexandria, EGYPT, pp. 116–121
20. Hong TP, Liu KY, Wang SL (2003) Fuzzy data mining for interesting generalized association rules. Fuzzy sets and systems 138: 255–269

21. Hüllermeier E (2005) Fuzzy methods in machine learning and data mining: Status and prospects. Fuzzy Sets and Systems 156 (3): 387–407
22. Ishibuchi H, Nakashima T, Nii M (2004) Classification and modeling with linguistic information granules Springer-Verlag, New York
23. Kavsek B, Lavrac N, Jovanoski V (2003) APRIORI-SD: Adapting association rule learning to subgroup discovery. In: Proceedings of the 5th International Symposium on Intelligent Data Analysis (IDA 2003). Berlin, pp. 230–241
24. Klösgen W (1996) Explora: A Multipattern and Multistrategy Discovery Assistant. In Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in Knowledge Discovery and Data Mining, AAAI Press, California, pp. 249–271
25. Klösgen W (2002) Subgroup Discovery. In Klösgen W, Zytkow J (eds) Handbook of Data Mining and Knowledge Discovery. Oxford University Press, New York, pp. 354–364
26. Klösgen W, May M (2002) Census Data Mining - An Application. In: 13th European Conference on Machine Learning (ECML'02) / 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02) workshop on on Mining Official Data. Helsinki, pp. 65–79
27. Kuok C, Fu A, Wong ML (1998) Mining fuzzy association rules in databases. ACM SIGMOD Record 27: 41–46.
28. Lavrac N, Flach P, Zupan B (1999) Rule evaluation measures: A unifying view. In: Proceedings of the 9th International Workshop on Inductive Logic Programming (ILP'99). Bled, Slovenia, pp. 174–185
29. Lavrac N, Kavsec B, Flach P, Todorovski L (2004) Subgroup discovery with CN2-SD. Journal of Machine Learning Research 5: 153–188
30. Lavrac N, Zelezny F, Flach P (2003) RSD: Relational subgroup discovery through first-order feature construction. In: Proceedings of the 13th International Conference on Inductive Logic Programming (ILP 2003). Szeged, Hungary, pp. 149–165
31. Michie D, Spiegelhalter DJ, Taylor CC (1994) Machine learning, neural and estatistical classification. Ellis Horwood
32. Piatetsky-Shapiro G, Matheus, C (1994) The interestingness of deviation. In: Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases. Seattle, Washington, pp. 25–36
33. Quinlan JR (1987) Generating Production Rules from Decision Trees. In: Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI'87). Milan, Italy, pp. 304–307
34. Raedt LD, Dehaspe L (1997) Clausal discovery. Machine Learning 26: 99–146
35. Wrobel S (1997) An algorithm for multi-relational discovery of subgroups. In Proceedings of the First European Symposion on Principles of Data Mining and Knowledge Discovery (PKDD-97). Trondheim, Norway, pp. 78–87
36. Zadeh LA (1965) Fuzzy sets. Information Control 8: 338–353
37. Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning, parts I, II, III. Information Sciences 8–9: 199–249, 301–357, 43–80
38. Zhang S, Lu J, Zhang C (2004) A fuzzy logic based method to acquire user threshold of minimum-support for mining association rules. Information Sciences 164: 1–16

# Fuzzy Prototypes: From a Cognitive View to a Machine Learning Principle

**Marie-Jeanne Lesot, Maria Rifqi and Bernadette Bouchon-Meunier**

**Abstract** Cognitive psychology works have shown that the cognitive representation of categories is based on a typicality notion: all objects of a category do not have the same representativeness, some are more characteristic or more typical than others, and better exemplify their category. Categories are then defined in terms of prototypes, i.e. in terms of their most typical elements. Furthermore, these works showed that an object is all the more typical of its category as it shares many features with the other members of the category and few features with the members of other categories.

In this paper, we propose to profit from these principles in a machine learning framework: a formalization of the previous cognitive notions is presented, leading to a prototype building method that makes it possible to characterize data sets taking into account both common and discriminative features. Algorithms exploiting these prototypes to perform tasks such as classification or clustering are then presented.

The formalization is based on the computation of *typicality degrees* that measure the representativeness of each data point. These typicality degrees are then exploited to define *fuzzy prototypes*: in adequacy with human-like description of categories, we consider a prototype as an intrinsically imprecise notion. The fuzzy logic framework makes it possible to model sets with unsharp boundaries or vague and approximate concepts, and appears most appropriate to model prototypes.

We then exploit the computed typicality degrees and the built fuzzy prototypes to perform machine learning tasks such as classification and clustering. We present several algorithms, justifying in each case the chosen parameters. We illustrate the results obtained on several data sets corresponding both to crisp and fuzzy data.

## 1 Introduction

Prototypes are elements representing categories, structuring and summarizing them, underlining their most important characteristics and their specificity as opposed to other categories. From a cognitive point of view, they are the basis for the categorization task, process that aims at considering as equivalent objects that are distinct but similar: cognitive science works [31, 32] showed that natural categories are organized around the notion of prototype and the related notion of typicality.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

In this paper, we propose to transpose the cognitive view of prototypes to a machine learning principle to extract knowledge from data. More precisely, we propose to characterize data sets through the construction of prototypes that realize the cognitive approach. The core of our motivation is the fact that, through this notion of prototype, a data subset (a category for instance) is characterized both from an internal and an external point of view: the prototype underlines both what is common to the subset members and what is specific to them in opposition to the other data. Using these two complementary components leads to context-dependent representatives that are more relevant than classic representatives that actually exploit only the internal view. Furthermore, the method makes it possible to determine the extent to which the prototype should be a central or a discriminative element, i.e. it allows to rule the relative importance of common and discriminative features, leading to a flexible prototype building method.

Another concern of our approach is the adequacy with human-like descriptions that are usually based on imprecise linguistic expressions. To that aim, the construction method we propose builds **fuzzy prototypes**: the fuzzy logic framework makes it possible to model sets with unsharp boundaries or vague and approximate concepts, as occur in this framework.

The paper is organized as follows: in Sect. 2, the cognitive definitions of typicality and prototypes are introduced. In Sect. 3, these principles are formalized to a prototype building method that makes it possible to characterize numerical data sets. These prototypes are then exploited both for supervised and unsupervised learning: Sect. 4 presents prototype-based classification methods and Sect. 5 describes a typicality-based clustering algorithm.

## 2 Cognitive Definition of Prototype

The cognitive definition of prototype was first proposed in the 70's [27] and popularized by E. Rosch [31, 32], in the context of the study of cognitive concept organization. Previously, a crisp relationship between objects and categories was assumed, based on the existence of necessary and sufficient properties to determine membership: according to this model, an object belongs to a category if it possesses the properties, interpreted as necessary and sufficient conditions; otherwise, it is not a member of the category. Now in the case of natural categories, it is often the case that no feature is common to all the category members: as modeled in the family resemblance model of Wittgenstein [36], each object shares different common features with other members of the category, but no globally shared feature can be identified.

The prototype view of concept organization [27, 31, 32] models categories as organized around a center, the prototype, that is described by means of properties that are characteristic, typical of the category members. Indeed, all objects in a category are not equivalent: some are better examples and more characteristic of the category than others. For instance, in the case of the mammal category, the dog is considered as a better example than a platypus. Thus, objects are spread over a scale, or a gradient of typicality; the prototype is then related to the individuals that maximize this gradient.

Rosch and her colleagues [31, 32] studied this typicality notion and showed it depends on two complementary components: an object is all the more typical of its category as it shares many features with the other members of the category and few features with the members of other categories. This can for instance be illustrated by platypuses and whales in the case of mammals: platypuses are atypical mainly because they have too few features in common with other mammals, whereas whales are atypical mainly because they have too many common features with members of the fish category. Due to this typicality definition, prototypes underline both the common features of the category members and their discriminative features as opposed to other categories: they characterize the category both internally and in opposition to other categories.

## 3 Realization of the Prototype View

### *3.1 Principle*

To construct a fuzzy prototype in agreement with the previous cognitive prototype view, we consider that the degree of typicality of an object depends positively on its total resemblance to others objects of its class (internal resemblance) and on its total dissimilarity to objects of other classes (external dissimilarity). This makes it possible to consider both the common features of the category members, and their distinctive features as opposed to other categories. More precisely, the fuzzy prototype construction principle consists in three steps [29]:

Step 1 Compute the *internal resemblance* degree of an object with the other members of its category and its *external dissimilarity* degree with the mem- bers of the outside categories.

Step 2 Aggregate the internal resemblance and the external dissimilarity degrees to obtain the *typicality degree* of the considered object.

Step 3 Aggregate the objects that are typical "enough", i.e. with a typicality degree higher than a predefined threshold to obtain the fuzzy prototype.

#### Internal Resemblance and External Dissimilarity

Step 1, that is illustrated on Fig. 1, requires the choice of a resemblance measure and a dissimilarity measure to compare the objects. These measures depend on the data nature and are detailed in Sects. 3.3 and 3.4 in the case of fuzzy and crisp data. Formally, denoting them $r$ and $d$ respectively, and denoting $x$ an object belonging

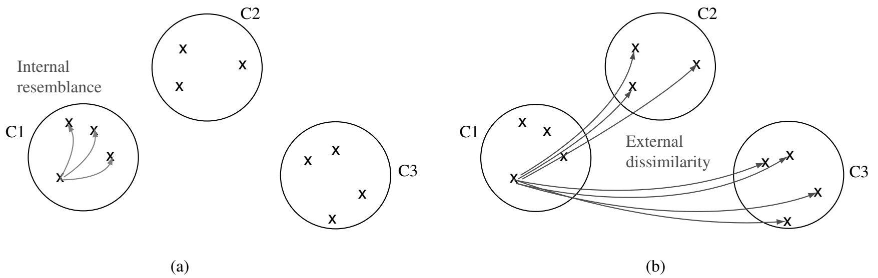



**Fig. 1** (**a**) Computation of the internal resemblance, as the resemblance to the other members of the category, (**b**) computation of the external dissimilarity, as the dissimilarity to members of the other categories

to a category $C$, $x$ internal resemblance with respect to $C$, $R(x, C)$, and its external dissimilarity, $D(x, C)$, are computed as

$$R(x, C) = avg(r(x, y), y \in C) \qquad D(x, C) = avg(d(x, y), y \notin C) \qquad (1)$$

i.e. as the average resemblance to other members of the category and the average dissimilarity to members of other categories. The average operator can be replaced with other operators [29].

Figure 2 illustrates these definitions in the case of the iris data base, using only one attribute (petal length) for visualization sake: the histograms represent the data distribution, $*$, $+$ and $\circ$ respectively depict the three classes; the y-axis shows for each point its internal resemblance and external dissimilarity. As expected, it can be seen that the points maximizing internal resemblance are, for each class, the central points, underlining the common features of the category members. On the contrary, the points maximizing the external dissimilarity are extreme points, at least

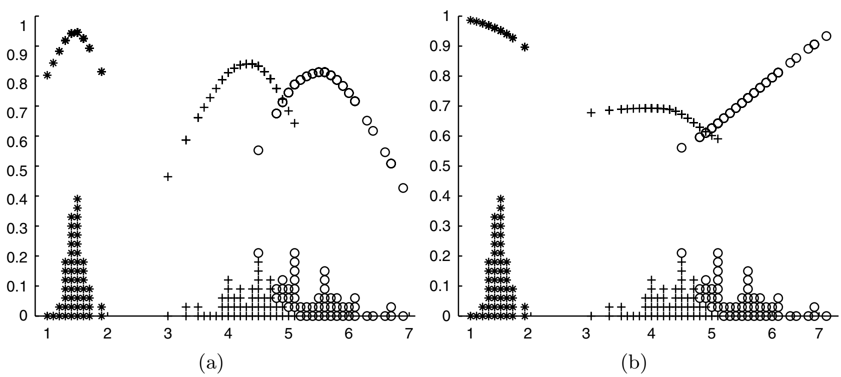



**Fig. 2** (**a**) Internal resemblance, (**b**) external dissimilarity, for the iris data set, considering only the petal length attribute [23]

for the two extreme classes: points in the middle class (+ class) get low external dissimilarity values, as they are too close to the other groups and correspond to an average behavior. Thus external dissimilarity underlines the specificity of the classes, for instance indicating that high petal length is characteristic for the ∘ class: it highlights the discriminative features of each category (or the absence of any, for the + class), and makes it possible to build caricatures of the classes.

Therefore combining both information to a typicality degree makes it possible to build representatives that simultaneously underline the common features of the category members, as well as their discriminative features as opposed to other categories.

### Aggregation to a Typicality Degree

Step 2 requires the choice of an aggregation operator to express the dependence of typicality on internal resemblance and external dissimilarity, that is formally written

$$T(x, C) = \varphi(R(x, C), D(x, C)) \tag{2}$$

where $\varphi$ denotes the aggregation operator. It makes it possible to rule the semantics of typicality and thus that of the prototypes, determining the extent to which the prototype should be a central or a discriminative element [23].

Figure 3 illustrates the typicality degrees obtained from the internal resemblance and external dissimilarity of Fig. 2 for four operators: the minimum (see Fig. 3a) is a conjunctive operator that requires both $R$ and $D$ to be high for a point to be typical, leading to rather small typicality degrees on average. On the contrary, for the maximum (see Fig. 3b), as any disjunctive operator, if either R or D is high, a point is considered as typical. This leads to higher values but to non-convex distributions, reflecting a double semantics for typicality: central points as well as extreme points are typical, but for different reasons.

Trade-off operators, such as the weighted mean (Fig. 3c), offer a compensation property: low R values can be compensated for by high D. This is illustrated by the leftmost point on Fig. 3c, whose typicality is higher than with the min operator, because its external dissimilarity compensates for its internal resemblance. The weights used in the weighted mean determine the extent to which compensation can take place, and rule the relative importance of internal resemblance and external dissimilarity, leading to more or less discriminative prototypes, underlying more the common or the distinctive features of the categories.

Lastly, variable behavior operators, such as the MICA operator [17] (see Fig. 3d) or the symmetric sum [33], are conjunctive, disjunctive or trade-off operators, depending on the values to be aggregated. They offer a reinforcement property: if both $R$ and $D$ are high, they reinforce each other to give an even higher typicality (see class $*$), if both are low, they penalize each other to give an even smaller value (see the leftmost point of the ∘ class).

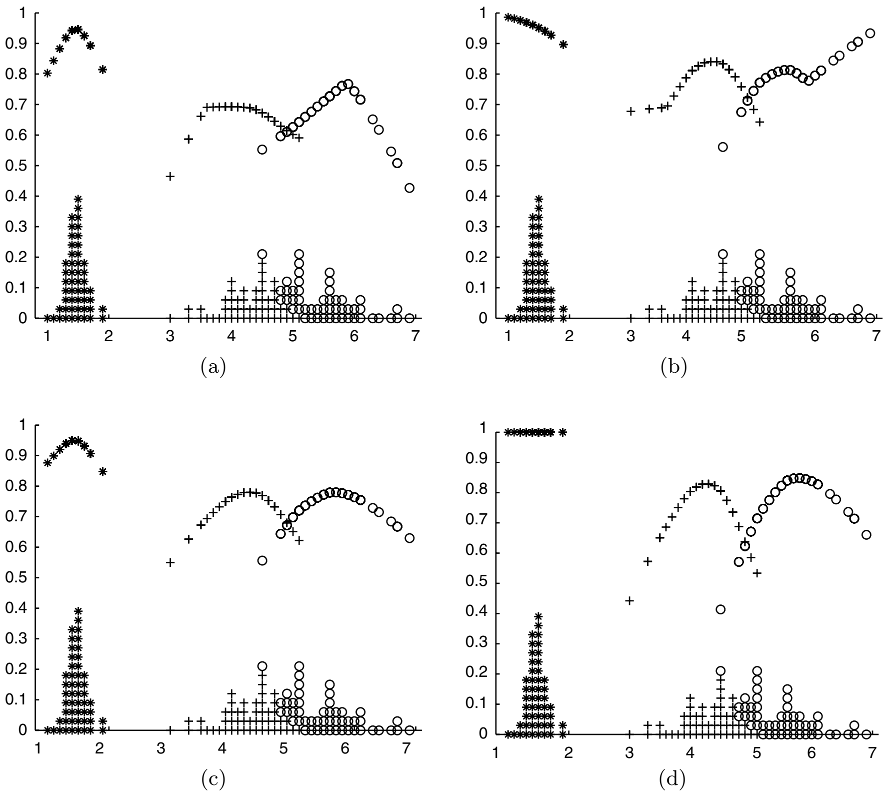



**Fig. 3** Typicality degrees obtained from the internal resemblance and the external dissimilarity shown on Fig. 2 using different aggregation operators: (**a**) $T = min(R, D)$, (**b**) $T = max(R, D)$, (**c**) $T = 0.6R + 0.4D$, (**d**) $T =$ MICA(R,D) [17]

Therefore, the aggregation operator determines the semantics of typicality, and rules the relative influence played by internal resemblance and external dissimilarity.

## Aggregation to a Prototype

Lastly step 3 builds the prototype of a category itself, as the aggregation of the most typical objects of the category: the prototype for category $C$ is defined in a general form as

$$p(C) = \psi(\{x/T(x, C) > \tau\}) \tag{3}$$

where $\tau$ is the typicality threshold and $\psi$ the aggregation operator. The latter can still depend on the typicality degrees associated to the selected points, taking for instance the form of a weighted mean. Its actual definition depends on the data nature, it is detailed in Sects. 3.3 and 3.4 in the case of fuzzy and crisp data.

**Remarks**

It is to be underlined that prototype can be built either from the typicality degrees of objects, as presented above, but also in an attribute per attribute approach: the global approach makes it possible to take into account attribute correlation, but it can also be interesting to enhance the typical values of the different properties used to describe the objects, without considering an object as an indivisible whole. This means, all the values of a property are considered simultaneously, without taking into account the objects they describe. For instance, if the objects are represented by means of a color, a size and a weight, the fuzzy prototype of the category of these objects is the union of typical values for the color, the size and the weight. It means that instead of computing internal resemblance, external dissimilarity and typicality degrees for each object, these quantities are computed for each attribute value. This approach is in particular applied in the case of fuzzy data, as described in Sect. 3.3.

## *3.2 Related Works*

There exist many works to summarize data sets, or build significant representatives. A first approach, taking the average as starting point, consists in defining more sophisticated variants to overcome the average's drawbacks, in particular its sensitivity to outliers (see for instance [11]). Among these variants, one can mention the Most Typical Value [11] or the representatives proposed in [26] or [37]. In all these cases, the obtained value is computed as a weighted mean, the difference comes from the ways the weights are defined.

Besides, a second approach, more concerned with interpretability, does not reduce the representative to a single precise value but builds so-called linguistic summaries [16, 15]. The latter identify significant trends in the data and represent them in the form "Q B's are A" where Q is a linguistic quantifier and A and B are fuzzy sets; for example, "most important experts are young".

Yet, both approaches build internal representatives that take into account the common characteristics of the data, but not their specificity. More precisely, they focus only on the data to be summarized, and do not depend on their context, which prevents them from identifying their particularity. On the contrary, the cognitive view highlights the discriminative features: the data are also characterized in opposition to the other categories.

Furthermore, the proposed realization of the cognitive approach described in Sect. 3.1 makes it possible to rule the relative importance common and discriminative features play and the trade-off between them through the choice of the aggregation operator, leading to a flexible method. As illustrated in Sects. 4 and 5, the choice of the aggregation operator depends on the considered use of the prototype.

## *3.3 Fuzzy Data Case*

In this section, we consider the case of fuzzy data, i.e. data whose attributes take as values fuzzy sets: formally, denoting $\mathcal{F}(\mathbb{R})$ the set of fuzzy subsets defined on the real line, the input space is $\mathcal{F}(\mathbb{R})^p$ where $p$ denotes the number of attributes. We describe the instantiation of the previous prototype construction method, discussing comparison measures for such data and aggregation operators to construct the prototypes from the most typical values. We then present an application in a medical domain, to mammographies.

### Comparison Measures

In the case of fuzzy data, the framework used to compute the internal resemblance as well as the external dissimilarity is the one defined in [4] generalizing the Tversky's "contrast model" [35].

In this framework a measure of resemblance comparing two fuzzy sets $A$ and $B$ is a function of three arguments: $M(A \cap B)$ (the common features), $M(A - B)$ and $M(B - A)$ (the distinctive features), where $M$ is fuzzy set measure [9] like the fuzzy cardinality for instance. More formally [4]:

**Definition 1** *A* measure of resemblance *r is*

- *non decreasing in* $M(A \cap B)$*, non increasing in* $M(A - B)$ *and* $M(B - A)$
- *reflexive:* $\forall A,\ r(A, A) = 1$
- *symmetrical:* $\forall A, B,\ r(A, B) = r(B, A)$

An example of measure of resemblance, proposed by [8], generalizing the Jacccard measure to fuzzy sets, is the following:

$$\dot{r}(A, B) = M(A \cap B)/M(A \cup B)$$

for$M$ such that : $M(A \cup B) = M(A \cap B) + M(A - B) + M(B - A)$.

We also refer to this framework to choose a dissmilarity measure:

**Definition 2** *A* measure of dissimilarity *d is:*

- *independent of* $M(A \cap B)$ *and non decreasing in* $M(A - B)$ *and* $M(B - A)$
- *minimal:* $\forall A,\ d(A, A) = 0$

An example of measure of dissimilarity, based on the generalized Minkowski's distance, is the following:

$$d(A, B) = \left( \frac{1}{Z} \left( \int |f_A(x) - f_B(x)|^n dx \right) \right)^{1/n}$$

where $Z$ is a normalizing factor, $n$ an integer and $f_X$ denotes the membership function of the fuzzy set $X$.

It is to be noticed that a dissimilarity measure is not necessarily deduced from a resemblance measure and vice versa because the purpose is to have different information when comparing two objects, the dissimilarity measure focusing on distinctive features.

### Aggregation into a Fuzzy Prototype

In the last step of the fuzzy prototype construction, fuzzy values of an attribute that are typical "enough" have to be aggregated in order to obtain a typical value for the considered attribute. An aggregation operator must be chosen among numerous existing operators, deeply studied by several authors like Mizumoto [24], [25], Detyniecki [7] or Calvo and her colleagues [5].

### Application to Mammography

Nowadays, mammography is the primary diagnostic procedure for the early detection of breast cancer. Microcalcification[1] clusters are an important element in the detection of breast cancer. This kind of finding is the direct expression of pathologies which may be benign or malignant. The description of microcalcifications is not an easy task, even for an expert. If some of them are easy to detect and to identify, some others are more ambiguous. The texture of the image, the small size of objects to be detected (less than one millimeter), the various aspects they have, the radiological noise, are parameters which impact the detection and the characterization tasks.

More generally, mammographic images present two kinds of ambiguity: *imprecision* and *uncertainty*. The *imprecision* on the contour of an object comes from the fuzzy aspect of the borders: the expert can define approximately the contour but certainly not with a high spatial precision. The *uncertainty* comes from the microcalcification superimpositions: because objects are built from the superimpositions of several 3D structures on a single image, we may have a doubt about the contour position.

The first step consists in finding automatically the contours of microcalcifications. This segmentation is also realized thanks to a fuzzy representation of imprecision and uncertainty (more details can be found in [28]). Each microcalcification is then described by means of 5 fuzzy attributes computed from its fuzzy contour. These attributes enable us to describe more precisely:

- the shape (3 attributes): elongation (minimal diameter/maximal diameter), compactness1, compactness2.
- the dimension (2 attributes): surface, perimeter.

[1] The microcalcifications are small depositions of radiologically very opaque materials which can be seen on mammography exams as small bright spots.

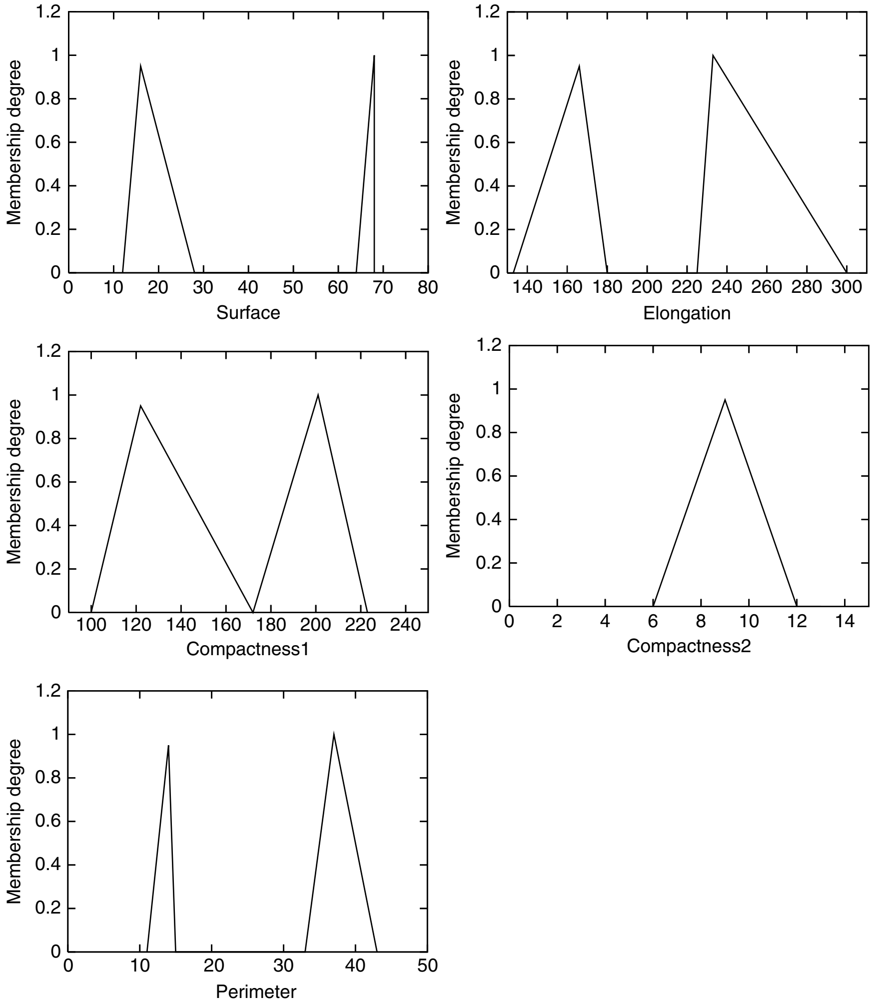



**Fig. 4** Description of a microcalcification by means of fuzzy values of its 5 attributes

Figure 4 shows an example of the membership functions of the values taken by a detected microcalcification. One can notice that the membership functions are not "standard" in the sense that they are not triangular or trapezoidal (as it is often the case in the literature) and this is because of the automatic generation of fuzzy values (we will not go into details here, interested readers may refer to [3]).

Experts have categorized microcalcifications into 2 classes: *round* microcalcifications and *not round* ones, because this property is important to qualify the malignancy of the microcalcifications. The aim is then to build the fuzzy prototypes of the classes *round* and *not round*. Figure 5 gives the obtained fuzzy prototypes with the internal resemblance and external dissimilarity computed using as aggregator the median (replacing the average in (1) by the median). The typicality degrees are

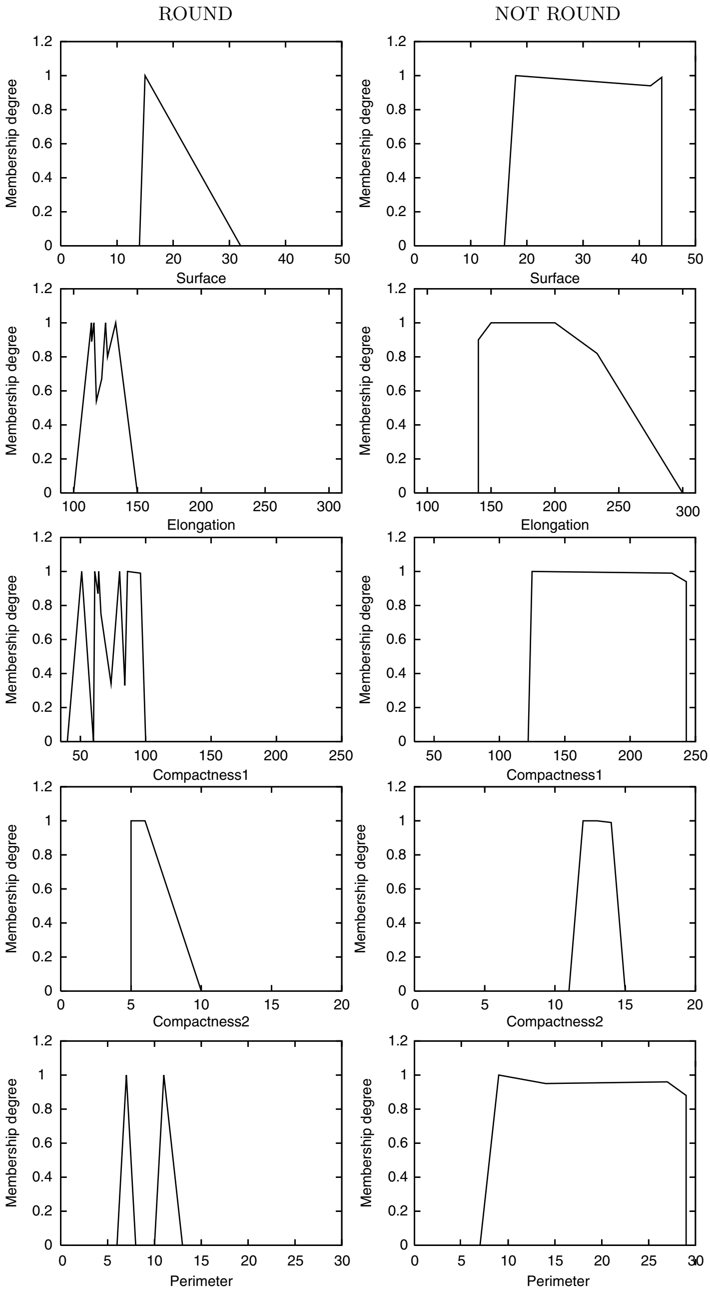



**Fig. 5** Prototypes of the classes *round* and *not round*

obtained by the probabilistic t-conorm ($\varphi(x, y) = x + y - x \cdot y$ in (2)). Lastly, the fuzzy values with maximal typicality degree are aggregated through the max-union operator to define the fuzzy prototype value of the corresponding attribute.

It can be seen that on the attributes *elongation*, *compactness1* or *compactness2*, the typical values of the two classes *round* and *not round*, are quite different: the intersection between them is low. This can be interpreted in the following way: *a round microcalcification typically has an elongation approximately between 100 and 150 whereas a not round microcalcification typically has an elongation approximately between 150 and 200*, etc. For the attributes *surface* and *perimeter*, contrary to the previous attributes, the typical values of the two classes are superimposed, it means that these attributes are not typical.

## 3.4 Crisp Data Case

In this section, we consider the case of crisp numerical data, i.e. data represented as real vectors: the input space is $\mathbb{R}^p$, where $p$ denotes the number of attributes. We describe the comparison measures that can be considered in this case, and the aggregation operator that builds prototypes from the most typical data; lastly we illustrate the results obtained on a real data set.

### Comparison Measures

Contrary to the previous fuzzy data, for crisp data, the relative position of two data cannot be characterized by their common and distinct elements respectively represented as their intersection and their set differences: the information is reduced, and depends on a single quantity, expressed as a distance or a scalar product of the two vectors.

Thus dissimilarity is simply defined as a normalized distance [19]

$$d(x, y) = \max\left(\min\left(\frac{\delta(x, y) - d_m}{d_M - d_m}, 1\right), 0\right) \tag{4}$$

where $\delta$ is a distance, for instance chosen among the Euclidean, the weighted Euclidean, the Mahalanobis or the Manhattan distances, depending on the desired properties (e.g. robustness, derivability). The parameters $d_m$ and $d_M$ are the normalization parameters, that can for instance be chosen as $d_m = 0$ and $d_M$ the maximum observed distance. More generally, $d_m$ can be interpreted as a tolerance threshold, indicating the distance below which the dissimilarity is considered as 0, i.e. no distinction is made between the two data points; $d_M$ corresponds to a saturation threshold, indicating the distance from which the two points are considered as totally dissimilar.

Regarding resemblance measures, two definitions can be considered. They can first be deduced from scalar products as the latter are related to the angle between the two vectors to be compared and are maximal when the two vectors are identical; they must be normalized too to define a resemblance measure. Besides, similarity can be defined as a decreasing function of dissimilarity, as for instance

$$r(x, y) = 1 - d(x, y) \qquad or \qquad r(x, y) = \frac{1}{1 + d(x, y)^{\gamma}} \tag{5}$$

where $\gamma$ is a user-defined parameter. Indeed, if dissimilarity is total, resemblance is 0 and reciprocally. The Cauchy function on the right expresses a nonlinear dependency between resemblance and dissimilarity, and in particular makes it possible to rule the discrimination power of the measure [19, 30].

It is to be noticed that, as in the case of fuzzy data, the resemblance measure is not necessarily the complement to 1 of the dissimilarity measure: the normalization parameters defining dissimilarity can be different from those used to define the dissimilarity from which the resemblance is deduced. Indeed, resemblance is used to compare data points belonging to the same category, whereas dissimilarity compares points from different categories. Thus it is expected that they apply to different distance scales, requiring different normalization schemes.

### Aggregation into a Fuzzy Prototype

After the typicality degrees have been computed, the prototype is defined as the aggregation of the most typical data. In the case of numerical data, one can for instance define the prototype as the weighted average of the data, using the typicality degrees as weights. Yet this reduces the prototype to a single precise value, which is not in adequacy with human-like description of categories: considering for instance data describing the height of persons from different countries, one would rather say that "the typical French person is around 1.70m tall" instead of "the typical French person is 1.6834m tall" (fictitious value): the prototype is not described with a single numerical value, but with a linguistic expression, "around 1.70m", which is imprecise. This is better modeled in the fuzzy set framework, that makes it possible to represent such unclear boundaries.

Therefore we propose to aggregate the most typical data in a fuzzy set [23]. To that aim, two thresholds are defined, respectively indicating the minimum typicality degree required to belong to the prototype kernel and its support; in-between, a linear interpolation is performed. In our experiments, the thresholds are set to high values (respectively 0.9 and 0.7), because the prototype aim is to characterize the data set, and extract its most representative components, and not to describe it as a whole.

### Application to Student Characterization

As an example, we consider a data set describing results obtained by 150 students to two exams [23]. It was decomposed into 5 categories by the fuzzy $c$-means [2, 10]: the central group corresponds to students having average results for both exams, the 4 peripheral clusters correspond to the 4 combinations success/failure for the

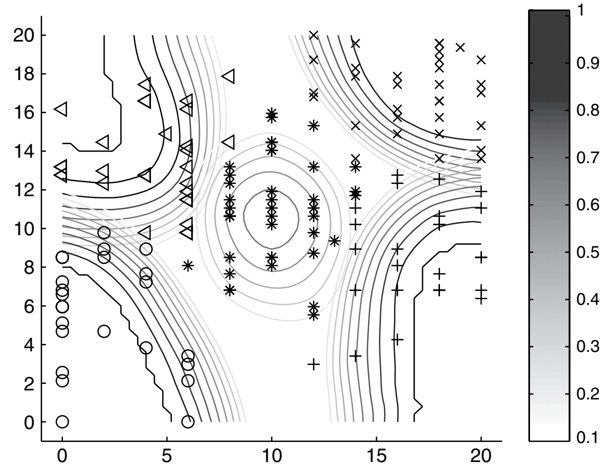



**Fig. 6** Level lines of the 5 fuzzy prototypes characterizing students described by their results on two exams [23]

two exams. Prototypes are built to characterize these 5 categories. The comparison measures are based on normalized Euclidean distances (with different normalization factors for the dissimilarity and the resemblance). The typicality degrees are derived from internal resemblance and external dissimilarity using the symmetric sum operator [33]. Lastly the prototype is derived from the most typical data points using the method described in the previous paragraph.

Figure 6 represents the level lines of the obtained prototypes and shows they provide much richer information than a single numerical value: they model the unclear boundaries of the prototypes. They also underline the difference between the central group and the peripheral ones: no point totally belongs to the prototype of the central group, because it has no real specificity as opposed to the other categories, as it corresponds to an average behavior.

In all cases, the fuzzy prototypes characterize the subgroups, capturing their semantics: they are approximately centered around the group means but take into account the discriminative features of the clusters and underline their specificity. For instance, for the lower left cluster, the student having twice the mark 0 totally belongs to the prototype, which corresponds to the group interpretation as students having failed at both exams and underline its specificity. It is to be noticed that this results from the chosen aggregation operator (symmetric sum [33]) that gives a high influence for external dissimilarity: it can be the case that such an extreme case should have a lower membership degree, which can be obtained changing the aggregation operator.

## 4 Extension to Supervised Learning: Classification

In this section, we exploit the previous formalization of the typicality degrees as well as the fuzzy prototype notion to perform a supervised learning task of classification. It is true that the major interest of a prototype comes from its power of description thanks to its synthetic view of the database. But, as Zadeh underlined [38], a fuzzy prototype can be seen as a schema for generating a set of objects. Thus, in a classification task, when a new object has to be classified, it can be compared to each fuzzy prototype and classified in the class of the nearest prototype (a sort of nearest neighbor algorithm where the considered neighbors are only the prototypes of the classes). Another approach consists in taking into account the degrees of typicality without considering the fuzzy prototypes, it means that the last step of our construction processed is missed. The difference of our approach with an instance-based learning algorithm [1] is that our methods are not lazy: the information learned during the typicality degree computation is taken into account in the classification task, more precisely in the class setting step, either by considering the nearest prototypes or by weighting the comparison by the typicality degrees.

We proposed three classification methods based on typicality or on prototype notions:

- The first one is the one described above giving the class of the nearest prototype of the object to be classified. The prototype is constructed with the fuzzy value maximizing the typicality degree.
- The second one is like the first one, but the fuzzy prototype is obtained aggregating by the union (the maximum) of the values with a high typicality degree whereas the first one considers only one value.
- In the third one, a new object is compared to each object of the learning database. The comparison is the aggregation of the attribute by attribute comparisons weighted by the degree of typicality of the attribute value of the object in the learning database. Then, the class given to the unknown object is the class of the most similar object in the learning database. It is also possible to consider the $k$ most similar objects but the realized experiments consider only the closest object relatively to the weighted similarity.

We tested these 3 methods on the microcalcifications database presented in Sect. 3.3 in 3 different classification problems: to classify the microcalcifications in round/not round, elongated/not elongated and small/not small. Table 1 gives the highest good classification rates obtained by each method and compares them with instance-based algorithm (IBL) with 10 neighbors. It shows that our method classifies better than IBL, highlighting the gain provided by the typicality-based approaches.

**Table 1** Classification results for the 3 typicality-based methods and for the instance-based learning algorithm (IBL) (in percentage of good classification)

| | Method 1 | Method 2 | Method 3 | IBL |
|---|---|---|---|---|
| round/not round | 75.00 | 79.63 | **82.41** | 79.63 |
| elongated/not elongated | 79.41 | **80.88** | 79.41 | 73.53 |
| mall/not small | 92.45 | **93.71** | 91.82 | 91.82 |

## 5 Extension to Unsupervised Learning: Clustering

In this section, we further exploit the notion of prototype as a machine learning principle, considering the unsupervised learning case and more precisely the clustering task.

### *5.1 Motivation*

Clustering [14] aims at decomposing a data set into subgroups, or clusters, that are both homogeneous and distinct: the fact that the subgroups are homogeneous (their compactness) implies that points assigned to the same cluster indeed resemble one another, which justifies their grouping. The fact that they are distinct (their separability) implies that points assigned to different subgroups are dissimilar one from another, which justifies their non-grouping and the individual existence of each cluster. Thus the cluster decomposition provides a simplified representation of the data set, that summarizes it and highlights its underlying structure.

Now these compactness and separability properties can be matched with the properties on which typicality degrees rely, namely internal resemblance and external dissimilarity: a cluster is compact if all its members resemble one another, which is equivalent to their having a high internal resemblance. Likewise, clusters are separable, if all their members are dissimilar from other clusters members, i.e. if they have a high external dissimilarity. Thus a cluster decomposition has a high quality if all points have a high typicality degree for the cluster they are assigned to.

This is illustrated using the artificial two-dimensional data set shown on Fig. 7a. Figures 7b and 7d present two data decompositions into 2 subgroups, respectively depicted with $+$ and $\triangleleft$. Figures 7c and 7e show their associated typicality degree distribution: for each point, represented by its identification number as indicated on Fig. 7a, its typicality degrees for the two clusters are indicated, the plain line corresponding to the $+$ cluster, the dashed one to the $\triangleleft$ cluster. It can be seen that typicality degrees take significantly higher values for the data partition of Fig. 7d than for the partition of Fig. 7b that is counter-intuitive and does not correspond to the expected decomposition: the most satisfying decomposition is the one for which each point is more typical of the cluster it belongs to.

Thus we propose to exploit the typicality degree framework to perform clustering: the typicality-based clustering algorithm (TBC) [20] looks for a decomposition such that each point is most typical of the cluster it is assigned to, and aims at maximizing the typicality degrees.

### *5.2 Typicality-based Clustering Algorithm*

Following the motivations described above, the typicality-based clustering algorithm consists in alternating two steps:

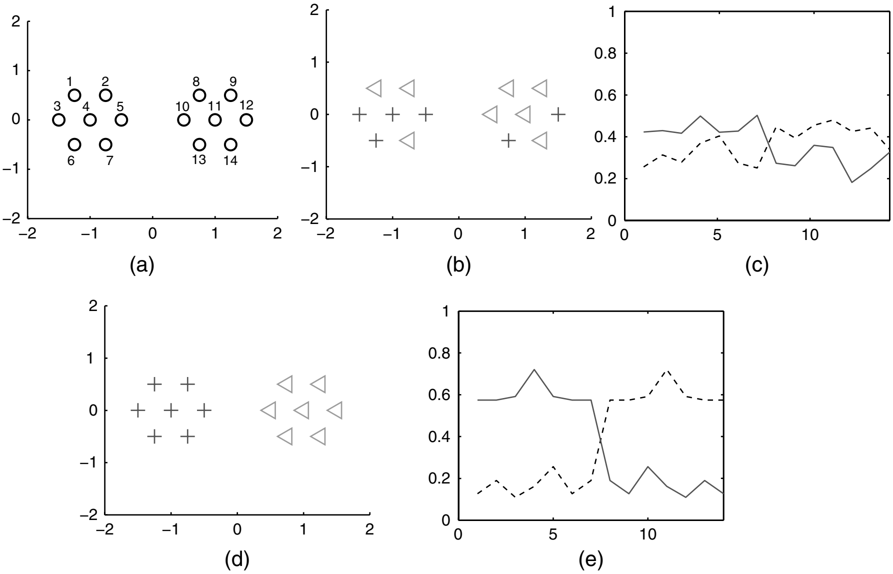



**Fig. 7** Motivation of the typicality-based clustering algorithm. (**a**) Considered data set and data point numbering. (**b**) Counter-intuitive decomposition into 2 clusters, respectively depicted with $+$ and $\triangleleft$, (**c**) associated typicality degrees with respect to the 2 clusters, for all data points represented by their identification number; the plain line indicates the typicality degree with respect to the $+$ cluster, the dashed one for the $\triangleleft$ cluster. (**d**) Decomposition into the 2 expected clusters and (**e**) associated typicality degrees

1. Assuming a data partition, compute typicality degrees with respect to the clusters,
2. Assuming typicality degrees, modify the data partition so that each point becomes more typical of the cluster it is assigned to.

This means, one reduces to the supervised case considering the candidate categories provided by the data decomposition, and one then evaluates these candidates using the computed typicality degrees. According to empirical tests, this alternated process converges very rapidly to a stable partition, that corresponds to the desired partition [20].

Among the expected advantages of this approach are robustness to outliers and ability to avoid cluster overlapping areas: both outliers and points located in overlapping areas can be identified easily, as they have low typicality degrees (respectively because of low internal resemblance and low external dissimilarity), leading to clusters that are indeed compact and separable. Moreover, after the algorithm has converged, the final typicality degree distribution can be exploited to build fuzzy prototypes characterizing the obtained clusters, offering an interpretable representation of the clusters.

Regarding the typicality computation step, two differences are to be underlined as compared to the supervised case described in Sect. 3. First, in the supervised learning framework, typicality is only considered for the category a point belongs to, and equals 0 for the other categories. In the clustering case, clusters are to be identified, and different assignments must be considered, thus typicality degrees are computed for all points and all clusters. The candidate partition is only used to determine which points must be taken into account for the computation of internal resemblance and external dissimilarity: for instance, for a given point, when its typicality with respect to cluster $C$ is computed, internal resemblance is based on points assigned to $C$ according to the candidate partition.

A second difference between supervised and unsupervised learning regards the choice of the aggregation operator defining typicality degrees from internal resemblance and external dissimilarity: it cannot be chosen as freely in both cases. Indeed, in the supervised case, on may be interested in discriminative prototypes (e.g. in classification tasks), thus a high importance may be given to external dissimilarity. In clustering, both internal resemblance and external dissimilarity must be influential, otherwise, outliers may be considered as highly typical of any cluster, distorting the clustering results. This means the aggregation operator must be a conjunctive operator or a variable behavior operator [20].

## *5.3 Experimental Results*

In order to illustrate the properties of the typicality-based clustering algorithm (TBC), its results on the artificial two-dimensional data set shown on Fig. 8a are presented. The latter is made of 3 Gaussian clusters and a small outlying group in the upper left corner.

### Typicality-based Clustering Algorithm

Figures 8b and 8c represent the level lines of the obtained typicality degree distribution and the associated fuzzy prototypes, when 3 clusters are searched for. Each symbol depicts a different cluster, the stars represent points assigned to the fictitious cluster that represents outliers and points in overlapping areas (more precisely, this cluster groups points with low typicality degrees [20]). Figure 8b shows that the expected clusters are identified, as well as the outliers. These results show the method indeed takes into account both internal resemblance and external dissimilarity.

The effect of these two components can also be seen in the typicality distribution: on the one hand, the distributions are approximately centered around the group center, due to the internal resemblance constraint. On the other hand, the distribution of the upper cluster for instance is more spread on the x-axis than on the y-axis: the overlap with the two other clusters leads to reduced typicality degrees, due to the external dissimilarity constraint.

The associated fuzzy prototypes shown on Fig. 8c provide relevant summaries of the clusters: they have small support, and characterize the clusters. Indeed, they are concentrated on the central part of the clusters, but also underline their distinctive

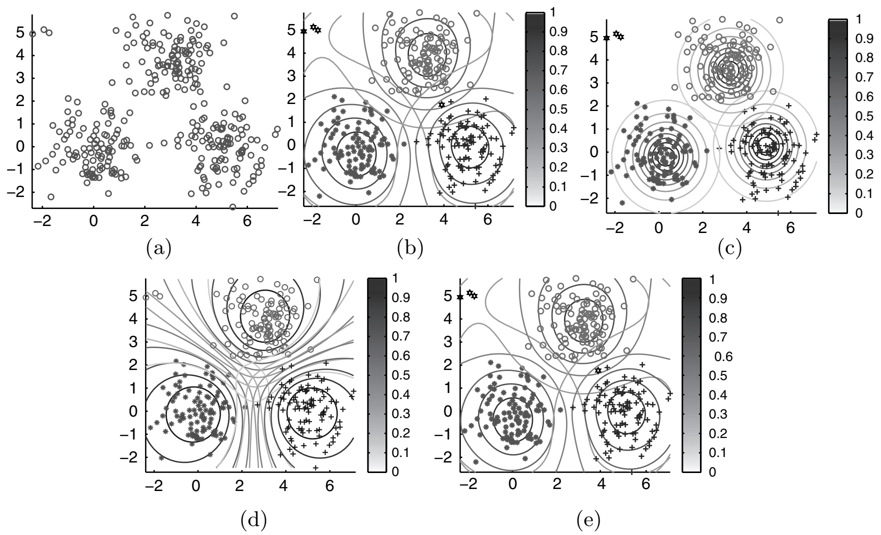


**Fig. 8** (**a**) Considered data set, (b-e) Level lines of several distributions: (**b**) typicality degrees, (**c**) fuzzy prototypes, (**d**) FCM membership degrees, (**e**) PCM possibilistic coefficients

features, i.e. the particularities of each cluster as compared to the others: for the rightmost cluster for instance, the prototype is more spread in the bottom right region, indicating such values are specific for this cluster, which constitutes a relevant characterization as opposed to the 2 other clusters.

#### Comparison with Fuzzy $c$-means and Possibilistic $c$-means

So as to compare the results with classic clustering algorithms, Figs. 8d and 8e respectively show the level lines of the fuzzy sets built by the fuzzy $c$-means (FCM) algorithm [2, 10] and the distribution of coefficients obtained with the possibilistic $c$-means (PCM) [18], both with $c = 3$.

Outliers have a bigger influence for FCM than for TBC, and tend to attract all three clusters in the upper left direction. This sensitivity of FCM is well-known and can be corrected using variants such as the noise clustering algorithm [6]. Apart from the outliers assigned to the upper cluster, the FCM partition is identical to that of TBC. The main difference between TBC and FCM concerns the obtained fuzzy sets: the FCM ones are much more spread and less specific than the typicality or the prototype distributions, they cover the whole input space. Indeed, FCM do not aim at characterizing the clusters, but at describing them as a whole, representing all data points: FCM fuzzy sets aim at modeling ambiguous assignments, i.e. the fact that a point can belong to several groups simultaneously. The associated distribution thus corresponds to membership degrees that indicate the extent to which a point belongs to each cluster. Fuzzy prototypes only represent the most typical points, they correspond to fuzzy sets that characterize each cluster, indicating the extent to which a point belongs to the representative description of the cluster.

The PCM coefficients (see Fig. 8e) correspond to a third semantics. It can be observed that their distributions are spherical for all 3 clusters: the underlying functions are indeed decreasing functions of the distance to the cluster center. This implies they are to be interpreted as internal resemblances, not taking into account any external dissimilarity component. Thus, contrary to FCM, PCM are not sensitive to outliers, the latter are indeed identified as such and not assigned to any of the three clusters. Yet due this weight definition PCM suffer from convergence problems: they sometime fail to detect the expected clusters and identify several times the same cluster [13]. To avoid this effect, Timm and Kruse [34] introduce in the PCM cost function a cluster repulsion term, so as to force clusters apart. The proposed typicality-based approach can be seen as another solution to this problem: the external dissimilarity component also leads to a cluster repulsion effect. The latter is incorporated in the coefficient definition itself and not only in the cluster center expression, enriching the coefficient semantics.

### *5.4 Algorithm Extensions*

The previous typicality-based clustering mechanism can be extended to adapt to specific data or cluster constraints, we briefly mention here two extensions.

TBC does not depend on the data points themselves, but only on their comparison through resemblance and dissimilarity measures: contrary to FCM or PCM, it does not require the computation of data means and in the course of the optimization process, clusters are only represented by the set of their members, and not cluster centers. This implies that the algorithm is independent of the data nature, and makes it possible to extend it to other distances [21]: on the one hand, non-Euclidean distances can be used, in particular to identify non-convex clusters; on the other hand, non-vectorial data, such as sequences, trees or graphs can be handled.

Another extension concerns the use of typicality degrees in a Gustafson-Kessel manner: the Gustafson-Kessel algorithm [12] is a variant of FCM that makes it possible to identify ellipsoidal clusters, whereas FCM restrict to spherical clusters, through the automatic extraction of the cluster covariance matrices: contrary to the previous methods, the appropriate distance function is not determined at the beginning of the algorithm, but is automatically learned from the data. The Gustafson-Kessel-like typicality-based clustering algorithm [22] modifies the optimization scheme presented in Sect. 5.2 to estimate cluster centers and cluster covariance matrices, using as weights the typicality degrees. As TBC, it is robust with respect to outliers and able to avoid overlapping areas between clusters, leading to both compact and separable clusters [22].

## 6 Conclusion

In this paper, we considered the cognitive definition of prototype and typicality and proposed to extend them to machine learning and data mining tasks. First, these notions make it possible to characterize data categories, underlining both the common features of the category members and their discriminative features, i.e. the category specificity, leading to an interpretable data summarization; they can be applied both to crisp and fuzzy data. Furthermore, these notions can be extended to extract knowledge from the data, both in supervised and unsupervised learning frameworks, to perform classification and clustering.

Perspectives of this work include the extension of the typicality notion to other machine learning tasks, and in particular to feature selection: when applied attribute by attribute, typicality makes it possible to identify properties that have no typical values for categories, and are thus not relevant for the category description. This approach has the advantage of defining local feature selection, insofar as attribute relevance is not defined for the whole data set, but locally for each category (or each cluster in unsupervised learning). Links with other feature selection methods, and particular entropy-based approaches are to be studied in details.

## References

1. D.W. Aha, D. Kibler, and M.K. Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.
2. J. Bezdek. *Fuzzy mathematics in pattern classification*. PhD thesis, Applied Mathematical Center, Cornell University, 1973.
3. S. Bothorel, B. Bouchon-Meunier, and S. Muller. Fuzzy logic-based approach for semiological analysis of microcalcification in mammographic images. *International Journ. of Intelligent Systems*, 12:814–843, 1997.
4. B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems*, 84(2):143–153, 1996.
5. T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In *Aggregation operators: new trends and applications*, pp. 3–104. Physica-Verlag, 2002.
6. R. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.
7. M. Detyniecki. *Mathematical Aggregation Operators and their Application to Video Querying*. PhD thesis, University of Pierre and Marie Curie, 2000.
8. D. Dubois and H. Prade. *Fuzzy Sets and Systems, Theory and Applications*. Academic Press, New York, 1980.
9. D. Dubois and H. Prade. A unifying view of comparison indices in a fuzzy set-theoretic framework. In R. R. Yager, editor, *Fuzzy set and possibility theory*, pp. 3–13. Pergamon Press, 1982.
10. J.C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journ. of Cybernetics*, 3:32–57, 1973.
11. M. Friedman, M. Ming, and A. Kandel. On the theory of typicality. *Int. Journ. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 3(2):127–142, 1995.
12. E. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *Proc. of IEEE CDC*, pp. 761–766, 1979.
13. F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis, Methods for classification, data analysis and image recognition*. Wiley, 2000.
14. A. Jain, M. Murty, and P. Flynn. Data clustering: a review. *ACM Computing survey*, 31(3): 264–323, 1999.
15. J. Kacprzyk and R. Yager. Linguistic summaries of data using fuzzy logic. *Int. Journ. of General Systems*, 30:133–154, 2001.
16. J. Kacprzyk, R. Yager, and S. Zadrozny. A fuzzy logic based approach to linguistic summaries of databases. *Int. Journ. of Applied Mathematics and Computer Science*, 10:813–834, 2000.

17. A. Kelman and R. Yager. On the application of a class of MICA operators. *Int. Journ. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 7:113–126, 1995.
18. R. Krishnapuram and J. Keller. A possibilistic approach to clustering. *IEEE Transactions on fuzzy systems*, 1:98–110, 1993.
19. M.-J. Lesot. Similarity, typicality and fuzzy prototypes for numerical data. In *6th European Congress on Systems Science, Workshop "Similarity and resemblance"*, 2005.
20. M.-J. Lesot. Typicality-based clustering. *Int. Journ. of Information Technology and Intelligent Computing*, 2006.
21. M.-J. Lesot and R. Kruse. Data summarisation by typicality-based clustering for vectorial data and nonvectorial data. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, Fuzz-IEEE'06*, pp. 3011–3018, 2006.
22. M.-J. Lesot and R. Kruse. Gustafson-Kessel-like clustering algorithm based on typicality degrees. In *Proc. of IPMU'06*, 2006.
23. M.-J. Lesot, L. Mouillet, and B. Bouchon-Meunier. Fuzzy prototypes based on typicality degrees. In B. Reusch, editor, *Proc. of the 8th Fuzzy Days 2004*, Advances in Soft Computing, pp. 125–138. Springer, 2005.
24. M. Mizumoto. Pictorial representations of fuzzy connectives, part I: Cases of $t$-norms, $t$-conorms and averaging operators. *Fuzzy Sets and Systems*, 31(2):217–242, 1989.
25. M. Mizumoto. Pictorial representations of fuzzy connectives, part II: Cases of compensatory operators and self-dual operators. *Fuzzy Sets and Systems*, 32(1):45–79, 1989.
26. N. Pal, K. Pal, and J. Bezdek. A mixed c-means clustering model. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, Fuzz-IEEE'97*, pp. 11–21, 1997.
27. M.I. Posner and S.W. Keele. On the genesis of abstract ideas. *Journ. of Experimental Psychology*, 77:353–363, 1968.
28. A. Rick, S. Bothorel, B. Bouchon-Meunier, S. Muller, and M. Rifqi. Fuzzy techniques in mammographic image processing. In Etienne Kerre and Mike Nachtegael, editors, *Fuzzy Techniques in Image Processing*, Studies in Fuzziness and Soft Computing, pp. 308–336. Springer Verlag, 2000.
29. M. Rifqi. Constructing prototypes from large databases. In *Proc. of IPMU'96*, pp. 301–306, 1996.
30. M. Rifqi, V. Berger, and B. Bouchon-Meunier. Discrimination power of measures of comparison. *Fuzzy sets and systems*, 110:189–196, 2000.
31. E. Rosch. Principles of categorization. In E. Rosch and B. Lloyd, editors, *Cognition and categorization*, pp. 27–48. Lawrence Erlbaum associates, 1978.
32. E. Rosch and C. Mervis. Family resemblance: studies of the internal structure of categories. *Cognitive psychology*, 7:573–605, 1975.
33. W. Silvert. Symmetric summation: a class of operations on fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 9:659–667, 1979.
34. H. Timm and R. Kruse. A modification to improve possibilistic fuzzy cluster analysis. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, Fuzz-IEEE'02*, 2002.
35. A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.
36. L. Wittgenstein. *Philosophical investigations*. Macmillan, 1953.
37. R. Yager. A human directed approach for data summarization. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, Fuzz-IEEE'06*, pp. 3604–3609, 2006.
38. L. A. Zadeh. A note on prototype theory and fuzzy sets. *Cognition*, 12:291–297, 1982.

# Improving Fuzzy Classification by Means of a Segmentation Algorithm

**Ana del Amo, Daniel Gómez, Javier Montero and Greg S. Biging**

**Abstract** In this chapter we consider remotely sensed images, where land surface should be classified depending on their uses. On one hand, we discuss the advantages of the fuzzy classification model proposed by Amo et al. (European Journal of Operational Research, 2004) versus standard approaches. On the other hand, we introduce a coloring algorithm by to Gómez et al. (Omega, to appear) in order to produce a supervised algorithm that takes into account a previous segmentation of the image that pursues the identification of possible homogeneous regions. This algorithm is applied to a real image, showing its high improvement in accuracy, which is then measured.

## 1 Introduction

Classification is a key stage towards knowledge, which used to require in the first stage an approximate identification of the entities we are facing (linguistic terms may represent useful tools). In a second stage we should be searching for relations between those entities. But knowledge should not be confused with decision making. We would all like to be ready to face our decision making problems, that is, to have in our mind a sufficient awareness of reality [28]. Different rules and requirements play in knowledge and decision making, and in fact Medicine has proven that these two activities are developed in different parts of the human brain: decision making is deeply related to emotion (as is creativity, which has also been proven as being physically separated from methodological thinking). A complete analytic capability does not imply being able to choose one particular alternative.

In this way, we can accept that a certain previous classification process is always needed in order to establish the terms in which each decision making problem is stated. In this sense, the way classification is understood and therefore modelled will be extremely relevant in decision making.

For example, our classification procedures are understood in terms of classical Aristotelic logic, for example, the entities we consider should always be precisely defined, and decision making problems will be stated in terms of those precise entities. Hence, if we accept that each human act can be either *good* or *bad*, and nothing else, the only question we can ask ourselves about our own possible acts is

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

to choose if such an act will be *good* or *bad*: on one hand, we are forced to accept that at least one *good* alternative is possible for every decision making (otherwise we may fall into desperation); and on the other hand, either we accept it is *good* to choose between several *good* alternatives, let's say at random or *by heart* (falling into a unethical behavior), or we are forced to think that such a *good* alternative is unique (falling into an extremely stressful search).

In a more technical framework, assuming that every entity that must be precisely defined may produce wrong results when precise definitions simply do not exist (we end up imposing precise but arbitrary definitions). When analyzing Earth land cover, for example, most of us may consider that beginning with three classes such as *natural space*, *urban area* and *wetland* could be a starting point. But it is obvious that a *natural space* vary between deep forest and desert, and frontiers between those classes will frequently be arbitrary, with extensive gradation zones.

Principal learning processes refer to the ability to manage more and more complex classification structures, which imply evaluating results and the possibility of modifying our classification structure, including the search for new classes if needed. In any case, it must be pointed out that such an evaluation implies a detailed analysis of results taking into account the whole classification structure under consideration.

In this paper we shall present a short overview of **aggregative fuzzy classification systems** in the sense of [1, 3, 5]), which allows evolution by avoiding Ruspini's [36] static approach (Sect. 2). Such a model is then considered for the analysis of digitalized images, by means of an unsupervised classification model, which is improved introducing a previous segmentation of the image according to possible homogeneous regions (Sect. 3). Such a model is applied to a particular remote sensing image (Sect. 4), followed by a final comments section.

## 2 Aggregative Fuzzy Classification Systems

Let us assume a discrete set of classes $\mathcal{C}$, each class being well defined, and a finite family of objects $X$ to be classified. From a classical point of view, the objective should be to assign, one and only one, of those classes to each of the objects we consider. The problems appear when the objects we present for classification do not fall into any of those classes under consideration. In complex classification problems (see [40]) we quite often observe that objects belong to several considered classes, but only to some extent. Following [39] (see also [10]) we can for example determine the degree $\mu_c(x) \in [0, 1]$ to which every object $x \in X$ belongs to class $c \in \mathcal{C}$, in such a way that a family of membership functions

$$\{\mu_c : X \to [0, 1]\}_{c \in \mathcal{C}}$$

are being defined.

It must be pointed out that, in practice, degrees of membership are assigned taking into account all available possibilities for classification. This is clear in the classical crisp case, where $\mathcal{C}$ defines a standard partition in such a way that it is required that $\mu_c(x) = 1$ for a unique class $c \in \mathcal{C}$ and $\mu_c(x) = 0$ otherwise. Most people will never assign an object to a particular class without having a look at the remaining classes (we choose the most appropriate class among the classes we are being offered). Ruspini [36] translates this approach into the fuzzy context, introducing fuzzy partitions (see also [11, 12, 18, 25]) in such a way that $\mu_c(x) \in [0, 1]$ for every object under consideration and

$$\sum_{c \in \mathcal{C}} \mu_c(x) = 1, \qquad \forall x \in X$$

holds. In this way it is suggested that the total degree of membership is distributed among the classes.

But as already pointed out in [3], great difficulties appear when Ruspini's proposal is implemented in practice: we should not expect that our set of classes $\mathcal{C}$ is capturing the whole problem, and in addition, in a compact way. In practice, classes used to overlap, and it may even be the case that they are almost the same, or an object has nothing to do with those considered classes, so an additional class should be searched. In most cases a fuzzy partition has not been defined, we do not know if it exists or if it is even a desired objective. It is only after a long learning (and lucky) process that users are able to get a family of fuzzy classes fully explaining every object and without superfluous information. In practice we need a first elementary family of classes, to be improved depending on its behavior with respect to considered objects.

Such a fuzziness is often present in classification problems of remotely sensed images, since most classes in nature do not present clear borders (see, e.g., [2, 4, 19, 27, 34]). In particular, we want to classify pixels in a digital image of a landscape that have been obtained from a satellite, each pixel representing several kilometers or meters (depending on the precision). A classical crisp approach will assign to each pixel the class with higher likelihood, and a standard Ruspini approach will distribute membership but forcing them to sum up to one, almost certain to be artificial. Moreover, classification of each pixel should take into account the behavior of surrounding pixels.

Note that two main characteristics have appeared in the previous arguments: the set of classes should explain each object as clearly as possible, but at the same time classes should avoid excessive overlapping. In addition, each class should of course be relevant (an empty class, for example, is useless). Hence, a classification system requires a certain logical structure behind it, in order to evaluate the degree to which each object is explained (by means of an *union* aggregation operator of partial information from subsets of classes), the degree to which classes do not overlap (by means of an *intersection* aggregation operator from subsets of classes), and the degree to which a subset of classes is relevant (by means of a *negation* operator). Hence, our classification structure should at least allow evaluation of the classification produced by $\mathcal{C}$, evaluation that should be the basis for the proposal of a new family of classes improving previous classification.

Of course, general considerations about aggregation rules apply here (see, e.g., [14], but also [21]), although it is important to point out that an aggregation rule should not be identified with an associative binary operator. For example, in [6, 17] an alternative model for building up operational aggregation rules was developed, based on a recursive binary calculus (see also [9]). Although this last approach was justified by a calculating argument, it should be noticed that such a model assumed a given structure of data, which in general may be non linear. In fact, such a structure defines a particular planar graph within the remote sensing framework. This discussion (see [31]) has much in common with Atanassov's intuitioinistic fuzzy sets [7, 8], and also with the necessary simultaneous presence of different aggregation rules in basic structures pointed out in [29, 30] (objects and classes may deserve different aggregation operators and logic).

In particular, a **recursive rule** $\phi$ was defined in [17] as a family of aggregation functions

$$\{\phi_n : [0, 1]^n \to [0, 1]\}_{n>1}$$

such that there exists an ordering rule $\pi$ and two sequences of binary operators

$$\{L_n : [0, 1]^2 \to [0, 1]\}_{n>1}$$

and

$$\{R_n : [0, 1]^2 \to [0, 1]\}_{n>1}$$

such that

$$\phi_n(\pi(x_1), \ldots, \pi(x_n)) = L_n(\phi_{n-1}(\pi(x_1), \ldots, \pi(x_{n-1})), \pi(x_n))$$

and

$$\phi_n(\pi(x_1), \ldots, \pi(x_n)) = R_n(\pi(x_1), \phi_{n-1}(\pi(x_2), \ldots, \pi(x_n)))$$

In this way recursiveness appears as a property of a sequence of operators $\{\phi_n\}_{n>2}$ allowing the aggregation of any number of items **one by one**, by means of a sequence of binary aggregation operators, but not necessarily always the same binary aggregation operator (see [6] for a characterization of quasi-linear recursive aggregation rules, based on a previous result given in [37]). In a more general context, this approach can be generalized in order to assure a sequential calculation, being consistent with the natural **contiguity** structure in the bi-dimensional picture, which is not a linear structure (the value to be associated with a certain pixel should be the aggregation of surrounding pixels, depending on the distance to that pixel).

Hence, an **aggregative fuzzy classification system** was defined in [3] as a family of fuzzy classes $\mathcal{C}$ (each $c \in \mathcal{C}$ with its associated membership function $\mu_c : X \to [0, 1]$), together with an **evaluation triplet** $(\phi, \varphi, N)$ where $\phi$ is a **union** aggregation, $\varphi$ is an **union** operator and $N$ is a **strong negation** [38]. Such an **aggregative fuzzy classification system** can be therefore denoted by

$$(\mathcal{C}; \phi, \varphi, N)$$

where $\varphi$ is a **disjunctive rule** in the sense that $\varphi_n(x_1, x_2, \ldots, x_n) = 1$ whenever $\exists j/x_j = 1$, and

$$\varphi(x_1, x_2, \ldots, x_n) = N(\phi(N(x_1), N(x_2), \ldots, N(x_n)))$$

for all $x_1, x_2, \ldots, x_n \in [0, 1]$.

There are two basic properties we would like in an aggregative fuzzy classification system, and a third is also desirable in order to produce **compact** models (and note that a complementary statistical analysis is suggested in every item):

1. On one hand, every object should be explained to some extent, so the higher all values,

$$\mu_{\mathcal{C}}(x) = \varphi\{\mu_c(x)/c \in \mathcal{C}\}$$

are, the better.
2. On the other hand, common information between classes should not be excessive, and in addition to standard statistical arguments to establish which classes are significatively different, we can point out that, for all $\mathcal{A}, \mathcal{B} \subset \mathcal{C}$, the lower all values

$$\phi\{\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)\}$$

are, the better.
3. And information should not be excessive either excessive, so whenever we suppress $\mathcal{B} \subset \mathcal{C}$ from our classification system, the lower all values

$$\varphi\{\mu_c(x), c \in \mathcal{B}\}$$

are, the better.

These three basic properties suggest the following formulation for the three key properties suggested above (note the great similitude of this approach to [13]).

## *2.1 Covering*

In case

$$\mu_{\mathcal{C}}(x) = \varphi\{\mu_c(x)/c \in \mathcal{C}\} = 1$$

it can be understood that object $x \in X$ is fully explained by our family of fuzzy classes $\mathcal{C}$.

Hence, the degree to which a family of objects $X$ is *covered* by a certain family of classes $\mathcal{C}$ can be associated with the value

$$\mu_X(\mathcal{C}) = min_{x \in X}\big(\varphi\{\mu_c(x)/c \in \mathcal{C}\}\big)$$

A low value of the above degree of covering will suggest a search for a new class not yet included in $\mathcal{C}$ (see [35]).

## 2.2 Redundancy

We should not only pursue an explanation for every object, but a *compact* family of classes avoiding overlapping. This issue implies of course certain statistical analysis to avoid useless information, like those **non discriminative** classes ($\mu_c(x) = a_c \in [0, 1], \forall x \in X$) or those replicated classes ($\mu_c(x) = \mu_k(x), \forall x \in X$), or any other close family not showing significantly different behavior between objects.

According to the arguments above we can propose

$$\mu_X(\mathcal{A}, \mathcal{B}) = max_{x \in X}\big(\phi\{\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)\}\big)$$

as a measure for the degree of redundancy between two families of classes $\mathcal{A}, \mathcal{B} \subset \mathcal{C}$.

In case

$$\mu_{\mathcal{A},\mathcal{B}}(x) = \phi\{\mu_c(x)/c \in \mathcal{C}\} = 0$$

it can be understood that those two families of classes are *orthogonal* with respect to object $x \in X$.

A high degree of redundancy may suggest a search for a new set of classes via recombination of classes in $\mathcal{C}$.

## 2.3 Relevancy

Redundancy between single classes may be enough in many practical problems, at least at an initial stage, and we may then proceed similarly to certain statistical procedures for the reduction of the amount of information. But such a search for a reduction of information can be also suggested from our model, whenever the information contained in a particular class or a subset of classes is not significantly relevant.

According to the above arguments, relevancy of a given subset of classes $\mathcal{A} \subset \mathcal{C}$ being a candidate for being deleted from our model will require a compared analysis between the following items:

(1) $\varphi\{\mu_c(x)/\, c \in \mathcal{C}\}$
(2) $\varphi\{\mu_k(x)/\, k \in \mathcal{A}\}$
(3) $\varphi\{\mu_k(x)/\, k \in \mathcal{C} - \mathcal{A}\}$

for every object $x \in X$ (see [1] for details).

It must be pointed out that a class may not show significance according to some standard statistical procedures, but that information may still be the only hint we have as to classify some pixels. And of course a sequential analysis of isolated irrelevant classes can suggest a wrong suppression of a relevant set of classes (a relevant family of classes may contain no relevant class).

### *2.4 Fuzzy Partitions*

From the above point of view, a fuzzy partition will appear when $\mu_X(\mathcal{C}) = 1$ and $\mu_X(\mathcal{A}, \mathcal{B}) = 0$ for all disjointed $\mathcal{A}, \mathcal{B} \subset \mathcal{C}$, i.e., $\mathcal{A} \cap \mathcal{B} = \emptyset$.

But in no way are we forced to work with these fuzzy partitions. The set of classes can evolve, producing a sequence of **aggregative fuzzy classification systems** being improved by experience, perhaps (but not necessarily) pursuing a fuzzy partition. Although meaningless classification systems will not be useful as a decision aid tool, any classification allowing some kind of discrimination between objects represents some knowledge.

## 3 Remote Sensing Classification

The key issue pointed out in the previous section is the following: we do not need, and we should not impose, that our classification system defines a fuzzy partition. We can still manage, and we should manage, classification systems far from being **good**. In fact, at the beginning of any problem we may assume a trivial **non informative** classification system, given by a single family of classes $\mathcal{C}_\emptyset = \{\emptyset\}$ such that $\mu_\emptyset(x) = 0, \forall x \in X$ (only one class telling us nothing). Then our first objective will be to find out a class telling us something about objects, but we can still find another trivial family containing only non discriminative classes, so in fact we should be looking for a set of **different** classes, none of them being trivial. This process can be continued until we get a satisfactory classification system, most likely approaching, but not defining, a fuzzy partition.

Hence, we can address classification problems without relevant model restrictions, at least at an initial stage. But it is also extremely important to keep clear in our mind the final objective we are pursuing, which can not be different from knowledge support. As already pointed above, we should separate decision making from the necessary rational analysis, most likely the only one subject to consistency restrictions. Our objective is therefore restricted to a decision making aid tool: helping decision making to understand the image under study (see [20]).

We shall now consider remotely sensed images, where land surfaces should be classified depending on their uses, and then discuss the advantages of the previous fuzzy classification model within a non-supervised framework. In accordance with this approach we shall produce at an initial stage a family of fuzzy classes, showing gradation zones and without imposing unrealistic assumptions. But since this approach managed with difficulty by decision makers, we introduce information about surrounding pixels, by means of a segmentation procedure that looks for regions with homogeneous behavior. In particular we consider the coloring algorithm given by to Gómez et al. [23]. Such a segmentation algorithm is based on a coloring process for valued and fuzzy graphs (see [32]). From this segmentation we gather useful information for training sites and pattern recognition steps, to be considered as complementary information. In this way we shall produce a supervised algorithm that can produce a simplified crisp classification, but based exclusively on data. Furthermore, the model allows a posterior learning process for the post-classification process (see, e.g., [26, 33] for some additional interesting complementary tools).

But we should first remind ourselves that although classification is the most important topic in remote sensing, numerous pre-processing procedures must always be previously applied, given the amazing complexity of the Earth's surface (see, e.g., [15, 16]):

- Sensor determination: the choice of the sensor must be made taking into account the objects or classes of the study, in order to discriminate classes. Fuzzy sets theory appears in a natural way when the preferences and aims of the decision maker are modelled.
- Management data and transformation: errors and fuzziness are often present in data acquisition processes. Furthermore, sometimes a reduction of the amount of information is needed when the image is extremely complex (for example more that 100 spectral bands).
- Training site and pattern recognition: in order to know the main features associated with each class, a previously unsupervised classification or expert classification is needed. Here the segmentation and non-supervised classification techniques (hard and soft) are the most common. In the following section a segmentation algorithm is presented as an alternative to solve this step.
- Supervised classification algorithm: this step ends with a classification (crisp or fuzzy) of the image. Each pixel or unit sample is classified into crisp or fuzzy classes, taking into account the information received from the training site.
- Post classification: in order to smooth the classification and improve the classification accuracy, some learning process will be needed. Some logical rules should be considered in order to improve and smooth results.
- Analysis results: once classification and post classification are finished, the accuracy of the process and classifiers must be determined. For example, by considering different agreement measures between the reference data set and the final classification. Recent researches give the fuzzy sets an important role in this analysis.

It will be shown below, in a particular image, the high improvement in accuracy (direct information about training sites can of course be considered at a third stage, but it will usually imply an additional relevant cost).

## 3.1 Segmentation as a Preprocessing Step

As already pointed out, before classifying pixels into known classes, complex data pre-processing steps are required. Once the pre-processed image is obtained, the standard immediate objective is the identification of homogeneous regions. These regions will allow the decision maker to identify training sites for further classification.

The most common techniques for determining homogeneous regions are based on statistical methods that only take into account the spectral information of each isolated pixel. In this way clusters are obtained, but since they are obtained taking into account only spectral information of each pixel, it may happen that the proposed cluster is not related to the real classes in which the decision maker is interested. Consequently, there is a need to include contextual information within those algorithms. With this aim, in this section a coloring algorithm is presented that takes into account the neighborhood of each pixel. In our opinion this information should be complementary with the classical information given by statistical methods.

The remote sensing image will be modelled here as a fuzzy graph where the nodes are crisp and the edges-links are fuzzy (see [23]). Mathematically, a remote sensing image with $k$ bands can be initially defined as a set

$$P = \left\{ p_{i,j} \;/\; 1 \le i \le r \;\; 1 \le j \le s \right\}$$

of $r \times s$ information units -pixels-, where

$$p_{i,j} = \left( p^1_{i,j}, p^2_{i,j}, \cdots, p^k_{i,j} \right)$$

is the pixel associated with the coordinate $(i, j)$.

In order to find homogeneous regions in the image we model this image by a planar graph (see again [23]). The graph is planar in the sense that two pixels $p_{i,j}$ and $p_{i',j'}$ are not connected if $\left|i - i'\right| + \left|j - j'\right| > 1$ (see Fig. 1).

A fuzzy graph $\tilde{G} = (P, \tilde{A})$ is then defined by the image of pixels $P$ and the set of fuzzy arcs $\tilde{A}$, being characterized by a matrix

$$\mu_{\tilde{A}} = \left( d_{p_{i,j}, p_{i',j'}} \right)_{p_{i,j}, p_{i',j'} \in \bar{P}} \tag{1}$$

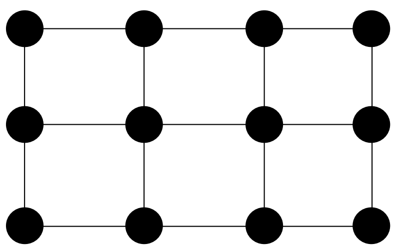


**Fig. 1** Planar graph with r = 3 and s = 4

where $\bar{P}$ is the set of connected pixels, i.e.,

$$\bar{P} = \left\{(p_{i,j}, p_{i',j'}) \in P^2 : \left|i - i'\right| + \left|j - j'\right| = 1; 1 \leq i \leq r, 1 \leq j \leq s\right\}$$

and the fuzzy number $d_{p_{i,j}, p_{i',j'}}$ represents the degree of dissimilarity between the adjacent pixels $p_{i,j}$ and $p_{i',j'}$. Some possible dissimilarity functions are given in [24].

The coloring algorithm for valued-fuzzy graphs we propose (see [23, 24]) is also a succession of binary coloring procedures. The first binary coloring analyzes the pixels set $P$ coloring each pixel as “0” or “1”. The second binary coloring is applied separately to the sub graph generated by those pixels colored as “0”, to obtain the color classes “00” and “01”, that can be understood as homogenous regions, and to the sub graph generated by those pixels colored as “1”, to obtain the color classes “10” and “11”. This hierarchical process of binary coloring is repeated a number of iterations until the image segmentation is obtained. It is important to note that we consider as homogeneous any region or subset of pixels whenever these pixels are connected and have the same color. Once this algorithm is finished, the segmentation information can be included in any of the standard unsupervised classification algorithms in order to improve the overall accuracy of classical algorithms that only take into account the spectral information of each pixel (see [22]). The key to this segmentation algorithm is therefore the basic binary coloring process we apply to each step.

## 3.2 Binary Coloring Algorithm

A binary coloring of a graph $G = (V, E)$ is a particular case of a 2-coloring: $col : V \rightarrow \{0, 1\}$. The binary coloring procedure that we propose as the basic procedure, colors two adjacent pixels as “0” and “1” depending on the fuzzy dissimilitude between them, when compared with a prescribed threshold (a procedure based on distribution percentiles is suggested in [23], in order to determine the different values of this prescribed threshold). Note that a standard crisp approach assigns the same color class to any two pixels whenever such a distance is small, no matter if they are not adjacent.

To define the first binary coloring procedure, a value $\alpha$ is fixed. We can start, for example, with pixel (1, 1) in the top-left corner of the image, and then pixels can be colored from left to right and from up to down, in the following way:

$$col(i+1, j) = \begin{cases} col(i, j) if d_{p_{i,j}, p_{i',j'}} \tilde{\geq} \alpha \\ 1 - col(i, j) if d_{p_{i,j}, p_{i',j'}} \tilde{\leq} \alpha \end{cases} \forall (i, j) \in \{1, \ldots r\} x \{1, \ldots s - 1\}$$

and

$$col(i, j+1) = \begin{cases} col(i, j) if d_{p_{i,j}, p_{i',j'}} \tilde{\geq} \alpha \\ 1 - col(i, j) if d_{p_{i,j}, p_{i',j'}} \tilde{\leq} \alpha \end{cases} \forall (i, j) \in \{1, \ldots r - 1\} x \{1, \ldots s\}$$

In order to determine if the fuzzy number $d_{p_{i,j},p_{i',j'}}$ is greater than $\alpha$, a ranking function will be used. Therefore, given a colored pixel $p_{i,j}$, the adjacent pixels $p_{i+1,j}$ and $p_{i,j+1}$ can then be colored in a similar way. However, since pixel $p_{i+1,j+1}$ can be colored either from pixel $p_{i+1,j}$ or from pixel $p_{i,j+1}$, both coloring processes may not lead to the same color. The we are talking about an *inconsistent* coloring process. This means that our binary coloring procedure is also dependent on the particular order we have chosen for coloring.

In the proposed algorithm, once all pixels are colored, we look for a value $\alpha^*$ that assures consistency. Then, pixels are classified either as color class "0" or "1", and afterwards we proceed to get a more precise color for both classes (class "0", for example, will switch either into "00" or "01"). This process is performed separately, by alternatively activating only one of the classes already colored at a previous stage. The same process will be applied to subsequent stages, in such a way that the above binary coloring process is carried out to the activated pixels under consideration, i.e., a subset $P'$ of pixels contained within $P$.

As any representation technique, the tool presented in this section gives decision makers an additional understanding in order to obtain a more accurate description of the images, in our case involving fuzzy classes. Our hierarchical output offers a systematic sequence of colored images that can be carefully analyzed by decision makers for a more global understanding of the image (depending on their objectives and capabilities).

As pointed out in [20], there is an absolute need for manageable descriptive tools in order to show fuzzy uncertainty. In fact, the information given by this algorithm can be included in different classical unsupervised classification methods in order to improve the training site description.

## 4 Some Computational Results

In this section we show the performance of the above segmentation technique for some common test digital images, and then we consider a particular Earth surface image obtained by remote sensing. The information contained in the segmentation process will be used in the subsequent fuzzy classification algorithm in order to improve the accuracy of the final classification.

As a first experiment, two standard images have been considered (see Fig. 2 and Fig. 4).

In order to model distance uncertainty, in these examples we consider symmetrical triangular fuzzy numbers for the fuzzy graph, based on the Euclidean distance, $d_{p,p'} = (d - er, d, d + er)$, where

$$d = \sqrt{\sum_{k=1}^{b} (x_p^k - x_{p'}^k)^2} \qquad \forall (p, p') \in P$$

**Fig. 2** Squares

is the deterministic Euclidean distance and "*er*" is the error measurement considered by the expert (see [24] for more details).

Hence, a sequence of binary classifications has been obtained, so a binary number is associated with each pixel and the image is divided into regions of adjacent pixels, all of them with the same associated color class. In order to be able to visualize these regions, we can paint regions taking advantage of the full RGB color range, obtained as the mean of the original color of pixels in each region (see segmented images in Fig. 3 and Fig. 5).

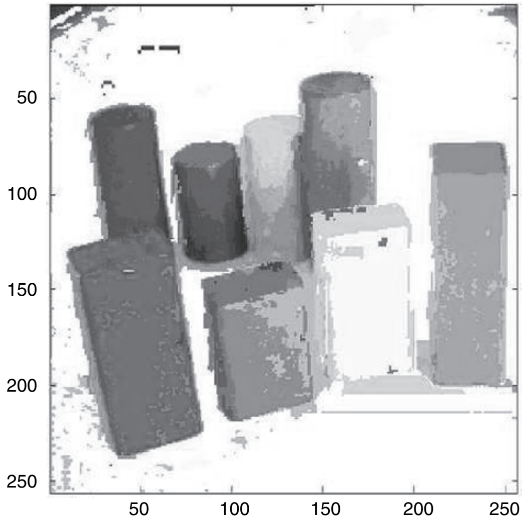
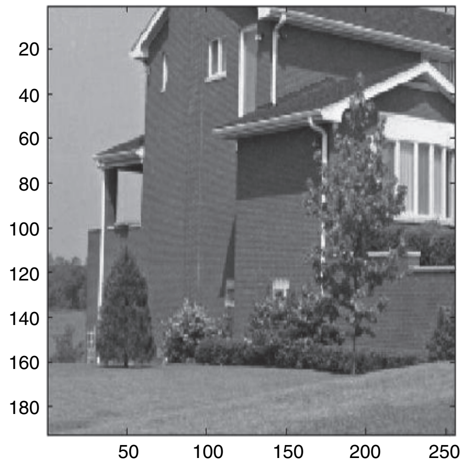



**Fig. 3** Segmented squares

**Fig. 4** House




## *4.1 Application to a Remote Sensing Image*

This section is devoted to summarize and describe how previous processes can be aggregated in order to obtain a final fuzzy classification.

As already pointed out, the classification process is complex and requires some preprocessing. We want to emphasize how representation techniques can be useful to improve such a *training sites and pattern recognition* step. Given a remote sensing image or digital image, we will describe shortly the interactive process proposed in this paper.

- The image is preprocessed in the sense described in Sect. 3. The ideal sensor is chosen and any necessary transformation of the date is made.
- After such a preprocessing, we model the remote sensing image as a fuzzy planar graph $\tilde{G} = (P, \tilde{A})$, where the nodes set $P$ represent the pixels of the image. The

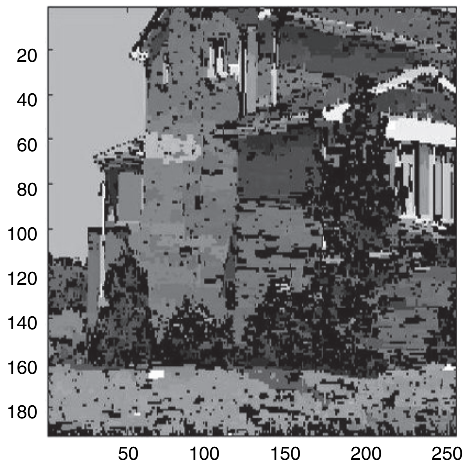



**Fig. 5** Segmented house

link set $\tilde{A}$ is given by a matrix of fuzzy numbers, each one of them representing the dissimilarity degree between two adjacent pixels.

- Once the model is built up, the coloring algorithm is applied to the fuzzy graph. This coloring process divides the set of nodes-pixels $P$ in homogeneous regions $R_k$, in such a way that $P = \cup R_k$ and $R_k$ are connected in the crisp graph $(P, \overline{P})$.
- The homogeneous regions determined in the above step, together with the information of the segmentation process, can be useful in order to improve standard non-classification algorithms. In particular, we can bring here the non-supervised fuzzy classification algorithm suggested in [3]. This process begins with an initial fuzzy classification given by the fuzzy c-means algorithm. This initial fuzzy classification can be improved by means of a learning algorithm that takes into account the concepts of relevance, redundancy and covering, previously defined in Sect. 2. But this algorithm classifies each pixel without taking into account behavior of neighborhood pixels. For this reason, we will combine the information

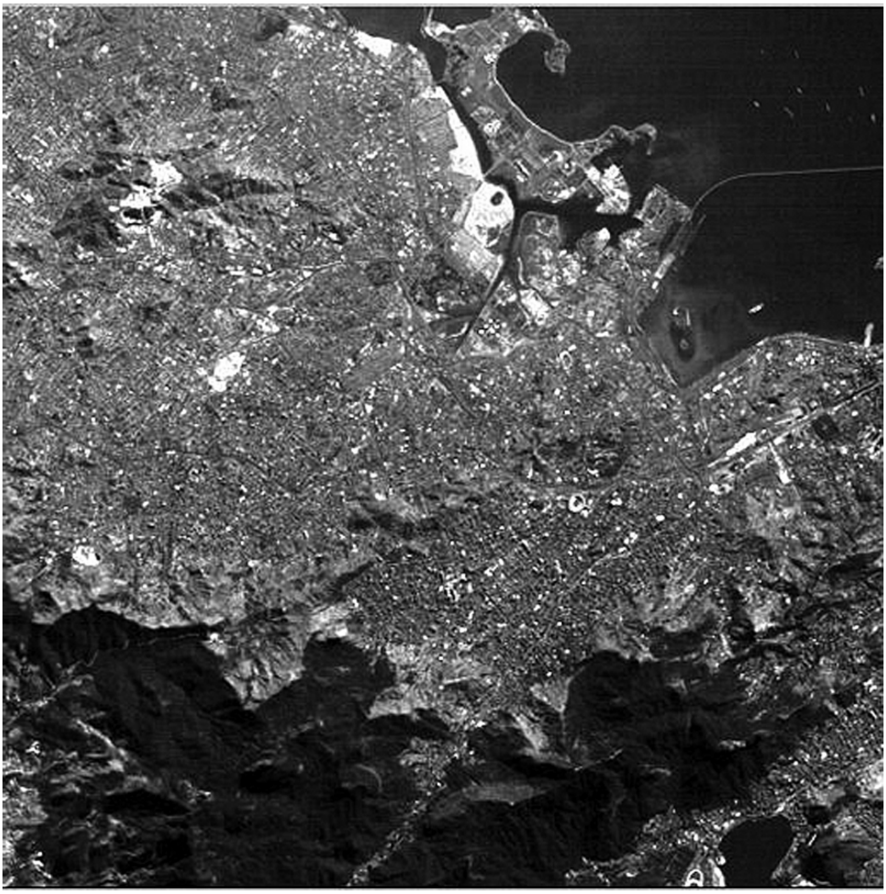


**Fig. 6** Remotely sensed image of Rio Janeiro

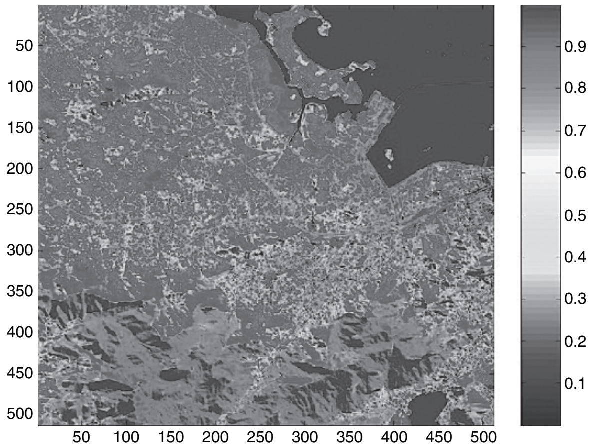



**Fig. 7** Classification by means of the fuzzy c-means

contained in the above steps with the algorithm proposed in [3]. In particular, fuzzy c-means algorithm (considered in [3] as the initial solution) has been modified, keeping the two bigger homogeneous regions found in the segmentation process. After that, classification learning can follow the algorithm described in [3].

## *4.2 A Particular Case*

The above two step process has been applied to a remote sensing image from Rio de Janeiro (coast of Brazil, see Fig. 6), first producing a segmented image by means of the above algorithm.

In order to compare the improvement introduced by our previous segmentation process, two different fuzzy classifications (Fig. 7 and Fig. 8) have been carried

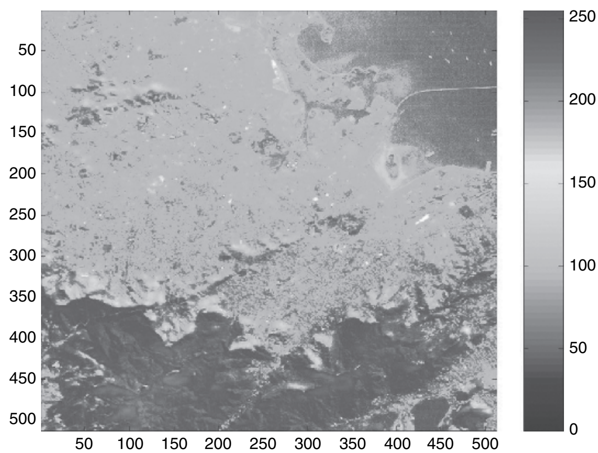



**Fig. 8** Fuzzy classification after segmentation

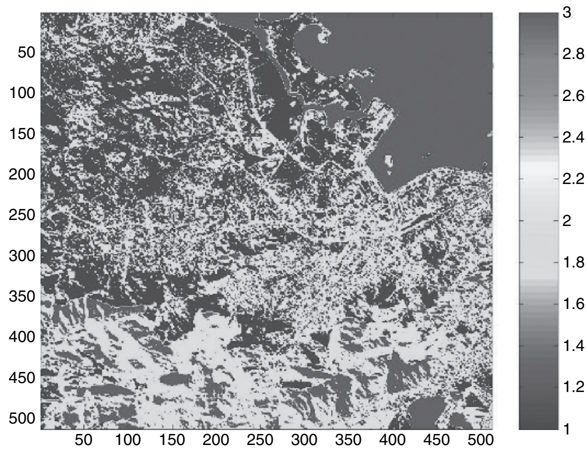



**Fig. 9** Crisp classification (fuzzy c-means)

out. The first (see Fig. 7) has been obtained by means of the classical *fuzzy c-means* algorithm [10]. We can observe that some pixels have been misclassified, most likely because this classification does not take into account the information associated with the neighborhood of each pixel. These errors can be appreciated with more detail in its associated simplified crisp classification (see Fig. 9).

The second classification (see Fig. 8) has been attained in accordance with the arguments of the above section. In this way, the information given by our segmentation process is taken into account, producing a fuzzy classified picture in three classes, which is shown in Fig. 8. From this fuzzy classification, a simplified crisp classification (Fig. 10) is analogously derived as before. Note the great analogy of this figure to the first classification the human eye will most likely make, distinguishing the main three concepts being present in such a picture: water, forest and urban area.

We also show below intermediate information that we can compare with both procedures: on one hand, degrees of membership to main three classes obtained

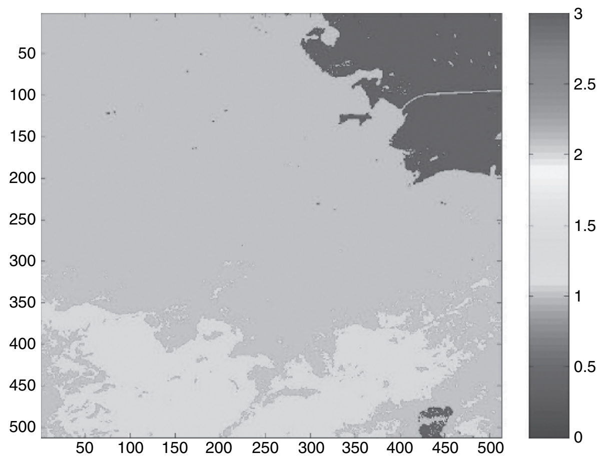



**Fig. 10** Crisp classification after segmentation

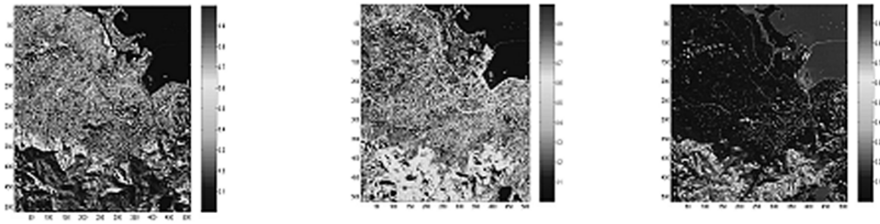


**Fig. 11** Degrees of membership of the three main classes in Fig.7

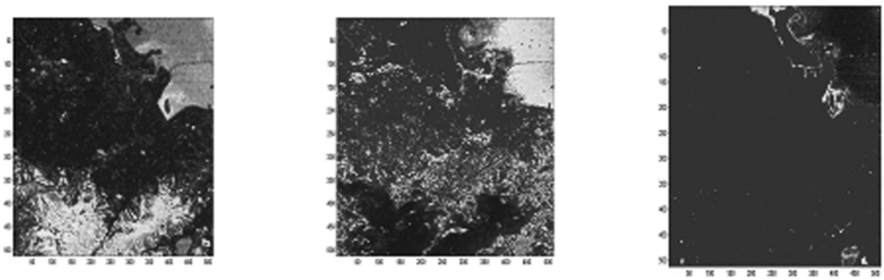


**Fig. 12** Degrees of membership of the three main classes in Fig. 8

by fuzzy c-means are shown in Fig. 11, where high degrees of membership appear associated with dark tones (red in our original colored picture). On the other hand, distances taking into account segmentation are shown in Fig. 12 (to "Forest", "Urban area" and "Water" classes, in this order). Note that in this case high degrees of membership are therefore associated to clear tones (blue in our original colored picture).

Although we present grey pictures here, it should be noted that their comprehension is not easy at all, even in their original colored version. Since both classifications are being made pixel by pixel, patterns are difficult to capture. The simplified crisp segmented image given in Fig. 10 is indeed helping us to capture the existence of the three main concepts explaining the image.

Some accuracy measures can be tried in order to quantify improvements between classification, and to get objective quality evaluations. If [22] and [15] are taken into account, for example, we find that for the crisp classification of Fig. 10 obtained by our approach, the accuracy measure is 0.97 (i.e., 97% of the total data have been classified correctly). This result represents a clear improvement if compared with the crisp classification given in Fig. 9, obtained by standard fuzzy c-means, where the total accuracy is 0.43.

## 5 Final Comments

The main objective of this paper is to point out that a *bad* classification system will usually be the beginning of our learning process, so it is extremely restrictive to force our model to fit desired theoretical conditions, which should only be understood as an objective to be reached through experience, if this is the case. This soft approach has allowed us to apply unsupervised classification systems without risky manipulations, and a particular application within remote sensing has been shown. Of course, we can take advantage of additional information obtained by other means, such as for example a segmentation procedure, in order to improve results.

It should be noted that the supervised classification being produced in this way takes into account the picture itself, not requiring any field experimentation such as standard expensive supervised methods in remote sensing. Of course this additional learning procedures can be implemented at a subsequent stage.

**Acknowledgment** This research has been partially supported by the Government of Spain, grants $MTM2005-08982-C04-01$ and $TIN2006-06190$, together with a **Del Amo** grant from a joint programme between Complutense University and the University of California.

## References

1. A. Amo, D. Gómez, J. Montero and G. Biging (2001): *Relevance and redundancy in fuzzy classification systems*. Mathware and Soft Computing 8:203–216.
2. A. Amo, J. Montero and G. Biging (2000): Classifying pixels by means of fuzzy relations. *Int. J. General Systems* 29:605–621.
3. A. Amo, J. Montero, G. Biging and V. Cutello (2004): Fuzzy classification systems. *European Journal of Operational Research* 156:459–507.
4. A. Amo, J. Montero, A. Fernández, M. López, J. Tordesillas and G. Biging (2002): Spectral fuzzy classification: an application. *IEEE Trans. Syst. Man and Cyb.* (C) 32:42–48.
5. A. Amo, J. Montero and V. Cutello (1999): On the principles of fuzzy classification. In *Proceedings North American Fuzzy Information Processing Society, NAFIPS Conference*, IEEE PRESS, Piscataway, NJ; 675–679.
6. A. Amo, J. Montero, E. Molina (2001): On the representation of recursive rules. *European Journal of Operational Research* 130:29–53.
7. K.T. Atanassov (1986): Intuitionistic fuzzy sets. *Fuzzy sets and Systems* 20:87–96.
8. K.T. Atanassov (1999): *Intuitionistic Fuzzy sets*, Physica-Verlag., Heidelberg.
9. G. Beliakov, R. Mesiar and L. Valaskova (2004): Fitting generated aggregation operators to empirical data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 12:219–236.
10. J.C. Bezdek (1981): Pattern Recognition with Fuzzy Objective Function Algorithms. *Plenum Press*, New York.
11. J.C. Bezdek and J.D. Harris (1978): Fuzzy partitions and relations: an axiomatic basis for clustering. *Fuzzy Sets and Systems* 1:111–127.
12. D. Butnariu (1983): Additive fuzzy measures and integrals. *Journal of Mathematical Analysis and Applications* 93:436–452.
13. P. Cintula (2005): Basics of a formal theory of fuzzy partitions. In *Proceedings European Society for Fuzzy Logic and Technology, EUSFLAT Conference*, Politechnic University of Catalunya, Barcelona, Spain; 884–888.
14. T. Calvo, A. Kolesarova, M. Komornikova and R. Mesiar, Aggregation operators (2002): properties, classes and construction methods. In T. Calvo, G. Mayor and R. Mesiar, Eds.: *Aggregation Operators*, Springer; 3–104.
15. R.G. Congalton and K. Green (1999): *Assessing the Accuracy of Remote Sensed Data, Principles and Practices*, Lewis Publishers, London.

16. R.G. Congalton and G. Biging (1992): A pilot study evaluating ground reference data collection efforts for use in forestry inventory. *Photogrametic Engineering and Remote Sensing* 58:1669-1671.
17. V. Cutello and J. Montero (1999): Recursive connective rules. *Int. J. Intelligent Systems* 14: 3–20.
18. D. Dumitrescu (1992): Fuzzy partitions with the connectives $T_\infty$,$S_\infty$. *Fuzzy Sets and Systems* 47:193–195.
19. G.M. Foody (1999): *The continuum of classification fuzziness in thematics mapping*. *Photogrammetric Engineering and Remote Sensing* 65:443–451.
20. D. Gómez, V. López and J. Montero (2006): The role of fuzziness in decision making. In D. Ruan, E. Kerre and P. Wang, eds., *Fuzzy Logic: an spectrum of applied and theoretical issues*, Springer.
21. D. Gómez and J. Montero (2004): A discussion on aggregation operators. *Kybernetika* 40: 107–120.
22. D. Gómez, J. Montero and G. Biging (2006): Accuracy measures for fuzzy clasification. *Proceedings Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU Conference*, Paris, France.
23. D. Gómez, J. Montero, J. Yáñez, C. Poidomani (2007): A graph coloring algorithm approach for image segmentation. *Omega 35,* 173–183.
24. D. Gómez, J. Montero and J. Yánez (2006): A coloring algorithm for image classification. *Information Sciences 176,* 3645–3657.
25. I. Iancu (1999): Connectives for fuzzy partitions. *Fuzzy Sets and Systems* 101:509–512.
26. E.E. Kerre and M. Nachtegael, eds. (2000): *Fuzzy Techniques in Image Processing*. Physica-Verlag, Heidelberg.
27. P. Matsakis, S. Andrèfouët and P. Capolsini (2000): Evaluation of fuzzy partitions. *Remote Sensing of Environment* 74:516–533.
28. J. Montero (2004): Classifiers and decision makers. In D. Ruan et al., eds., *Applied Computacional Intelligence*, World Scientific, Singapore; 19–24.
29. J. Montero (1986): Comprehensive fuzziness. *Fuzzy Sets and Systems* 20:89–86.
30. J. Montero (1987): Extensive fuzziness. *Fuzzy Sets and Systems* 21:201–209.
31. J. Montero and D. Gómez (2003): Preferences, classification and intuitionistic fuzzy sets. *Proceedings EUSFLAT Conference*, University of Applied Sciences, Zittau, Germany; 187–192.
32. S. Muñoz, T. Ortuño, J. Ramírez and J. Yáñez (2005) : Coloring fuzzy graphs. *Omega* 33: 211–221.
33. S.K. Pal, A. Gosh and M.K. Kundu (2000): *Soft Computing for Image Processing*. Physica-Verlag, Heidelberg.
34. F.E. Petry, V.B. Robinson and M.A. Cobb (2005): *Fuzzy Modeling with Spatial Information for Geographic Problems*. Springer, Berlin.
35. B. Roy (1993): Decision science or decision-aid science. *European Journal of Operational Research* 66:184–203.
36. E.H. Ruspini (1969): A new approach to clustering. *Information and Control* 15:22–32.
37. B. Schwizer and A. Sklar (1983): *Probabilistic Metric Spaces*. North-Holland, New York.
38. E. Trillas (1979): Sobre funciones de negación en la teoría de conjuntos difusos. *Stochastica* 3:47–59, (in Spanish). Reprinted (English version) in S. Barro, A. Sobrino and A. Bugarin, eds. (1998): *Advances of Fuzzy Logic*, University of Santiago de Compostela, Spain; 31–43.
39. L.A. Zadeh (1965): Fuzzy Sets. *Information and Control* 8:338–353.
40. L.A. Zadeh (1973): Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics* 1:28–44.

# FIS2JADE: A New Vista for Fuzzy-oriented Agents

**Vincenzo Loia and Mario Veniero**

**Abstract** In this chapter we present a new generalized agent-based model to implement fuzzy logic controllers (FLCs) distributed according to the Agents Paradigm. The discussed model is presented along with FIS2JADE, a tool developed to convert Simulink® FIS-format controllers into a Multi-Agent System adhering to the defined model, in order to improve developers efficiency when dealing with distributed control. We present the model describing first its founding elements and concepts. Then we present the model design by means of a simple and well-known case study, the stationary inverted pendulum control problem, allowing us to focus on the model definition rather than on the control problem to which it is applied. Lastly, we show snapshots from the detailed design and implementation process when choosing the JADE Agents Platform to implement the derived Multi-Agent System.

## 1 Introduction

Artificial Intelligence (AI) techniques have frequently been used to tackle the most difficult automation problems. This is particularly true in industrial applications where conventional software and teams of operators were unable to cope with the demands of rapidly changing, complex environments. This explains the richness of decision-support systems reported in scientific and industrial literature and designed to assist different aspects of the control engineer's job. Different approaches apart, a "first" phase, ranging from the late 1970s to early 1990s, and characterized by the use of AI in industrial application, ends with a unanimous learned lesson, the difficulty in balancing effective realization of multi-control modelling with design simplicity.

Since 1985, a new tendency emerged in the field of AI: "agent"-based approaches became the alternative to main-stream "knowledge-based" systems. An agent is an entity capable of carrying out goals as a component of a wider community of agents that interact and cooperate with each other. Agents work by exploiting concurrent and distributed processing, and performing some activities autonomously. Autonomy improves system modularization by organizing the modelling of the system in terms of delegation and distribution activities. Autonomy of agents evolves with the life of the system: when the agent shows an ability to improve its behaviour over

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

the time then the agent is said to be adaptive. These features influence the design choices of recent intelligent systems, also thanks to these unstoppable trends:

- Computers are smaller, more powerful, and cheaper (everywhere powerful machines are no longer prohibitively expensive);
- Improvements in network technology (real advantage of distributed computing);
- Emerging of object-oriented programming paradigms (active objects, delegation, reusing and asynchronous message passing support naturally agent-based design).

Since agents embody inner features that underlie adaptive, robust and effective behaviour, it is more natural for designer to mix together heterogeneous techniques in order to improve representation and handling of dynamic environments at different level of granularity and complexity. Furthermore, the possibility to integrate useful pieces of intelligence into embedded artefact makes realizable scenario of ubiquitous computing. This vision has strongly stimulated the research community in envisaging agents stemmed with physical devices. These agents are named embedded agents, that is agents that run in an embedded system or device. Kaelbling and Rosenschein refer embedded agents (or situated automata) in a finite-state-machine structure.

The effectiveness of agent decision depends on 1) the "quality" of the knowledge of its world and 2) on the robustness of its inference mechanism. One solution is to enrich the environment with as many sensors as possible (in fact, large sensory sets are an issue for ubiquitous computing), but the drawback is the explosion of data and, consequently, of the difficulty in finding "relevant" data. The problem cannot be solved only in terms of size of information: even in case of a space full of sensors, we must remember that people act in a non-deterministic way and often their actions are characterized by qualitative, subjective information.

Fuzzy Logic can play a fundamental role in coping with this difficult problem thanks to its ability in bringing together human expertise and deals with the uncertainties and imprecision typical of complex environments. Often, human beings take decisions based on rules: from an operating viewpoint this approach is extremely useful, since rule-based systems, consisting of if-then forms, represent knowledge in terms of rules that indicate what to do in different situations instead of representing knowledge in a static way. In a similar way, fuzzy rules tend to mimic expert behaviours by using a series of if-then statements: the decision itself and the means of choosing that decision are replaced by fuzzy sets and the rules are replaced by fuzzy rules. This simple but powerful idea, explains why fuzzy control is the most successful and dynamic branch of fuzzy logic system technology, in terms of both theoretical research and practical application. Fuzzy logic controller (FLC) is used whenever conventional control methods are difficult to apply or show limitations in the performances, for example, due to a complex system structure. FLC approach is closer to the human, thanks to its ability to determine outputs for a given set of inputs without using conventional, mathematical models. FLC, by using one set of control rules and membership functions, converts linguistic variables intonumeric values required in most applications, revealing robustness with regard to noise and variations of the system parameters.

When designing very complex control strategy using hybrid technology, one usually faces the challenge of balancing effective realization of multi-control modelling with design simplicity. MATLAB® is a popular high-performance language for technical computing. Initially skilled for problems such as inverting matrices, solving linear equations and so on, MATLAB® quickly has evolved towards a powerful general language, dutied of set, list and string processing, and a rich suite of plotting and graphic commands. The MATLAB website provides hundreds of MATLAB® programmes supported multilingual textbooks. Inside this wide suite, we find specialized areas such as neural nets, fuzzy logic, and statistics. This could be enough to understand the worldwide success of MATLAB®, and to explain its central role in most of simulation and control projects, in academic as well as industrial contexts.

We argue that, nevertheless MATLAB®, Fuzzy Logic and agent paradigm play individually a strong role, their integration can be considered still at embryonic level. This chapter tries to foster a progress in this direction. Starting from our experiences in designing hybrid complex system with the agent paradigm, in this chapter we propose a new integrated architecture for building distributed fuzzy control systems by means of agents. The platform we describe enables to conceptualize MathWorks Simulink® fuzzy inference systems in terms of ontology-driven Java agents.

## *1.1 System Overview*

Here we summarize the system design of a Distributed Fuzzy Logic Controller as generated by “FIS2JADE” released by LASA (LAboratorio Sistemi ad Agenti, Department of Mathematics and Computer Science of the University of Salerno) in open source software under the terms of the LGPL (Lesser General Public License Version 2). This chapter is the first publication devoted to FIS2JADE and its produced model, whose version 1.0 is available for download at http://www.lasa.dmi.unisa.it/.

FIS2JADE and its produced model can be viewed as an efficient architecture for building distributed ontology-based fuzzy agent controllers. FIS2JADE has been developed as an answer to the need for efficient, inexpensive, elegant, and extensible multi-agents system-based applications encapsulating fuzzy inference engines developed and tested with MathWorks Simulink®. FIS2JADE application can generate the classes tree to execute fuzzy control activities in a FIPA (Foundation for Intelligent, Physical Agents) compliant agent platform at any level of complexity. The open source software includes a parsing and encoding application tool that can be configured to allow users to produce distributed fuzzy controllers using Java$^{TM}$, NRC FuzzyJ Toolkit [15] and JADE [16] content. FIS2JADE can be employed by researchers and developers to build robust application by separating the fuzzy control logic side, delegated to the well-known MATLAB® environment, from implementation details concerned to the agent-oriented paradigm.

#### Multi-Agent Fuzzy Inference Framework Pattern

A pattern is a piece of literature that describes a design problem along with a general solution for the problem in a given context. A pattern captures stable practices and conventional problem solving approaches. In this chapter we define the agent pattern for fuzzy inference systems generation, proving its "generative" property (according with the objective of usable software). We discuss how this approach is useful to enhance productivity while reducing development time, thanks to FIS2JADE applications.

#### Analysis, Design and Implementation Model

FIS2JADE has been analyzed and developed adopting the Rational Unified Process model, using UML as modeling language. The generated classes tree implementing the distributed agent-based fuzzy logic controller is the outcome of a common analysis and design process performed using the Gaia [11] methodology explicitly proposed for Agent-Oriented Software Engineering. The Gaia metodology has been extended to overcome its weaknesses, as described in [9], when applied to JADE Agents Platform.

#### Reusability

The FIS2JADE application tool has been designed for reuse and extensibility. New parsers and converters and their associated interfaces can be added to enhance or extend the translation engine and object definitions in order to perform new actions or implement new agents patterns.

### *1.2 Dependencies*

This section identifies external protocols and software offerings upon which this design relies.

*Java$^{TM}$*. FIS2JADE software is written using Java$^{\text{TM}}$, an object-oriented network programming language originally introduced by Sun Microsystems in 1995. The FIS2JADE framework software is intended for use with the Java 2 Standard Edition, available at http://java.sun.com/j2se/index.jsp.

*MathWorks Simulink®*. Simulink® is a platform for multi-domain simulation and Model-Based Design for dynamic systems. It provides an interactive graphical environment and a customizable set of block libraries, and can be extended for specialized applications. Further information are available at http://www.mathworks.com/products/simulink/.

*JADE.* JADE (Java Agent DEvelopment Framework) is a software framework fully implemented in Java language. It simplifies the implementation of multi-agent systems through a middle-ware that complies with the FIPA specifications and through a set of graphical tools that supports the debugging and deployment phases. The JADE platform is available for download at http://jade.cselt.it/.

*NRC FuzzyJ Toolkit.* The NRC FuzzyJ Toolkit from the National Research Council of Canada's Institute for Information Technology is a set of Java classes that provide the capability for handling fuzzy concepts and reasoning. The toolkit is available for download at http://www.iit.nrc.ca/IR_public/ fuzzy/fuzzyJToolkit2.html.

## 2 The FIS2JADE View of a Distributed Fuzzy Controller

This section introduces main concepts of FIS2JADE view of a fuzzy controller. It starts by defining the notion of distributed fuzzy controller as seen by FIS2JADE, which is required to understand the generated classes and relationships among them. Then the main concepts and relationships of the view are described.

Moreover, the section describes the interactions occurring among roles specifying the ontology-based communication interactions that have been defined by means of the adopted distributed fuzzy control analysis and design model to help the analyst/designer using FIS2JADE to structure and extend the generated classes.

### *2.1 Foundations*

The FIS2JADE interpretation of a distributed fuzzy controller can be described as follows. A distributed fuzzy controller is a collection of fuzzy sensor and consumer roled agents exchanging fuzzy values and interacting on a pure reactive base. The Role concept allows the part played by an agent to be logically separated from the identity of the Agent itself. The distinction between Role and Agent is analogous to that between Interface and (object) Class: a Role describes the external characteristics of an Agent in a particular context. This is useful in defining re-usable patterns. Some sensor-like agents are specialized as physical sensors whilst some consumer-like agents are specialized as physical actuators. Moreover, the same agent may be capable of playing several roles, both consumer and sensor typed, this may be the case of inference engine consumers acting as input fuzzy values consumer and output fuzzy values producers (sensors). Finally, multiple Agents may be able to play the same Role. Each interaction among roles is driven by specific communicative ontologies induced by involved fuzzy values and initiator/responder agent roles. The rest of this section describes the knowledge level concepts and interactions that feature most prominently in the FIS2JADE view of a fuzzy controller as it stands at the moment.

### *2.2 Knowledge-level Concepts*

Most of FIS2JADE knowledge-level entity concepts can be grouped into the following main categories: **Fuzzy Concepts**, **Fuzzy Sensors**, **Fuzzy Consumers**. Figure 1 gives an informal overview of how these concepts are inter-related.

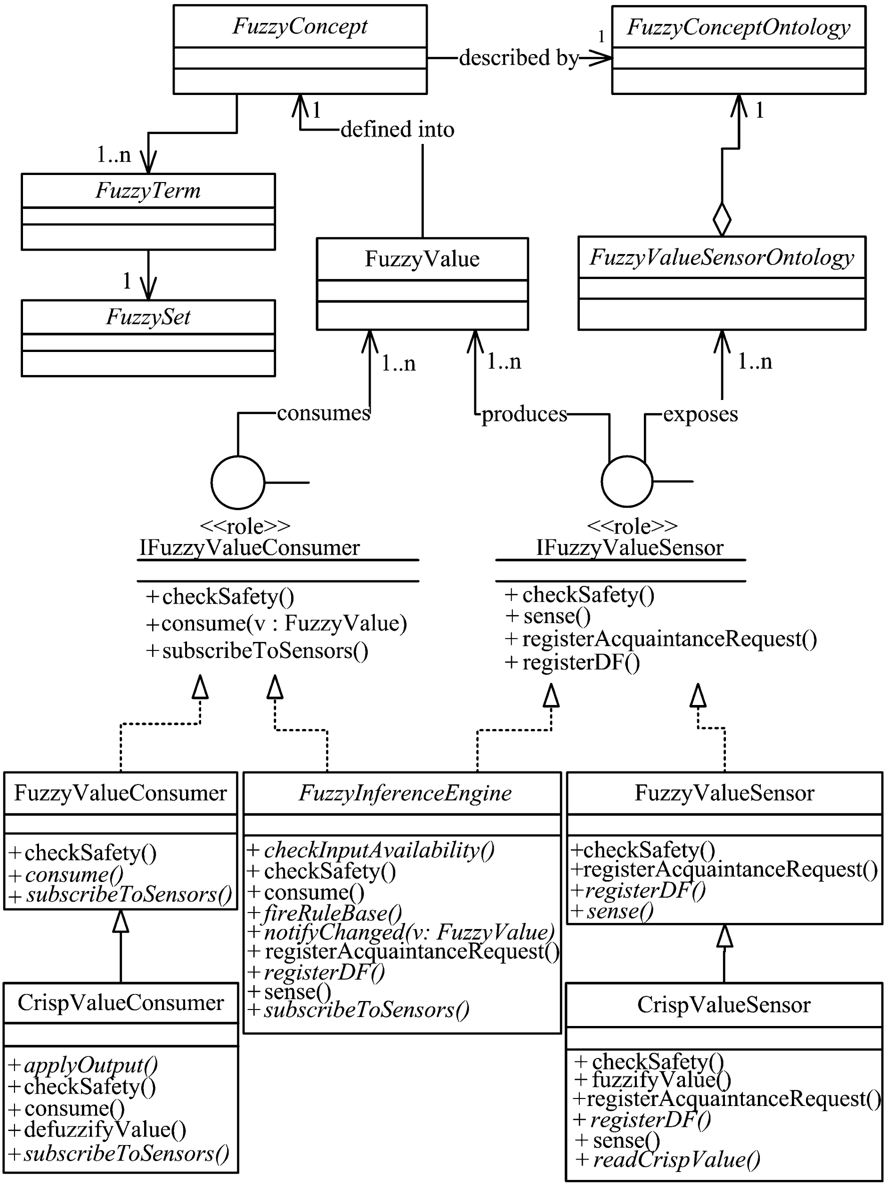



**Fig. 1** FIS2JADE view of a fuzzy controller

*Fuzzy Concept.* A Fuzzy Concept entity defines the language that will be used to discuss a fuzzy concept such as **temperature**, **pressure**, **age**, or **height**. Each Fuzzy Concept is characterized by providing a name (for example, **"temperature"**), the units of the variable if required (for example, **"degrees C"**), the universe of discourse for the concept (for example a **"range from 0 to 100"**), and a set of primary fuzzy terms (like **"hot"**, **"cold"** and **"warm"**) that will be used when describing the specific concept's components associated with the fuzzy variable. The name and units are elements that are mainly used to generate the induced ontology to describe the Fuzzy Concept. Alternately, the universe of discourse defines a set of upper and lower bounds for the values of the fuzzy sets used to describe the concept.

*Fuzzy Term.* As already stated, fuzzy terms are used to describe a specific fuzzy component associated with a more general Fuzzy Concept. A fuzzy term is described using a **term name** such as **"hot"**, along with a fuzzy set that represents that term, i.e. a mapping of a set of real numbers ($x_i$) onto membership values ($u_i$) that (generally) lie in the range [0, 1]. This description is made a part of the ontology induced by the associated fuzzy variable.

*Fuzzy Value.* The FuzzyValue defines the same concept identified by a **Fuzzy Term** even though from a more generalized view point. In particular, a Fuzzy Value is an association between a FuzzyConcept and a linguistic expression used to describe a specific compomponent of it. However, a FuzzyValue can be created using a FuzzyConcept and a FuzzySet, loosing any human language-like expression associated with it. The FIS2JADE framework uses it to define a crisp value-derived term to be used by the inferential process. Here a crisp value is a precise numerical value obtained from a physical sensor and fuzzified (translated into a Fuzzy Value) by a CrispFuzzyValueSensor located somewhere in the network and wired to the physical sensor.

*FuzzyValueSensor.* A FuzzyValueSensor for the distributed FLC is a common design interface (or stereotype) aimed to identify each MAS role acting as a fuzzy input value sensor for the FLC. A FuzzyValueSensor exposes and communicates through specific ontologies describing involved fuzzy values and concepts. Futhermore it is specialized by a CrispValueSensor when dealing with physical sensors of the system.

*FuzzyValueConsumer.* Analogous to sensor, a controller consumer is a common design interface (or stereotype) defined to identify each MAS role acting as a sensed fuzzy value consumer. It worths noticing here the producer/consumer approach adopted to define the distributed fuzzy controller design pattern. A FuzzyValueConsumer registers to specific sensors in order to obtain the measured fuzzy values. As for sensors, it is specialized into CrispValueConsumer when dealing with physical actuators of the system.

## 3 Case Study-driven Design

The following paragraph describes a Case Study used to design the modeling pattern adopted by FIS2JADE while trying to illustrate the benefits coming from converting a single MathWorks Simulink® FIS controller into a multiagent distributed controller.

This case study is applied to the well known inverted pendulum as a very classical control problem. From an application point of view, the inverted pendulum is a common mechatronic [7] application existing in many different forms, all sharing a common goal: to balance a link on end using feedback control.

In our application context, the system (Fig. 2) being controlled is a computer-simulated stationary inverted pendulum composed by a variable mass on the top of a flexible pole hinged to a stationary variable power strenght motor through a free joint with one degree of freedom. The motor will try to balance the pendulum depending on the bopping and bobbing forces excerted on its top mass. This last force makes the top mass appear as if it were on a spring rather than a rigid stick. Moreover, the top mass is pushed by a bopping force $f_1$ so to determine an angle from the vertical that is greater than zero, obtaining an angular falling rate $\omega$ increasing or decreasing as the force of gravity applies in conjunction with an eventual bobbing force $f_2$ tapping the upper mass down to the lower one thus shrinking and extending the pole of the pendulum. The typical balancing act is performed by the lower variable mass strenght motor aiming to reverse the falling rate while trying to achieve an angle from the vertical that is close to zero. In parametric values terms, the desired angle and angular rate are both zero. In the absence of any other perturbation, if the initial angle and rate are both zero then the balanced position is maintained. However, if there is a perturbation induced by the bopping force $f_1$, then the gravity increases the falling rate forcing the angle to a side of the vertical. Without any control action, the pole will fall down making the goal unachieved. Obviously, if the initial position in unbalanced then the control is required immediately.

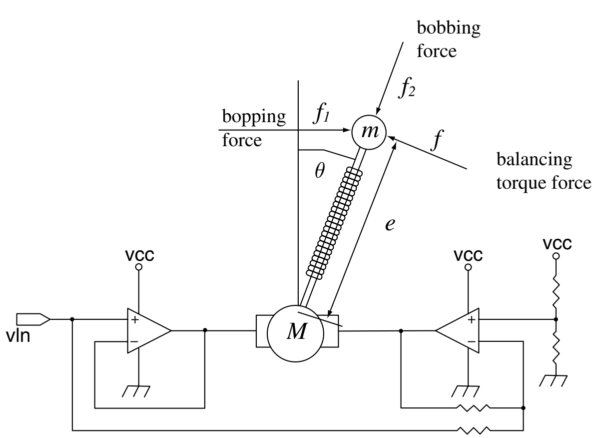



**Fig. 2** Computer-simulated stationary inverted pendulum

### 3.1 A Fuzzy Logic Controller for Inverted Pendulum

Classical control techniques require a deep mathematical background. Nevertheless, it can be recognized that a small set of common sense rules suffice to solve the problem (we will see later, a few basic intuitive rules used and tuned to handle the balancing act). With respect to this consideration, fuzzy controller was introduced and developed to solve this control problem.

Since our main interest is not how to build the fuzzy logic controller to solve the inverted pendulum control problem, but is primarly involved in showing how a generic control problem can be solved using a fuzzy logic controller implemented according to the agent paradigm and adopting the FIS2JADE vista of fuzzy control, this phase is skipped and a fine-tuned system will be given directly.

The controller interaction with the system is primarly dependent on its state. This appears defined in terms of two main input variariables: angle $\theta$ with the vertical and angular rate $\omega$ of falling. A further characterization of the system is obtained from the maximum power of the motor $M$ and the mass $m$, respectively, used as bottom motor power and top stick mass, as well as the original stick length $l$ and the eventual stick shrinking and expansion due to the bobbing force $f_2$. Each of these affects the main inputs of the controller during the balancing activity. Obviously, the controller's output should be the torque force $f$ to be excerted by the motor to achieve the balancing goal.

With respect to the angle with the vertical, we can think at its maximum value being $+\frac{\pi}{2}$, at which the pole lies flat on a side (for instance, the right hand side) and the minimum value being $-\frac{\pi}{2}$, at which the pole lies flat on the opposite side. Thus the universe of discourse for the angle parameter $\theta$ is $\left[-\frac{\pi}{2}, +\frac{\pi}{2}\right]$. To define the universe of discourse for the angular velocity $\omega$ we could reason, for sake of simplicity fixing some constant values. Let us assume the pole has an allowed extension in the range [0.7, 1.3]m whilst the size of the mass $m$ can range in [0.5, 3.5]kg. Furthermore, let us assume the bumping is such to introduce a perturbation able to set the mass starting position to an angle $\theta = +\frac{\pi}{10}$ with the vertical. Finally let us range the motor size in the interval [60, 180] Amps. With these constraints is reasonable to assume a maximum angular rate equal to $\frac{4}{3}\pi$, leading us to set the universe of discourse $\left[-\frac{4}{3}\pi, \frac{4}{3}\pi\right]$/s.

To deal with the controller's output, we must consider that the presented scheme adopts a variable strenght motor with a linear power driving scheme realized by two amplifiers with a bridge configuration. The example shows, on a principle base, the left amplifier shaped like a tension pursuer (Vout(1) = Vin), while the right amplifier is acting as a differentiator (if the four resistors are equal, Vout(2) = Vcc - Vin). This configuration let us set modulus and direction of the Vm input tension driving in this way the motor rotation. In the light of these assumption, letting the maximum working tension be 115V, and the motor power range from 60W to 180W, respectively 0.52A and 1.56A at the fixed maximum working tension, we could set the universe of discourse of controller's output to [0, 115]/V. Having defined the fuzzy labels, shown in Table 1, the membership functions for each input and output and the controller's rule base are depicted in Fig. 3.

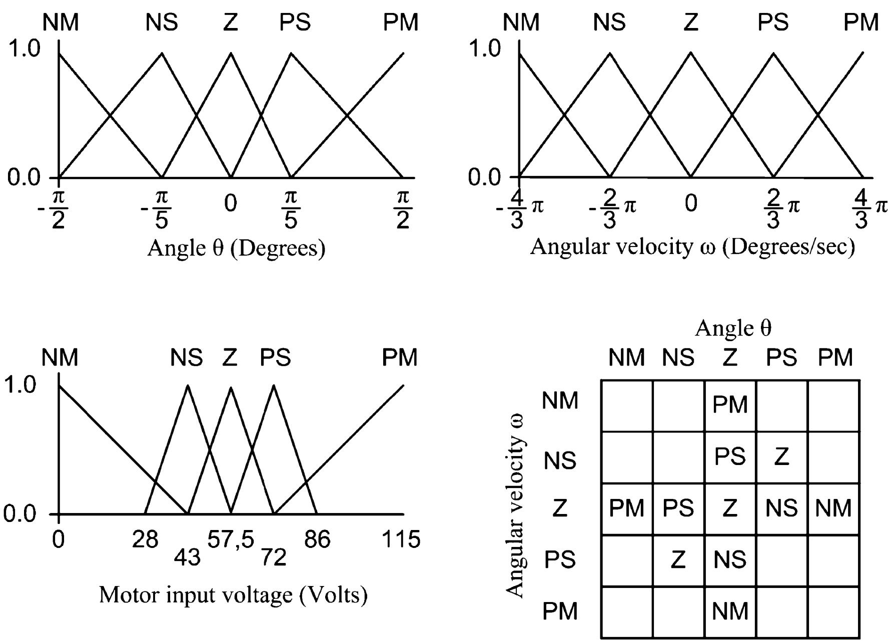
**Table 1** Fuzzy labels for inverted pendulum FLC rule base definition

| Label | Description |
|---|---|
| NM | Negative Medium |
| NS | Negative Small |
| Z | Zero |
| PS | Positive Small |
| PM | Positive Medium |

The adopted fuzzy inference scheme is the classical Mamdani min-max composition and the choosen defuzzification method is the centroid as shown in (1).

$$COG = \frac{\sum_{i=1}^{N} x_i^{'} \cdot A_i}{\sum_{i=1}^{N} A_i} \tag{1}$$

So far the stationary inverted pendulum fuzzy logic controller has been completed.




| Angular velocity ω \ Angle θ | NM | NS | Z | PS | PM |
|---|---|---|---|---|---|
| NM | | | PM | | |
| NS | | | PS | Z | |
| Z | PM | PS | Z | NS | NM |
| PS | | Z | NS | | |
| PM | | | NM | | |

**Fig. 3** Membership functions and rule base for stationary inverted pendulum

### 3.2 Simulink® FIS for Inverted Pendulum FLC

To fully demostrate the controller performance, a suitable simulation environment is provided by MATLAB®/Simulink®. The environment allows us to simulate the whole system to control by means of composition of functional building blocks. Among these we find the Simulink® Fuzzy Inference System implementing the defined FLC.

Since our main interest is focused on the conceptualization of a MathWorks Simulink® Fuzzy Inference System in terms of ontology-driven Java agents, in the following we will directly report (Listing 1) the FLC controller via its FIS serialization format.

**Listing 1** Simulink® FIS file for stationary inverted pendulum

```
[System]
Name='InvertedPendulum'
Type='mamdani'
Version=2.0
NumInputs=2
NumOutputs=1 NumRules=11
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'

[Input1]
Name='Theta'
Range=[-1.57 1.57]
NumMFs=5
MF1='NM':'trimf',[-3.14 -1.57 -0.62]
MF2='NS':'trimf',[-1.57 -0.62 0]
MF3='Z':'trimf',[-0.620 0.62]
MF4='PS':'trimf',[0 0.62 1.57]
MF5='PM':'trimf',[0.62 1.573.14]

[Input2]
Name='Omega'
Range=[-4.18 4.18]
NumMFs=5
MF1='NM':'trimf',[-6.28 -4.18 -2.09]
MF2='NS':'trimf',[-4.18 -2.090]
MF3='Z':'trimf',[-2.09 0 2.09]
MF4='PS':'trimf',[0 2.09 4.18]
MF5='PM':'trimf',[2.09 4.18 6.28]

[Output1]
Name='VIn'
Range=[0 115]
NumMFs=5
MF1='NM':'trimf',[-1 0 43.125]
MF2='NS':'trimf',[28.7543.125 57.5]
MF3='Z':'trimf',[43.125 57.5 71.875]
```

**Listing 1** (continued)

```
MF4='PS':'trimf',[57.5 71.875 86.125] MF5='PM':'trimf',[71.875
115120]

[Rules]
3 1, 5 (1) : 1
3 2, 4 (1) : 1
4 2, 3 (1) : 1
1 3, 5 (1) : 1
2 3, 4 (1) : 1
3 3, 3 (1) : 1
4 3, 2 (1) : 1
5 3, 1 (1) : 1
2 4, 3 (1) : 1
3 4, 2 (1) : 1
3 5, 1 (1) : 1
```

## *3.3 Agent-based Approach for Inverted Pendulum FLC*

In order to better understand the FIS2JADE view of a distribuited ontology-based fuzzy controller, we will present how the previously defined Simulink® FIS file will be handled to obtain an agent-based distributed controller. We will accomplish this task showing how the system is analyzed, designed and implemented adopting the FIS2JADE view.

The aim of this system is to provide a distributed fuzzy controller based on the given Simulink® FIS file. The MAS analysis and design will be conducted using the Gaia [11] methodology and implemented using the JADE [16] Agents Platform and the FuzzyJ Toolkit [15].

### The Analysis Phase

The analysis phase identifies six roles: ThetaFuzzyValueSensor and OmegaFuzzyValueSensor sensing, respectively, the pendulum's error from the vertical and the top mass angular velocity; ThetaFuzzyValueConsumer and OmegaFuzzyValueConsumer collecting the sensed values needed to infer control actions; VInFuzzyValueSensor using the sensed values to infer control actions to sense the VIn required to the balance act; VInFuzzyValueConsumer, that applies the sensed VIn to the bottom strenght motor in order to balance the stationary pendulum. A Gaia roles model for the system is presented in Table 2. In particular, recognizing a common structure schema descriptions are presented via a role template approach for both $\theta/\omega$ fuzzy value sensors and consumers.

It worths noticing here that, in these definitions, all interactions with the Directory Facilitator (DF) are identified as actions because of the DF interaction model provided by JADE and based on static methods invocation.

**Table 2** Distributed FLC roles schemas

**Role Schema:** ThetaFuzzyValueSensor (OmegaFuzzyValueSensor)

**Description:**

It senses the input crisp value (angle $\theta$ with the vertical, angular velocity $\omega$).
It also acquaints specific subscribed agents.

**Protocols and Activities:**

RegisterDF, ReadCrispValue, FuzzifyValue, SensorValueAcquaintingRequest<T,FVS>, RegisterAcquaintanceRequest, SensorValueNotification<T,FVS>.

**Permissions:**

| | | |
|---|---|---|
| reads | measured *sensedCrispValue<T>* | // theta angle (omega rate) |
| | supplied *consumerAID* | // acquaintance subscriber |
| | supplied *conceptTName* | |
| generates | fuzzyfied *sensedFuzzyValue<T>* | // fuzzy theta (omega rate) |
| writes | supplied *consumerAID* | |
| | supplied *conceptTName* | |

**Responsibilities:**

**Liveness:**

MONITOR = Sense<T> || SubscribeConsumers
SUBSCRIBE CONSUMERS =
RegisterDF.( AcquaintanceInformRequest<T,FVS>.
( RegisterAcquaintanceRequest.
( AcquaintanceInformAgreement<T,FVS>|
InformAboutSensedValue<T,FVS> ) )|
AcquaintanceInformRefusal<T,FVS> ) )$^{\omega}$
SENSE<T> = ReadCrispValue.FuzzifyValue.
InformAboutSensedValue<T,FVS>*

**Safety:** A successful DF registration will be held during the whole life time.

**Role Schema:** VInFuzzyValueSensor

**Description:**

It senses the VIn value needed to balance the stationary inverted pendulum. It also acquaints specific subscribed agents.

**Protocols and Activities:**

RegisterDF, CheckInputAvailability,FireRuleBase, SensorValueAcquaintingRequest<VIn,FVS>, RegisterAcquaintanceRequest, SensorValueNotification<VIn,FVS>.

**Permissions:**

| | | |
|---|---|---|
| reads | supplied *consumerAID* | // acquaintance subscriber |
| | supplied *conceptTName* | |
| generates | fuzzyfied *sensedFuzzyValue<VIn>* | // fuzzy voltage to provide |
| writes | supplied *consumerAID* | |
| | supplied *conceptTName* | |

**Responsibilities:**

**Liveness:**

MONITOR = Sense<VIn> || SubscribeConsumers
SUBSCRIBE CONSUMERS =
RegisterDF.( AcquaintanceInformRequest<VIn,FVS>.
( RegisterAcquaintanceRequest.

**Table 2** (continued)

**Role Schema:** ThetaFuzzyValueSensor (OmegaFuzzyValueSensor)

( AcquaintanceInformAgreement<VIn,VS>|
InformAboutSensedValue<VIn,FVS> ) )|
AcquaintanceInformRefusal<VIn,FVS> ) )$^{\omega}$

SENSE<VIn> =
( CheckInputAvailability.
[ FireRuleBase.InformAboutSensedValue<VIn,FVS>* ] )$^{\omega}$

**Safety:**

A successful DF registration will be held during the whole life time.
No needed input is empty when firing rule base.
At least an input has changed before firing rule base.

**Role Schema:** ThetaFuzzyValueConsumer (OmegaFuzzyValueConsumer)

**Description:**

It senses the input fuzzy value (angle $\theta$ with the vertical, angular velocity $\omega$).

**Protocols and Activities:**

SearchDF<T>, SensorValueAcquaintingRequest<T,FVC>, SensorValueNotification<T,FVC>, NotifyChanged.

**Permissions:**

| | | |
|---|---|---|
| reads | supplied *sensedFuzzyValue<T>* | // fuzzy value |
| generates | supplied *sensedFuzzyValue<T> interrupt* | |

**Responsibilities:**

**Liveness:**

PERFORM = SearchDF<T>.
AcquaintanceInformRequest<T,VC>
[ AcquaintanceInformAgreement<T,FVC> ]
( InformAboutSensedValue<T,FVC>.Consume )$^{\omega}$

CONSUME = NotifyChanged

**Safety:**

A successful DF search for T concept ontology supporting sensor has been performed.
sensedFuzzyValue<T> is not empty when firing rule base.

**Role Schema:** VInFuzzyValueConsumer

**Description:**

It consumes VIn fuzzy value to obtain the defuzzyfied crisp value to provide to the strenght motor in order to balance the falling pendulum. It also gets acquainted with specific sensor agents.

**Protocols and Activities:**

SearchDF<VIn>, DefuzzifyValue, ApplyOutput, SensorValueAcquaintingRequest<VIn,FVC>, SensorValueNotification<VIn,FVC>.

**Permissions:**

| | | |
|---|---|---|
| reads | supplied *sensedFuzzyValue<VIn>* | // fuzzy output |
| generates | defuzzified *sensedCrispValue<VIn>* | // output to provide |
| applies | defuzzified *sensedCrispValue<VIn>* | // output provided |

**Responsibilities:**

**Liveness:**

PERFORM = SearchDF<VIn>.

**Table 2** (continued)

| **Role Schema:** ThetaFuzzyValueSensor (OmegaFuzzyValueSensor) |
|---|
| AcquaintanceInformRequest<VIn,FVC> |
| [ AcquaintanceInformAgreement<VIn,FVC> ] |
| ( InformAboutSensedValue<VIn,FVC>.Consume )$^{\omega}$ |
| CONSUME = DefuzzifyValue.ApplyOutput |
| **Safety:** |
| A successful DF search for VIn concept ontology supporting sensor has been performed. |
| sensedFuzzyValue<VIn> is not empty when applying voltage. |

The adopted interaction model is presented in Fig. 4. This model consists of a set of protocol definitions and has been used to fine tune the roles model. The reported roles, for simplicity, have been identified as FVS<T> for <T>FuzzyValueSensor and FVC<T> for <T>FuzzyValueConsumer.

## The Design Phase

The design phase allows to achieve the Agent model (Fig. 5) that leads to the identification of the agent types that will make up the system and the agent instances from these. The agent model for our system will include four agent types: the ThetaSensor and OmegaSensor, fulfilling the homologous FuzzyValueSensor roles;

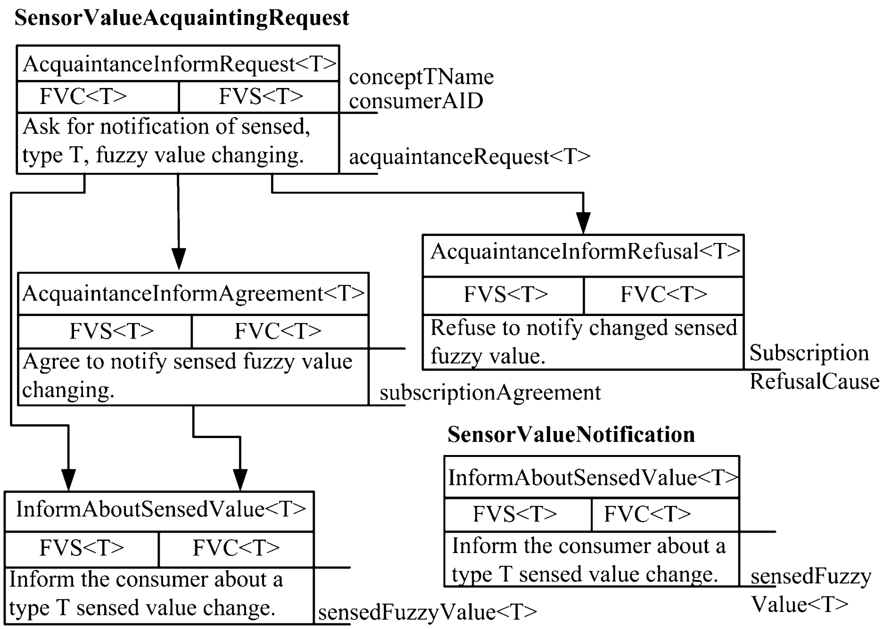


**Fig. 4** Gaia interaction model

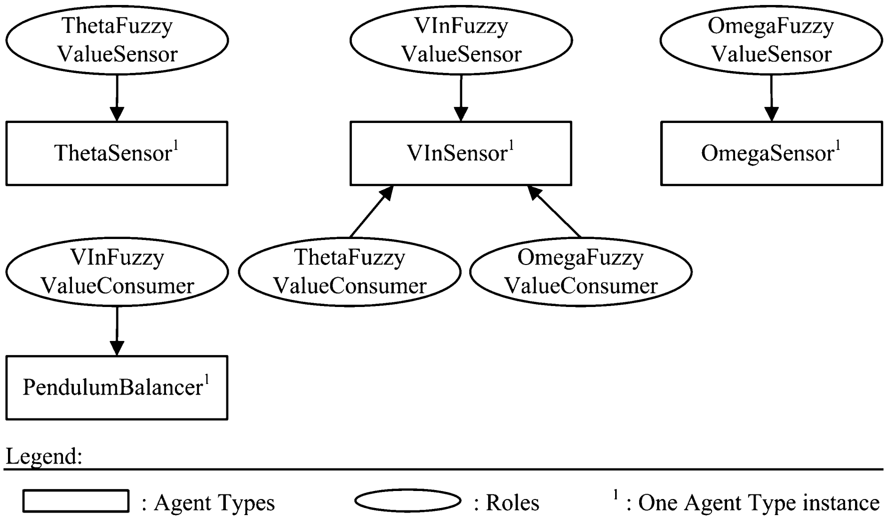



**Fig. 5** Gaia agent model

the VInSensor, who fulfills the VInFuzzyValueSensor, ThetaFuzzyValueConsumer and OmegaFuzzyValueConsumer roles; and the PendulumBalancer, who fulfills the VInFuzzyValueSensor role. Furthermore, there will be one single instance of each agent type.

Services and acquaintances model are presented, respectively, in Table 3 and Table 4, thus identifying the main services required to realize the agent's role and the lines of communication between the different agents types.

This completes the abstract analysis and design of the system enabled by Gaia.

## JADE Design and Implementation

This section shows the design and implementation activities performed by the FIS2JADE conversion tool in order to produce the JADE based MAS. This phase has been conducted according to the roadmap presented in [13]. Thus the first step is to define the ACL Messages by using the identified protocols and interaction models. Table 5 presents the more relevant ACL messages types among those defined by the previously described interactions. In Table 5, <T>FuzzyValueSensorOntology and <T>FuzzyValueConsumerOntology are specialized from the FIS2JADE Fuzzy-

**Table 3** Gaia Services Model

| | |
|---|---|
| Service | Balance Inverted Stationary Pendulum |
| Inputs | $\theta$ angle, $\omega$ rate |
| Outputs | Voltage to provide to the strength motor |
| Pre conditions | A multi-agent FLC is instantiated and associated to the physical sensors |
| Post conditions | $\theta$ and $\omega$ are reduced |

**Table 4** Gaia acquaintances model

| | ThetaSensor | OmegaSensor | VInSensor | PendulumBalancer |
|---|---|---|---|---|
| *ThetaSensor* | | | I | |
| *OmegaSensor* | | | I | |
| *VInSensor* | I,A | I,A | | I |
| *PendulumBalancer* | | | I,A | |

Legend (read occurrences in rows):
I: *Interacts*, the row indentified agent type interacts with the column identified one.
A: *Is acquainted*, the row indentified agent type has the column identified one in his acquaintances data structure.

**Table 5** ACL messages definition

| | |
|---|---|
| ACLMessage | AcquaintanceInformRequest<T> |
| Sender | FVC<T> |
| Receiver | FVS<T> |
| FIPA performative | SUBSCRIBE |
| Protocol | SensorValueAcquaintingRequest |
| Language | SL |
| Ontology | <T>FuzzyValueSensorOntology |
| Content | Ontology action: Subscribe |
| ACLMessage | InformAboutSensedValue<T> |
| Sender | FVS<T> |
| Receiver | FVC<T> |
| FIPA performative | INFORM REF |
| Protocol | SensorValueNotification |
| Language | SL |
| Ontology | <T>FuzzyValueConsumerOntology |
| Content | Ontology action: Consume |

ConceptOntology. This ontology describes fuzzy concepts, terms and sets adhering to the FuzzyJ Toolkit proposed taxonomy. Following, FIS2JADE defines data structures and software modules that are going to be used by the agents according to the played roles. Moreover JADE behaviours needed to implement liveness formulas from the roles model are written down in a bottom-up implementation process. Furthermore, for each safety properties block of the roles is defined a specific behaviour responsible for monitoring it and suspending or adding the liveness-dedicated behaviour to the agent scheduling depending on the safety condition value. Last FIS2JADE produces a static class utility aimed to startup the distributed agent platform and fuzzy logic controller. FIS2JADE, when applied to this specific case study writes down *consume* and *produce* empty placeholders the MAS developer should fill in order complete the MAS implementation (Listing 2).

**Listing 2** Customizations needed for implementation completion.

```
public class ThetaSensor extends CrispValueSensor {
    [...]
    public float readCrispValue() {
        // empty placeholder for theta crisp value sensing
    }
}
    public class OmegaSensor extends CrispValueSensor {
    [...]
    public float readCrispValue() {
        // empty placeholder for omega crisp value sensing
    }
}
    public class VInActuator extends CrispValueConsumer {
    [...]
    public void applyOutput(float value) {
        // empty placeholder for VIn value applying
    }
}
```

## 4 Conclusive Remarks and Future Work

Most real fuzzy control applications require a limited number of inputs to the rule base making the control surface suitable to define, test and implement simple fuzzy control rules in simple environments. Generally this is due to hardware, software or, more generally, performance requirements contraints. In more complex surroundings, the number of inputs would not be so limited when wishing to perform realistic tasks aided by fuzzy controllers. A typical scenario is provided by autonomous systems which typically need many sensors of different types, resulting in a huge input space.

This chapter presented a distributed agent-based model to implement FLCs by means of JADE agents platform. The model can be succesfully applied to context requiring more complex interactions among distributed inference engines. In particular, the proposed model aims to answer to some important issues about the fuzzy control system scaling up capabilities. These would allow programmers to define reactive fuzzy system able to respond to a complex environment without throwing out useful data or decreasing the input space by non-fuzzy means. This task can be achieved by defining many parallel and independent fuzzy controllers linked together by producer-consumer relationships to make up a fuzzy control network (FCN). In this FCN each fuzzy consumer node is feeded by one or more fuzzy sensors, acting as FCN sources. Inner FCN nodes can be implemented both as fuzzy consumers and sensors. Such a node accepts the outputs of other controllers as inputs and feeds with its output assignment of rules the input of other fuzzy consumers value. In this way, when such a rule set fires, the output of this control node can be added into the centroid calculation of sink fuzzy consumers allowing them to behave as fuzzy multiplexers by making smooth transitions between multiple recommendations according to qualitative rules. This decomposition, also known as fuzzy pre-processing, allows the mapping to be performed at each stage to be kept much simpler and performative than the usual approach while keeping unchanged the input space features and granularity. Moreover, the model allows to distribute FCN nodes onto autonomous, mobile and intelligent agents enabling improved distribuited inference handling strategies.

Due to the generality of the model, FIS2JADE tool has been developed to convert Simulink® FIS-format controllers into a Multi-Agent System adhering to the given model in order to improve developers efficiency when dealing with distributed control. As future works we plan to improve the model and tool allowing more complex and automatized ontology driven processes to distribute intelligence in a multi-agent based system.

## References


1. E.Gamma, R.Helm, R.Johnson, J.Vlissides. (1995) “Design Patterns: Elements of Reusable Object-Oriented Software.” Addison-Wesley, Reading, MA.
2. V.Loia - “Soft Computing Agents: A New Perspective for Dynamic Information Systems (Frontiers in Artificial Intelligence and Applications, 83)” - Ios Pr Inc (January 2003)
3. F.Bellifemine, G.Caire, T.Trucco (TILAB S.p.A., formerly CSELT), G.Rimassa (FRAMe-Tech s.r.l.), R.Mungenast (PROFACTOR GmbH). “JADE Administrator’s GUIDE”.
4. F.Bellifemine, G.Caire, T.Trucco (TILAB, formerly CSELT), G.Rimassa (University of Parma). “JADE Programme’s GUIDE”.
5. G.Caire (TILAB, formerly CSELT), D.Cabanillas (Technical University of Catalonia - UPC). “JADE Tutorial. Application-defined content languages and ontologies.”
6. The MathWorks. Fuzzy Logic Toolbox for Use with MATLAB®. User’s Guide.
7. K.Craig, S.Awtar. “Inverted Pendulum Systems: Rotary and Arm-Driven. A Mechatronic System Design Case Study”, Department of Mechanical Engineering, Aeronautical Engineering and Mechanics Rensselaer Polytechnic Institute, Troy, NY12180, USA.
8. B.Sc.Jie Mao “Reduction of the Quantization Error in Fuzzy Logic Controllers by Dithering”, Ottawa-Carleton Institute for Electrical and Cornputer Engineering School of Information Technology and Engineering, Faculty of Engineering - University of Ottawa. November 1998.
9. P.Moraitis, N.Spanoudakis. “Combining Gaia and JADE for Multi-Agent Systems Development”.
10. V.Loia and S.Sessa (Eds.) “Soft Computing Agents: New Trends for Designing Autonomous Systems”, Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer, Vol. 75, 2001.
11. M.Wooldridge, N.R.Jennings, D.Kinny. “The Gaia Methodology fro Agent-Oriented Analysis and Design”. JAAMAS Vol. 3. No. 3, 2000, pp. 285–312.
12. P.Moraitis, E.Petraki, N.Spanoudakis. “Engineering JADE Agents with the Gaia Methodology”. In R. Kowalszyk, et al. (eds) “Agents Technologies, Infrastructures, Tools, and Applications for e-Services”, LNAI 2592, Springer-Verlag, 2003, pp. 77–91.
13. P.Moraitis and N.Spanoudakis. “Combining Gaia and JADE for Multi-Agent Systems Design”. 4th Int. Symp. ”From Agent Theory to Agent Implementation” (AT2AI4), in: Proceedings of the 17th European Meeting on Cybernetics and Systems Research (EMCSR 2004), Vienna, Austria, April 13-16, 2004.
14. D.Lea, “Cristopher Alexander: An Introduction to Object Oriented Designers”, http://gee.cs.oswego.edu/dl/ca/ca/ca.html.
15. NRC FuzzyJ Toolkit National Research Council of Canada’s Institute for Information Technology
16. JADE (Java Agent DEvelopment Framework) (TILAB, formerly CSELT), - http://jade.cselt.it/.

# An Overview on the Approximation Quality Based on Rough-Fuzzy Hybrids

**Van-Nam Huynh, Tu-Bao Ho and Yoshiteru Nakamori**

**Abstract** The so-called measure of approximation quality plays an important role in many applications of rough set based data analysis. In this chapter, we provide an overview on various extensions of approximation quality based on rough-fuzzy and fuzzy-rough sets, along with highlighting their potential applications as well as future directions for research in the topic.

## 1 Introduction

After nearly twenty years since the introduction of fuzzy sets theory [51], Pawlak [33] introduced the notion of a rough set as a new mathematical tool to deal with the approximation of a concept in the context of incomplete information. Basically, while a fuzzy set models the ill-definition of the boundary of a concept often described linguistically, a rough set characterizes a concept by its lower and upper approximations due to indiscernibility between objects arose because of incompleteness of available knowledge. Since its inception, the rough set theory has been proven to be of substantial importance in many areas of application [34, 39, 45].

During the last decades, many attempts to establish the relationships between the two theories, to compare each to the other, and to simultaneously hybridize them have been made, e.g., [10, 30, 31, 35, 40, 46, 47, 49, 50]. Among these lines of research, rough fuzzy hybridization has emerged as a promising new paradigm for decision-making related applications [17, 18, 31, 32], data analysis [22, 25] and many others. This is due to rough-fuzzy hybrids can encapsulate two distinct aspects of imperfection of knowledge being vagueness and indiscernibility, which may simultaneously occur in many situations of practical application [10].

On the other hand, one of issues of great practical importance in data analysis is discovering dependencies between attributes in datasets. In rough set theory, the notion of **approximation quality** (also called **degree of dependency**) is often used to evaluate the classification success of attributes in terms of a numerical evaluation of the dependency properties generated by these attributes. Particularly, it has been used as a useful tool, for instance, for discovering data dependencies and for semantics-preserving feature reduction using only the given data without any additional information as required by other theories [13, 25, 34]. This chapter aims at providing an overview on various extensions of approximation quality based on rough-fuzzy hybrids, along with highlighting their potential applications and future directions for research in the topic as well.

The structure of the rest of this chapter is as follows. Section 2 briefly introduces necessary notions of fuzzy sets and rough sets. Section 3 recalls Pawlak's notions of approximation quality and significance of attributes. In Sect. 4, the notions of rough fuzzy sets and fuzzy rough sets are reviewed in relation to their applications in practice. Sect. 5 devotes to an overview on rough-fuzzy hybrids based extensions of approximation quality, accompanying with illustrative examples. Finally, some concluding remarks and future work are presented in Sect. 6.

## 2 Basic of Rough Sets and Fuzzy Sets

In this section we briefly recall basic notions of fuzzy sets and rough sets. For the purpose of this paper, it is sufficient to consider the finite version of universes of discourse.

### *2.1 Fuzzy Sets*

Let $\mathbb{U}$ be a finite and non-empty set called universe of discourse. A fuzzy set $F$ of $\mathbb{U}$ is a mapping $\mu_F : \mathbb{U} \longrightarrow [0, 1]$, where for each $x \in \mathbb{U}$ we call $\mu_F(x)$ the membership degree of $x$ in $F$.

Given a number $\alpha \in (0, 1]$, the $\alpha$-cut, or $\alpha$-level set, of $F$ is defined as follows

$$F_\alpha = \{x \in \mathbb{U} | \mu_F(x) \geq \alpha\}$$

which is a subset of $\mathbb{U}$. Let us denote $\text{rng}(\mu_F) = \mu_F(\mathbb{U}) \setminus \{0\}$ and assume that $\text{rng}(\mu_F) = \{\alpha_1, \ldots, \alpha_n\}$, where $\alpha_i > \alpha_{i+1}$, for $i = 1, \ldots, n-1$. Then the membership function $\mu_F$ can be expressed as [12]

$$\mu_F(x) = \sum_{x \in F_{\alpha_i}} (\alpha_i - \alpha_{i+1}) \tag{1}$$

Clearly, $\alpha_1 = 1$ if $F$ is normal, i.e. $\exists x$ such that $\mu_F(x) = 1$. This representation of a fuzzy set is considered as providing a probability based semantics for membership function of fuzzy sets, where $m_i = (\alpha_i - \alpha_{i+1})$, with $\alpha_{n+1} = 0$ by convention, can be viewed as the probability that $F_{\alpha_i}$ stands as a crisp representative of $F$. Then $\{(F_{\alpha_i}, m_i) | i = 1, \ldots, n\}$ is usually referred to as a finitely discrete **consonant random set**, or **body of evidence** [41]. Note that the normalization assumption of $F$ insures the body of evidence does not contain the empty set. This view of fuzzy sets has been also used in [2] to introduce the so-called **mass assignment** of a fuzzy set, with relaxing the normalization assumption of fuzzy sets. Namely, the mass assignment of $F$, denoted by $m_F$, is a probability distribution on $2^{\mathbb{U}}$ defined by

$$\begin{aligned} m_F(\emptyset) &= 1 - \alpha_1, \\ m_F(F_{\alpha_i}) &= m_i, \text{ for } i = 1, \ldots, n. \end{aligned} \tag{2}$$

## 2.2 Rough Sets

Pawlak's theory of rough sets begins with the notion of an approximation space, which is a pair $\langle \mathbb{U}, R \rangle$, where $\mathbb{U}$ is a non-empty set (the universe of discourse) and $R$ an equivalence relation on $U$, i.e., $R$ is reflexive, symmetric, and transitive. The relation $R$ decomposes the set $\mathbb{U}$ into disjoint classes in such a way that two elements $x, y$ are in the same class iff $(x, y) \in R$. If two elements $x, y$ in $\mathbb{U}$ belong to the same equivalence class, we say that $x$ and $y$ are indistinguishable. For $X \in 2^{\mathbb{U}}$, in general it may not be possible to describe $X$ precisely in $\langle \mathbb{U}, R \rangle$. One may then characterize $X$ by a pair of lower and upper approximations defined as follows [33]

$$\underline{R}(X) = \{x \in \mathbb{U} | [x]_R \subseteq X\}; \qquad \overline{R}(X) = \{x \in \mathbb{U} | [x]_R \cap X \neq \emptyset\}$$

where $[x]_R$ stands for the equivalence class of $x$ by $R$. The pair $(\underline{R}(X), \overline{R}(X))$ is the representation of an ordinary set $X$ in the approximation space $\langle U, R \rangle$ or simply called the rough set of $X$.

In the context of rough set based data analysis, the equivalence relation in an approximation space is often interpreted via the notion of information systems. An **information system** $\mathcal{I}$ is a pair $\mathcal{I} = \langle \mathbb{U}, \mathcal{A} \rangle$, where $\mathbb{U}$ is a set of objects, $\mathcal{A}$ is a set of attributes, and each attribute $a \in \mathcal{A}$ associated with the set of attribute values $V_a$ is understood as a mapping $a : \mathbb{U} \to V_a$. An information system is called a **decision system** if assuming that the set of attributes $\mathcal{A} = \mathcal{C} \cup \mathcal{D}$ and $\mathcal{C} \cap \mathcal{D} \neq \emptyset$, where $\mathcal{C}$ is the set of **conditional attributes** and $\mathcal{D}$ is the set of **decision attributes**. Given an information system $\mathcal{I}$, each subset $P$ of the attribute set $\mathcal{A}$ induces an equivalence relation $\text{IND}(P)$ called *$P$-indiscernibility relation* as follows

$$\text{IND}(P) = \{(x, y) \in \mathbb{U}^2 | a(x) = a(y), \text{ for all } a \in P\}$$

and $\text{IND}(P) = \bigcap_{a \in P} \text{IND}(\{a\})$. If $(x, y) \in \text{IND}(P)$ we then say that objects $x$ and $y$ are indiscernible with respect to attributes in $P$. In other words, we cannot distinguish $x$ from $y$, and vice versa, in terms of attributes in $P$. Note that the partition of $\mathbb{U}$ generated by $\text{IND}(P)$, denoted by $\mathbb{U}/\text{IND}(P)$, can be calculated in terms of those partitions generated by single attributes in $P$ as follows [24]

$$\mathbb{U}/\text{IND}(P) = \bigotimes_{a \in P} \mathbb{U}/\text{IND}(\{a\}) \tag{3}$$

where

$$\mathcal{X} \otimes \mathcal{Y} = \{X \cap Y | X \in \mathcal{X}, Y \in \mathcal{Y}, X \cap Y \neq \emptyset\}$$

For simplicity of notation, from now on we use the same notation $P$ to denote the equivalence relation induced from a set $P$ of attributes, instead of IND$(P)$.

## 3 Pawlak's Approximation Quality

As mentioned in [13], one of the strengths of rough set theory is the fact that all its parameters are directly obtained from the given data. That is, in rough set theory the numerical value of imprecision is calculated by making use of the granularity structure of the data only, while other uncertainty theories like Dempster-Shafer theory [41] or fuzzy set theory [26] require probability assignments and membership values respectively.

In [34], Pawlak firstly introduces two numerical characterizations of imprecision of a subset $X$ in the approximation space $\langle \mathbb{U}, P \rangle$: *accuracy* and *roughness*. Accuracy of $X$, denoted by $\alpha_P(X)$, is simply the ratio of the number of objects in its lower approximation to that in its upper approximation; namely

$$\alpha_P(X) = \frac{|\underline{P}(X)|}{|\overline{P}(X)|} \tag{4}$$

where $|\cdot|$ denotes the cardinality of a set. Then the roughness of $X$, denoted by $\rho_P(X)$, is defined by subtracting the accuracy from 1 as

$$\rho_P(X) = 1 - \alpha_P(X) = 1 - \frac{|\underline{P}(X)|}{|\overline{P}(X)|} \tag{5}$$

Note that the lower is the roughness of a subset, the better is its approximation. In [48], Yao has interpreted Pawlak's accuracy measure in terms of a classic distance measure based on sets, called *Marczewski-Steinhaus (MS) metric* [27], which is defined by

$$D_{MS}(X, Y) = \frac{|X \cup Y| - |X \cap Y|}{|X \cup Y|} = 1 - \frac{|X \cap Y|}{|X \cup Y|}$$

Using MS metric, the roughness measure of a set $X$ in $\langle \mathbb{U}, P \rangle$ is the distance between its lower and upper approximations.

Suppose now that two views of universe $\mathbb{U}$ are given, which may come from two different subsets $P$ and $Q$ of attributes, by means of associated equivalence relations. Then an interesting question arises to be how well the knowledge from one view can be expressed by that from the other. In other words, we are concerned here with the issue of measuring dependencies between attributes. This issue is very important in many tasks of data analysis. In rough set theory, the so-called **approximation quality measure** $\gamma$ [34] is often used for such a situation to describe the degree of partial dependency between attributes.

Particularly, let $P$ and $Q$ be equivalence relations over $\mathbb{U}$, then the approximation quality of $Q$ by $P$, also called *degree of dependency*, is defined by

$$\gamma_P(Q) = \frac{|\mathrm{POS}_P(Q)|}{|\mathbb{U}|} \tag{6}$$

where

$$\mathrm{POS}_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P}(X) \tag{7}$$

is called the **positive region** of the partition $\mathbb{U}/Q$ with respect to $P$. We then say that $Q$ depends on $P$ in a degree $k = \gamma_P(Q)$ $(0 \leq k \leq 1)$ and denote as $P \Rightarrow_k Q$. If $k = 1$, $Q$ totally depends on $P$; if $0 < k < 1$, $Q$ partially (or roughly) depends on $P$, and if $k = 0$, $Q$ is totally independent from $P$.

Note that the approximation quality $\gamma_P(Q)$ can be also represented in terms of accuracy as follows

$$\gamma_P(Q) = \sum_{X \in \mathbb{U}/Q} \frac{|\overline{P}(X)|}{|\mathbb{U}|} \alpha_P(X) \tag{8}$$

Then, $\gamma_P(Q)$ is regarded as the weighted mean of the accuracies of approximation of sets $X \in \mathbb{U}/Q$ by $P$ [13].

Another issue of great practical importance is that of identifying how significant a specific attribute (or a group of attributes) is in respect of the classification power. This information is captured by calculating the change in dependency when an attribute is removed from the set of considered conditional attributes. In particularly, we can measure the significance of an attribute $a \in P$ with respect to the classification induced from $Q$ by the difference

$$\sigma_P(Q, a) = \gamma_P(Q) - \gamma_{P \setminus \{a\}}(Q) \tag{9}$$

This measure expresses how influence on the quality of approximation if we drop the attribute $a$ from $P$. The higher the change in dependency, the more significant the attribute is. If the significance is 0, the attribute is dispensable. A subset $S$ of $P$ is called a $Q$-reduct of $P$ (or a reduct of $P$ with respect to $Q$) if $\gamma_S(Q) = \gamma_P(Q)$.

In [13], the authors have also used the MS metric to re-interpret the rough approximation quality $\gamma$ and ascertain its statistical significance. The approximation quality measure and its extended variants have been extensively studied and used in many applications, especially in feature selection, e.g., [4, 9, 22, 23, 24, 25, 43, 44] and ranking problems, e.g., [14, 15, 16].

# 4 Rough-Fuzzy Hybrids

As argued by Dubois and Prade [10], rough sets and fuzzy sets capture two distinct aspects of imperfection of knowledge: indiscernibility and vagueness, that may be simultaneously present in a given application. Therefore, it is necessary to find out hybrid models which combines these notions for knowledge representation and integration in such situations. Among many possibilities for rough-fuzzy hybridization, the most typical ones are to fuzzify sets to be approximated and/or to fuzzify the equivalence relation in an approximation space [10, 11]. The first case allows to obtain rough approximations of fuzzy sets which results in the so-called **rough fuzzy sets**; while the second case allows to obtain approximations of (fuzzy) sets by means of fuzzy similarity relations resulting in the so-called **fuzzy rough sets**.

## *4.1 Rough Fuzzy Sets*

Given an approximation space $\langle \mathbb{U}, P \rangle$. Let $F$ be a fuzzy set in $\mathbb{U}$ with the membership function $\mu_F$. The upper and lower approximations $\overline{P}(F)$ and $\underline{P}(F)$ of $F$ by $P$ are fuzzy sets in the quotient set $\mathbb{U}/P$ with membership functions defined by, for each $F_i \in \mathbb{U}/P$,

$$\mu_{\overline{P}(F)}(F_i) = \sup_{x \in F_i} \{\mu_F(x)\}$$

$$\mu_{\underline{P}(F)}(F_i) = \inf_{x \in F_i} \{\mu_F(x)\}$$

The pair $(\underline{P}(F), \overline{P}(F))$ is then called a rough fuzzy set [11].

Furthermore, the rough fuzzy set $(\underline{P}(F), \overline{P}(F))$ naturally induces two fuzzy sets $P^*(F)$ and $P_*(F)$ in $\mathbb{U}$ with membership functions are defined respectively as follows

$$\mu_{P^*(F)}(x) = \mu_{\overline{P}(F)}([x]_P) \text{ and } \mu_{P_*(F)}(x) = \mu_{\underline{P}(F)}([x]_P) \tag{10}$$

That is, $P^*(F)$ and $P_*(F)$ are fuzzy sets with constant membership degree on the equivalence classes of $\mathbb{U}$ by $P$, and for any $x \in \mathbb{U}$, $\mu_{P^*(F)}(x)$ (respectively, $\mu_{P_*(F)}(x)$) can be viewed as the degree to which $x$ possibly (respectively, definitely) belongs to the fuzzy set $F$ [3]. Conceptually, the pair $(P_*(F), P^*(F))$ can be viewed as "extension" of rough fuzzy set $(\underline{P}(F), \overline{P}(F))$.

Rough fuzzy sets could find many applications in practical situations where a fuzzy classification or a fuzzy concept must be approximated by available knowledge expressed in terms of a Pawlak's approximation space, for instance as in pattern recognition and image analysis problems [1, 3, 5, 6, 7, 36, 37, 38, 42].

## 4.2 Fuzzy Rough Sets

Let us consider another extension of rough sets corresponding to the second case mentioned above. In this extension, instead of equipping the universe $\mathbb{U}$ with an equivalence relation $P$, we consider a fuzzy similarity relation $R$, i.e., a fuzzy set $R$ of $\mathbb{U}^2$, such that the properties of reflexivity ($\mu_R(x, x) = 1$), symmetry ($\mu_R(x, y) = \mu_R(y, x)$), and $\wedge$-transitivity of the form

$$\mu_R(x, z) \geq \mu_R(x, y) \wedge \mu_R(y, z)$$

are holded [52]. In order to define fuzzy rough approximation operators, the counterpart of equivalence classes called fuzzy equivalence classes must be defined first. According to Zadeh [52], the fuzzy equivalence class $[x]_R$ of objects close to $x$ is defined by

$$\mu_{[x]_R}(y) = \mu_R(x, y), \forall y \in \mathbb{U} \tag{11}$$

Interestingly, this definition degenerates to the usual definition of equivalence classes when $R$ is a non-fuzzy relation. Furthermore, Höhle [19] also proposed a definition of what should be a fuzzy equivalence class $X$ by means of the following axioms

(i) $\mu_X$ is normalized, i.e. $\exists x, \mu_X(x) = 1$,
(ii) $\mu_X(x) \wedge \mu_R(x, y) \leq \mu_X(y)$,
(iii) $\mu_X(x) \wedge \mu_X(y) \leq \mu_R(x, y)$.

Then, according to [10], a fuzzy set $[x]_R$ as in (11) is a fuzzy equivalence class in the sense of Höhle.

The family of fuzzy equivalence classes $\{[x]_R | x \in \mathbb{U}\}$, also denoted by $\mathbb{U}/R$, forms a "fuzzy partition" of $\mathbb{U}$. Also, a more direct way is to define a family $\mathcal{F} = \{F_1, \ldots, F_n\}$ of normal fuzzy sets of $\mathbb{U}$, with $m < |\mathbb{U}|$, which covers $\mathbb{U}$ sufficiently in the following sense

$$\inf_{x \in \mathbb{U}} \max_i \mu_{F_i}(x) > 0$$

Further, a disjointness property between $F_i$'s can be requested as

$$\forall i, j, \sup_{x \in \mathbb{U}} \min\{\mu_{F_i}(x), \mu_{F_j}(x)\} < 1$$

In the literature, a stronger restriction is often adopted

$$\sum_{i=1}^{n} \mu_{F_i}(x) = 1 \tag{12}$$

for any $x \in \mathbb{U}$. Then $\mathcal{F}$ plays the role of the family of fuzzy equivalence classes induced from a similarity relation $R$, i.e., $\mathcal{F} = \mathbb{U}/R$.

Given a fuzzy approximate space $\langle \mathbb{U}, R \rangle$, a fuzzy set $F$ can be approximated by means of the fuzzy partition $\mathbb{U}/R$ in terms of an $R$-upper and an $R$-lower approximation $\overline{R}(F)$ and $\underline{R}(F)$ as follows [10]

$$\mu_{\overline{R}(F)}(F_i) = \sup_{x \in \mathbb{U}} \min\{\mu_{F_i}(x), \mu_F(x)\} \tag{13}$$

$$\mu_{\underline{R}(F)}(F_i) = \inf_{x \in \mathbb{U}} \max\{1 - \mu_{F_i}(x), \mu_F(x)\} \tag{14}$$

for any $F_i \in \mathbb{U}/R$. The pair $(\underline{R}(F), \overline{R}(F))$ is then called a fuzzy rough set. When $F_i$'s are crisp, i.e., $R$ is an equivalence relation, we obtain the rough approximation of $F$ which results in a rough fuzzy set defined previously.

As noted in [24], these definitions given in (13)–(14) differ a little from the crisp rough approximations, as the memberships of individual objects to the approximations are not explicitly available. As a result of this, fuzzy rough approximations are redefined as fuzzy sets of $\mathbb{U}$ [24] by

$$\mu_{\overline{R}(F)}(x) = \sup_{F_i \in \mathbb{U}/R} \min\left(\mu_{F_i}(x), \sup_{y \in \mathbb{U}} \min\{\mu_{F_i}(y), \mu_F(y)\}\right) \tag{15}$$

$$\mu_{\underline{R}(F)}(x) = \sup_{F_i \in \mathbb{U}/R} \min\left(\mu_{F_i}(x), \inf_{y \in \mathbb{U}} \max\{1 - \mu_{F_i}(y), \mu_F(y)\}\right) \tag{16}$$

These definitions have been often used in application of fuzzy rough sets to dimensionality reduction [22, 23, 24, 25, 44].

*Remark 1.* Note that (15)–(16) can be viewed as the "extension" of the fuzzy rough set $(\underline{R}(F), \overline{R}(F))$, which was defined in [10] making use of the knowledge of fuzzy similarity relation $R$ directly, instead of fuzzy equivalence classes induced by $R$. Particularly, according to Dubois and Prade [10], we have

$$\mu_{\overline{R}(F)}(x) = \sup_{y \in \mathbb{U}} \mu_F(x) * \mu_R(x, y) \tag{17}$$

$$\mu_{\underline{R}(F)}(x) = \inf_{y \in \mathbb{U}} \mu_R(x, y) \to \mu_F(y) \tag{18}$$

where $*$ is a $t$-norm and $\to$ is an $S$-implication operator. However, in practical applications of fuzzy rough sets in data analysis, the knowledge of fuzzy similarity relation $R$ may not be available, but a fuzzy linguistic partition of attribute domain which plays the role of the family of fuzzy equivalence classes is often pre-assumed. This practically explains why (15)–(16) is often used in application.

For a more general and comprehensive treatment of fuzzy rough sets, the readers can refer, e.g., to [10, 11, 40, 49].

# 5 Approximation Quality Based on Rough-Fuzzy Sets

As we have mentioned previously, rough fuzzy sets arise naturally when we want to approximate a fuzzy set or a fuzzy classification by means of the available knowledge expressed in terms of an approximation space $\langle \mathbb{U}, P \rangle$.

The first case may often occur in, for example, problems of image analysis, where $\mathbb{U}$ denotes a gray image or feature space and $\mathbb{U}/P$ is a partition of $\mathbb{U}$, a fuzzy set $F$ can be viewed to represent ill-defined pattern classes or some imprecise image property such as brightness, darkness, smoothness, etc [3, 7]. In such a situation, roughness (or accuracy) of a fuzzy set $F$ may be used to provide the information of how well its approximation is in $\langle \mathbb{U}, P \rangle$. Regarding to this, Banerjee and Pal [3] have proposed a roughness measure for fuzzy sets and have discussed the issue of how to use this measure in tasks of image analysis.

The second case may come up in a natural way when a linguistic classification must be expressed by means of already existing knowledge $P$. For example, let us consider two attributes "experience" and "salary" in a database of employees. Then the attribute "experience" may take values in a finite set of labels such as **good**, **poor**, **very good**, etc., and the attribute "salary" may have numerical values. Then it is natural to intuitively infer a "partial" dependence between "experience" and "salary" as (the better the experience, the higher the salary). However, such a dependency could not be expressed in terms of traditional data dependencies, because there may be different employees having the same value of "experience" but different salaries, even in small magnitude. Therefore, it is necessary and useful to look for measures such as the approximation quality that may support us as numerical characteristics to realize partial dependency between attributes in such situations.

## *5.1 Roughness of a Fuzzy Set*

### Banerjee and Pal's Approach

In [3], Banerjee and Pal have proposed a roughness measure for fuzzy sets in a given approximation space. Essentially, this measure of roughness of a fuzzy set depends on parameters that are designed as thresholds of definiteness and possibility in membership of the objects in $\mathbb{U}$ to the fuzzy set.

More explicitly, let us be given an approximation space $\langle \mathbb{U}, P \rangle$ and a fuzzy set $F$ in $\mathbb{U}$. We now consider parameters $\alpha, \beta$ such that $0 < \beta \leq \alpha \leq 1$. The $\alpha$-cut $P_*(F)_\alpha$ and $\beta$-cut $P^*(F)_\beta$ of fuzzy sets $P_*(F)$ and $P^*(F)$, respectively, are called to be the $\alpha$-*lower approximation* and the $\beta$-*upper approximation* of $F$ in $\langle \mathbb{U}, P \rangle$, respectively. Then a roughness measure of the fuzzy set $F$ with respect to parameters $\alpha, \beta$, with $0 < \beta \leq \alpha \leq 1$, and the approximation space $\langle \mathbb{U}, P \rangle$ is defined by

$$\rho_P^{\alpha,\beta}(F) = 1 - \frac{|P_*(F)_\alpha|}{|P^*(F)_\beta|}$$

By the assumption made on parameters, we have

1. $0 \leq \rho_P^{\alpha,\beta}(F) \leq 1$.
2. If $F$ is a fuzzy set such that there is a member $x$ in each equivalence class of $\mathbb{U}/P$ with $\mu_F(x) < \alpha$, then $\rho_P^{\alpha,\beta}(F) = 1$.
3. If $F$ is a definable fuzzy set, i.e., $\mu_F$ is a constant function on each equivalence class of $\mathbb{U}/P$ and $\alpha = \beta$, then $\rho_P^{\alpha,\beta}(F) = 0$.

Note that while the third statement seems interesting as it says that the measure $\rho_P^{\alpha,\beta}(\cdot)$ inherits a property of Pawlak's roughness measure, the second one may not be well-justified. Furthermore, the following property of $\rho_P^{\alpha,\beta}(\cdot)$ proved in [3] may be also undesired, unless the support of a constant fuzzy set, i.e. its strong 0-cut, is definable in the approximation space.

**Proposition 1.** *If $F$ is a constant fuzzy set, say $\mu_F(x) = \delta$, for all $x \in \mathbb{U}$, then $\rho_P^{\alpha,\beta}(F) = 0$, with the exception when $\beta < \delta < \alpha$, in which $\rho_P^{\alpha,\beta}(F) = 1$.*

Properties of the measure $\rho_P^{\alpha,\beta}(\cdot)$ and its potential applications in the field of pattern recognition have been reported and mentioned in [3], and more recently in [53].

### An Alternative Approach

In [20], the authors have introduced a parameter-free measure of roughness of a fuzzy set that in fact is a generalization of Pawlak's notion of roughness measure and avoids the undesirable properties held by Banerjee and Pal's roughness measure as mentioned above. Basically, this approach is based on the random set based representation of a fuzzy set and defines its roughness as the weighted mean of roughness measures of its crisp representatives.

In particularly, let $\text{rng}(\mu_F)$ and $m_F$ be the range of the membership function $\mu_F$ and the mass assignment of $F$, respectively. Recall that in this representation of fuzzy set $F$, for each $\alpha \in \text{rng}(\mu_F)$, $m_F(F_\alpha)$ is viewed as the probability that $F_\alpha$ stands as a crisp representative of $F$. Under such a representation, the roughness measure of $F$ with respect to the approximation space $\langle \mathbb{U}, P \rangle$ is defined as follows

$$\hat{\rho}_P(F) = \sum_{\alpha \in \text{rng}(\mu_F)} m_F(F_\alpha)\Big(1 - \frac{|\underline{P}(F_\alpha)|}{|\overline{P}(F_\alpha)|}\Big) \equiv \sum_{\alpha \in \text{rng}(\mu_F)} m_F(F_\alpha)\rho_P(F_\alpha) \qquad (19)$$

*Remark 2.* With this definition of roughness, we have

- $0 \leq \hat{\rho}_P(F) \leq 1$.
- $\hat{\rho}_P(\cdot)$ is a natural extension of Pawlak's roughness measure for fuzzy sets, i.e., if $F$ is a crisp subset of $U$ then $\hat{\rho}_P(F) = \rho_P(F)$.
- $F$ is a definable fuzzy set, i.e., if $\underline{P}(F) = \overline{P}(F)$, if and only if $\hat{\rho}_P(F) = 0$.

**Table 1** The approximations of the fuzzy set $\mu_{small}$

| | {0,2,4} | {1,3,5} | {6,8,10} | {7,9} |
|---|---|---|---|---|
| $\mu_{small_*}$ | 0.25 | 0 | 0 | 0 |
| $\mu_{small^*}$ | 1 | 1 | 0 | 0 |

Let us consider a simple example depicting the introduced notions.

*Example 1.* Suppose we are given an approximation space $\langle \mathbb{U}, P\rangle$, where $U = \{0, 1, 2, \ldots, 10\}$ and $P$ is such that

$$\mathbb{U}/P = \{\{0, 2, 4\}, \{1, 3, 5\}, \{6, 8, 10\}, \{7, 9\}\}$$

Let us consider a linguistic value *small* whose membership function is defined by

| $u$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mu_{small}(u)$ | 1 | 1 | 0.75 | 0.5 | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 |

The approximations of the fuzzy set $\mu_{small}$ in $\langle \mathbb{U}, P\rangle$ are given in Table 1. Then we obtain the mass assignment for the linguistic value *small*, and approximations of its focal sets given in Table 2.

Using Banerjee and Pal's notion, we obtain

$$\rho_P^{\alpha,\beta}(small) = \begin{cases} 1 & \text{for } \alpha > 0.25 \\ 0.5 & \text{for } 0.25 \geq \alpha > 0 \end{cases}$$

where the constraint $\alpha \geq \beta > 0$ is always assumed. On the other hand, the roughness by (19) yields

$$\hat{\rho}_P(small) = \sum_{\alpha \in \mathrm{rng}(\mu_{small})} m_{small}(small_\alpha)\left(1 - \frac{|\underline{P}(small_\alpha)|}{|\overline{P}(small_\alpha)|}\right) = 0.875$$

Let $P^*(F)$ and $P_*(F)$ be fuzzy sets of $\mathbb{U}$ induced from the rough fuzzy set $(\underline{P}(F), \overline{P}(F))$ as in preceding section. Denote

**Table 2** Mass assignment for *small* and approximations of its focal sets

| $\mathrm{rng}(\mu_{small})$ | 1 | 0.75 | 0.5 | 0.25 |
|---|---|---|---|---|
| $small_\alpha$ | {0, 1} | {0, 1, 2} | {0, 1, 2, 3} | {0, 1, 2, 3, 4} |
| $m_{small}(small_\alpha)$ | 0.25 | 0.25 | 0.25 | 0.25 |
| $\underline{P}(small_\alpha)$ | ∅ | ∅ | ∅ | {0, 2, 4} |
| $\overline{P}(small_\alpha)$ | {0, 1, 2, 3, 4, 5} | {0, 1, 2, 3, 4, 5} | {0, 1, 2, 3, 4, 5} | {0, 1, 2, 3, 4, 5} |

$$\text{rng}(\mu_{P_*(F)}) \cup \text{rng}(\mu_{P^*(F)}) = \{\omega_1, \ldots, \omega_p\}$$

such that $\omega_i > \omega_{i+1} > 0$ for $i = 1, \ldots, p-1$. Obviously, $\{\omega_1, \ldots, \omega_p\} \subseteq \text{rng}(\mu_F)$. With these notations, the following holds [20]

**Lemma 1.** *For any* $1 \leq j \leq p$*, if there exists* $\alpha_i, \alpha_{i'} \in \text{rng}(\mu_F)$ *such that* $\omega_{j+1} < \alpha_i < \alpha_{i'} \leq \omega_j$ *then we have* $F_{\alpha_i} \approx_P F_{\alpha'_i}$*, i.e.* $\underline{P}(F_{\alpha_i}) = \underline{P}(F_{\alpha'_i})$ *and* $\overline{P}(F_{\alpha_i}) = \overline{P}(F_{\alpha'_i})$*, and so* $\rho_R(F_{\alpha_i}) = \rho_R(F_{\alpha'_i})$.

Further, the following lemma is due to Dubois and Prade [10]

**Lemma 2.** *For any* $\alpha \in (0, 1]$*, we have*

$$P^*(F)_\alpha = \overline{P}(F_\alpha) \text{ and } P_*(F)_\alpha = \underline{P}(F_\alpha)$$

It then follows from Lemmas 1 and 2 that $\hat{\rho}_P(F)$ can be represented in terms of level sets of fuzzy sets $P_*(F)$ and $P^*(F)$ as the following proposition shows.

**Proposition 2.** $\hat{\rho}_P(F) = \sum_{j=1}^{p} (\omega_j - \omega_{j+1})(1 - \frac{|P_*(F)_{\omega_j}|}{|P^*(F)_{\omega_j}|})$*, where* $\omega_{p+1} = 0$*, by convention.*

*Example 2.* Let us continue with the approximation space $\langle U, P\rangle$ and the fuzzy set *small* given in Example 1. We have

$$\text{rng}(\mu_{small}) = \{1, 0.75, 0.5, 0.25\}$$

By Table 1, we obtain

$$\text{rng}(\mu_{P_*(small)}) \cup \text{rng}(\mu_{P^*(small)}) = \{1, 0.25\}$$

which makes a partition of $\text{rng}(\mu_{small})$ as $\{\{1, 0.75, 0.5\}, \{0.25\}\}$. It is easily to see that Table 2 illustrates for Lemma 1, and by Proposition 2 we get

$$\hat{\rho}_R(small) = (1 - 0.25)(1 - \frac{P_*(small)_1}{P^*(small)_1}) + 0.25(1 - \frac{P_*(small)_{0.25}}{P^*(small)_{0.25}}) = 0.875$$

which coincides with that given in Example 1.

Similar to the case of roughness of a crisp set, we have also the following proposition [20].

**Proposition 3.** *If fuzzy sets F and G in U are roughly equal in* $\langle U, R\rangle$*, then we have* $\hat{\rho}_R(F) = \hat{\rho}_R(G)$.

### 5.2 Approximation Quality of a Fuzzy Classification

Let $P$ and $Q$ be two equivalence relations over universal set $\mathbb{U}$. As mentioned above, $P$ and $Q$ may be induced respectively by sets of attributes applied to objects in $\mathbb{U}$. Then the approximation quality $\gamma_P(Q)$ of $Q$ by $P$ defined by (6) can be rewritten as

$$\gamma_P(Q) = \frac{1}{|\mathbb{U}|} \sum_{X \in \mathbb{U}/Q} |\underline{P}(X)| \tag{20}$$

In [34], Pawlak also defines the so-called *approximation accuracy* of $Q$ by $P$, which extends the approximation accuracy of sets, by

$$\alpha_P(Q) = \frac{\sum_{X \in \mathbb{U}/Q} |\underline{P}(X)|}{\sum_{X \in \mathbb{U}/Q} |\overline{P}(X)|} \tag{21}$$

which is easily represented in terms of accuracies of sets as follows

$$\alpha_P(Q) = \sum_{X \in \mathbb{U}/Q} \frac{|\overline{P}(X)|}{\sum_{Y \in \mathbb{U}/Q} |\overline{P}(Y)|} \alpha_P(X)$$

That is, the approximation accuracy of a classification can be regarded as the convex sum of accuracies of its classes.

Furthermore, as mentioned in [34], the measure of approximation quality $\gamma_P(Q)$ does not capture how this partial dependency is actually distributed among classes of $\mathbb{U}/Q$. To capture this information we need the so-called *precision measure* $\pi_P(X)$, for $X \in \mathbb{U}/Q$, defined by

$$\pi_P(X) = \frac{|\underline{P}(X)|}{|X|} \tag{22}$$

Clearly, we have $\pi_R(X) \geq \alpha_P(X)$, for any $X \in \mathbb{U}/Q$. The two measures $\gamma_P(Q)$ and $\pi_P(X)$, $X \in \mathbb{U}/Q$, give us full information about the "classification power" of the knowledge $P$ with respect to the classification $\mathbb{U}/Q$.

Now let us consider a fuzzy classification $\widetilde{Q}$ of $\mathbb{U}$ instead of a crisp one $Q$, i.e., $\mathbb{U}/\widetilde{Q}$ is a fuzzy partition of $\mathbb{U}$. This situation may naturally occur when a linguistic classification must be approximated in terms of already existing knowledge $P$. For example, assume that we have a personnel database given as $\mathbb{D}$ = PERSONNEL[$ID$; $Name$; $Position$; $Salary$], and attribute **Position** induces an approximation space $\langle \mathbb{D}, \mathrm{IND}(Position) \rangle$. Given a linguistic description on the attribute **Salary**, say '$high$', it defines a fuzzy set on $\mathbb{D}$ denoted by $\mathbb{D}_{high}$. Then the accuracy of the fuzzy set $\mathbb{D}_{high}$, namely

$$\hat{\alpha}_{\mathrm{IND}(Position)}(\mathbb{D}_{high}) = 1 - \hat{\rho}_{\mathrm{IND}(Position)}(\mathbb{D}_{high})$$

may express the degree of completeness of our knowledge about the statement "Salary is $high$", given the granularity of $\mathbb{D}/\mathrm{IND}(Position)$. Further, a linguistic classification, say $\{low, medium, high\}$, may be imposed on the attribute **Salary** that induces a fuzzy partition of $\mathbb{D}$. Now one may want to measure a degree of dependency between "knowledge on attribute **Salary** expressed linguistically" and "knowledge on attribute **Position**".

In such a situation, guided by (20)–(21) and the random set based interpretation of a fuzzy set, the approximation quality and accuracy of a fuzzy classification $\widetilde{Q}$ by a crisp classification $P$ can be defined [20, 21] as

$$\hat{\gamma}_P(\widetilde{Q}) = \frac{1}{|\mathbb{U}|} \sum_{X \in \mathbb{U}/\widetilde{Q}} \sum_{\alpha \in \mathrm{rng}(\mu_X)} m_X(X_\alpha)|\underline{P}(X_\alpha)| \tag{23}$$

and

$$\hat{\alpha}_P(\widetilde{Q}) = \frac{\sum_{X \in \mathbb{U}/\widetilde{Q}} \sum_{\alpha \in \mathrm{rng}(\mu_X)} m_X(X_\alpha)|\underline{P}(X_\alpha)|}{\sum_{X \in \mathbb{U}/\widetilde{Q}} \sum_{\alpha \in \mathrm{rng}(\mu_X)} m_X(X_\alpha)|\overline{P}(X_\alpha)|} \tag{24}$$

respectively, where for $X \in \mathbb{U}/\widetilde{Q}$, $m_X$ stands for the mass assignment of $X$.

On the other hand, for each fuzzy class $X \in \mathbb{U}/\widetilde{Q}$, viewing $\underline{P}(X)$ as the induced fuzzy set $P_*(X)$ of $\mathbb{U}$ (refer to (10)) defined by

$$\mu_{P_*(X)}(x) = \mu_{\underline{P}(X)}([x]_P)$$

we can then define a counterpart of (7) for $\mathrm{POS}_P(\widetilde{Q})$ as a fuzzy set of $\mathbb{U}$ by

$$\mu_{\mathrm{POS}_P(\widetilde{Q})}(x) = \max_{X \in \mathbb{U}/\widetilde{Q}} \mu_{P_*(X)}(x) \tag{25}$$

Thus, guided by (6), another extension of the approximation quality can be also defined as

$$\hat{\gamma}'_P(\widetilde{Q}) = \frac{|\mathrm{POS}_P(\widetilde{Q})|}{|\mathbb{U}|} = \frac{\sum_{x \in \mathbb{U}} \mu_{\mathrm{POS}_P(\widetilde{Q})}(x)}{|\mathbb{U}|} \tag{26}$$

Similarly, rewriting (21) as

$$\alpha_P(Q) = \frac{|\bigcup_{X \in \mathbb{U}/Q} \underline{P}(X)|}{\sum_{X \in \mathbb{U}/Q} |\overline{P}(X)|}$$

suggests another extension of approximation accuracy of $\widetilde{Q}$ by $P$ defined by

$$\hat{\alpha}'_P(\widetilde{Q}) = \frac{|\mathrm{POS}_P(\widetilde{Q})|}{\sum\limits_{X\in\mathbb{U}/\widetilde{Q}} |\overline{P}(X)|} = \frac{\sum\limits_{x\in\mathbb{U}} \mu_{\mathrm{POS}_P(\widetilde{Q})}(x)}{\sum\limits_{X\in\mathbb{U}/\widetilde{Q}} \sum\limits_{\alpha\in\mathrm{rng}(\mu_X)} m_X(X_\alpha)|\overline{P}(X_\alpha)|} \tag{27}$$

It is worth noting [20] here that the approximation quality and accuracy of $\widetilde{Q}$ by $P$ defined by (23)–(24) can be respectively represented as

$$\hat{\gamma}_P(\widetilde{Q}) = \frac{1}{|\mathbb{U}|} \sum_{X\in\mathbb{U}/\widetilde{Q}} |P_*(X)| = \frac{1}{|\mathbb{U}|} \sum_{X\in\mathbb{U}/\widetilde{Q}} \sum_{x\in\mathbb{U}} \mu_{P_*(X)}(x) \tag{28}$$

$$\hat{\alpha}_P(\widetilde{Q}) = \frac{\sum\limits_{X\in\mathbb{U}/\widetilde{Q}} |P_*(X)|}{\sum\limits_{X\in\mathbb{U}/\widetilde{Q}} |P^*(X)|} = \frac{\sum\limits_{X\in\mathbb{U}/\widetilde{Q}} \sum\limits_{x\in\mathbb{U}} \mu_{P_*(X)}(x)}{\sum\limits_{X\in\mathbb{U}/\widetilde{Q}} \sum\limits_{x\in\mathbb{U}} \mu_{P^*(X)}(x)} \tag{29}$$

which interestingly turn out to be natural extensions of (20) and (21), respectively, for the crisp case.

Clearly, two different, but equivalent, representations of $\gamma_P(Q)$ and $\alpha_P(Q)$ lead to various different extensions in the fuzzy case. Therefore, the natural question arises is that what extension should be used in practice. Theoretically, it seems difficult to give a satisfactory answer to the question, however, an appropriate selection could be made on the basis of experimental evaluations as usual for a given application.

In the following we consider a simple example to illustrate discussed extensions.

*Example 3.* Let us consider a relation in a relational database as shown in Table 3 (this database is a variant of that found in [8]).

Let $P$ be the set of attributes **D** (degree) and **E** (experience). We then obtain an approximation space $\langle \mathbb{U}, P\rangle$, where $\mathbb{U} = \{1, \ldots, 16\}$, with the corresponding partition

**Table 3** Relation in a relational database

| ID | Degree | Experience | Salary | ID | Degree | Experience | Salary |
|---|---|---|---|---|---|---|---|
| 1 | Ph.D. | good | 63K | 9 | M.S. | poor | 41K |
| 2 | Ph.D. | very poor | 47K | 10 | M.S. | very good | 68K |
| 3 | M.S. | good | 53K | 11 | M.S. | good | 50K |
| 4 | B.S. | very poor | 26K | 12 | B.S. | very poor | 23K |
| 5 | B.S. | poor | 29K | 13 | M.S. | good | 55K |
| 6 | Ph.D. | very poor | 50K | 14 | M.S. | good | 51K |
| 7 | B.S. | poor | 35K | 15 | Ph.D. | good | 65K |
| 8 | M.S. | poor | 40K | 16 | M.S. | very good | 64K |

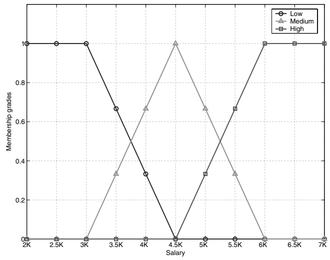



**Fig. 1** A linguistic partition of attribute *salary*

$$\mathbb{U}/P = \{\{1, 15\}, \{2, 6\}, \{3, 11, 13, 14\}, \{4, 12\}, \{5, 7\}, \{8, 9\}, \{10, 16\}\}$$

Further, consider now for example a linguistic classification over attribute **S** (salary), i.e. $\widetilde{Q} = \{\mathrm{S}\}$, with membership functions of linguistic classes **Low**, **Medium**, **High** graphically depicted as in Fig. 1. Then the linguistic classification induces a fuzzy partition $\mathbb{U}/\widetilde{Q}$ whose membership functions of fuzzy classes are shown in Table 4.

Then approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in the approximation space $\langle\mathbb{U}, P\rangle$ are given in Table 5.

Using (28) and (29) we obtain

$$\hat{\gamma}_P(\widetilde{Q}) = \frac{13.46}{16} = 0.84, \text{ and } \hat{\alpha}_P(\widetilde{Q}) = \frac{13.46}{18.21} = 0.739$$

respectively. That is, we have the following partial dependency in the database

$$\{\mathrm{D}, \mathrm{E}\} \Rightarrow_{0.84} \mathrm{S} \tag{30}$$

Note that making use of (26) and (27) gives us

**Table 4** Induced fuzzy partition of $\mathbb{U}$ based on *salary*

| $\mathbb{U}$ | $\mu_{Low}$ | $\mu_{Medium}$ | $\mu_{High}$ | $\mathbb{U}$ | $\mu_{Low}$ | $\mu_{Medium}$ | $\mu_{High}$ |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 9 | 0.27 | 0.73 | 0 |
| 2 | 0 | 0.87 | 0.13 | 10 | 0 | 0 | 1 |
| 3 | 0 | 0.47 | 0.53 | 11 | 0 | 0.67 | 0.33 |
| 4 | 1 | 0 | 0 | 12 | 1 | 0 | 0 |
| 5 | 1 | 0 | 0 | 13 | 0 | 0.33 | 0.67 |
| 6 | 0 | 0.67 | 0.33 | 14 | 0 | 0.6 | 0.4 |
| 7 | 0.67 | 0.33 | 0 | 15 | 0 | 0 | 1 |
| 8 | 0.33 | 0.67 | 0 | 16 | 0 | 0 | 1 |

**Table 5** The approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in $\langle \mathbb{U}, P \rangle$

| $X_i$ | {1, 15} | {2, 6} | {3, 11, 13, 14} | {4, 12} | {5, 7} | {8, 9} | {10, 16} |
|---|---|---|---|---|---|---|---|
| $\mu_{P_*(High)}$ | 1 | 0.13 | 0.33 | 0 | 0 | 0 | 1 |
| $\mu_{P^*(High)}$ | 1 | 0.33 | 0.67 | 0 | 0 | 0 | 1 |
| $\mu_{P_*(Medium)}$ | 0 | 0.67 | 0.33 | 0 | 0 | 0.67 | 0 |
| $\mu_{P^*(Medium)}$ | 0 | 0.87 | 0.67 | 0 | 0.33 | 0.73 | 0 |
| $\mu_{P_*(Low)}$ | 0 | 0 | 0 | 1 | 0.67 | 0.27 | 0 |
| $\mu_{P^*(Low)}$ | 0 | 0 | 0 | 1 | 1 | 0.33 | 0 |

$$\hat{\gamma}'_P(\widetilde{Q}) = \frac{11.34}{16} = 0.709, \text{ and } \hat{\alpha}'_P(\widetilde{Q}) = \frac{11.34}{13.88} = 0.82$$

Now in order to show how the influence of, for example, attribute **E** on the quality of approximation, let us consider the partition induced by the relation $R = P \setminus \{\text{E}\} = \{\text{D}\}$ as follows

$$U/R = \{\{1, 2, 6, 15\}, \{3, 8, 9, 10, 11, 13, 14, 16\}, \{4, 5, 7, 12\}\}$$

Then we obtain approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in the approximation space $\langle \mathbb{U}, R \rangle$ given in Table 6.

Thus we have

$$\hat{\gamma}_R(\widetilde{Q}) = \hat{\gamma}_{P\setminus\{\text{E}\}}(\widetilde{Q}) = \frac{3.2}{16} = 0.2$$

Similarly, we also easily obtain

$$\hat{\gamma}_{P\setminus\{\text{D}\}}(\widetilde{Q}) = \frac{5.06}{16} = 0.316$$

As we can see, both attributes **D** and **E** are highly significant because without each of them the approximation quality $\hat{\gamma}_P(\widetilde{Q})$ changes considerably.

**Table 6** The approximations of the fuzzy partition $\mathbb{U}/\widetilde{Q}$ in $\langle U, R \rangle$

| $X_i$ | {1, 2, 6, 15} | {3, 8, 9, 10, 11, 13, 14, 16} | {4, 5, 7, 12} |
|---|---|---|---|
| $\mu_{R_*(High)}$ | 0.13 | 0 | 0 |
| $\mu_{R^*(High)}$ | 1 | 1 | 0 |
| $\mu_{R_*(Medium)}$ | 0 | 0 | 0 |
| $\mu_{R^*(Medium)}$ | 0.87 | 0.73 | 0.33 |
| $\mu_{R_*(Low)}$ | 0 | 0 | 0.67 |
| $\mu_{R^*(Low)}$ | 0 | 0.33 | 1 |

## 6 Approximation Quality Based on Fuzzy-Rough Sets

Let us turn to a fuzzy approximation space $\langle \mathbb{U}, P \rangle$, where $P$ is a fuzzy similarity relation over universe $\mathbb{U}$. This fuzzy similarity relation induces a fuzzy partition over $\mathbb{U}$ denoted by $\mathbb{U}/P$ as mentioned previously. Assume now that $\mathbb{U}/Q$ is another (fuzzy) partition of $\mathbb{U}$. In order to have a counterpart of (6) for the approximation quality of $Q$ by $P$ in this situation, one needs to define the fuzzy positive region $\mathrm{POS}_P(Q)$ which is regarded as a fuzzy set of $\mathbb{U}$. Then, once having defined the fuzzy positive region, an extension of the approximation quality of $Q$ by $P$ can be defined [24, 28] as follows

$$\hat{\gamma}_P(Q) = \frac{|\mathrm{POS}_P(Q)|}{|\mathbb{U}|} = \frac{\sum_{x \in \mathbb{U}} \mu_{\mathrm{POS}_P(Q)}(x)}{|\mathbb{U}|} \tag{31}$$

where $\Sigma$-count is used for the cardinality of a fuzzy set.

In the case that the knowledge of $P$ is not given directly but, instead, a fuzzy partition $\mathbb{U}/P$ is predefined, Jensen and Shen [24, 25] have defined the membership function of fuzzy positive region $\mathrm{POS}_P(Q)$, for any object $x \in \mathbb{U}$, as

$$\mu_{\mathrm{POS}_P(Q)}(x) = \sup_{X \in \mathbb{U}/Q} \mu_{\underline{P}(X)}(x) \tag{32}$$

where the membership function $\mu_{\underline{P}(X)}(x)$ of fuzzy lower approximations can be defined by (16). Note that when $\mathbb{U}/P$ is a crisp partition, (31) is identical to (26) above. This approach has been successfully used for the task of feature reduction for crisp and real-valued datasets in various applications of data mining [22, 23, 24, 25, 44].

In particularly, regarding the issue of feature reduction in crisp and real-valued datasets, each real-valued attribute $a$ is first associated with a fuzzy linguistic partition denoted by $\mathbb{U}/\{a\}$, then the fuzzy partition $\mathbb{U}/P$ induced by a set $P$ of attributes defined over objects in $\mathbb{U}$ is defined as a fuzzy counterpart of (3) as follows

$$\mathbb{U}/P = \bigotimes_{a \in P} \mathbb{U}/\{a\} \tag{33}$$

where $t$-norm min is used for the fuzzy intersection. On the basis of these above extensions, a fuzzy-rough based method of attribute reduction described by the so-called fuzzy-rough QuickReduct algorithm has been proposed and applied to Web categorization in [24] and complex systems monitoring [25].

The following simple example taken from [24] will illustrate how these extensions work.

*Example 4.* Let us consider an example data set and fuzzy sets $N$ and $Z$ given in Fig. 2. Here, for illustrative simplicity, the fuzzy sets are viewed as fuzzy classes defined for all real-valued attributes.

Then we have the following partitions induced from corresponding individual attributes

| Object | a | b | c | q |
|---|---|---|---|---|
| 1 | -.4 | -.3 | -.5 | no |
| 2 | -.4 | .2 | -.1 | yes |
| 3 | -.3 | -.4 | -.3 | no |
| 4 | .3 | -.3 | 0 | yes |
| 5 | .2 | -.3 | 0 | yes |
| 6 | .2 | 0 | 0 | no |

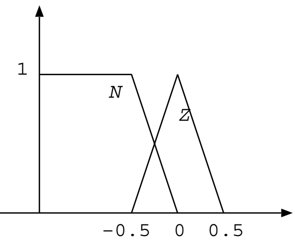



**Fig. 2** Data set and corresponding fuzzy sets

$$\mathbb{U}/A = \{N_a, Z_a\},\ \mathbb{U}/B = \{N_b, Z_b\},$$
$$\mathbb{U}/C = \{N_c, Z_c\},\ \mathbb{U}/Q = \{\{1,3,6\},\{2,4,5\}\},$$

where $A = \{a\}, B = \{b\}, C = \{c\}, Q = \{q\}$ and membership functions of corresponding fuzzy classes are given in Table 7.

The following fuzzy partitions induced from subsets of conditional attributes are obtained by (33)

$$\begin{aligned}
\mathbb{U}/\{a,b\} &= \{N_a \cap N_b, N_a \cap Z_b, Z_a \cap N_b, Z_a \cap Z_b\},\\
\mathbb{U}/\{b,c\} &= \{N_b \cap N_c, N_b \cap Z_c, Z_b \cap N_c, Z_b \cap Z_c\},\\
\mathbb{U}/\{a,c\} &= \{N_a \cap N_c, N_a \cap Z_c, Z_a \cap N_c, Z_a \cap Z_c\},\\
\mathbb{U}/\{a,b,c\} &= \{N_a \cap N_b \cap N_c, N_c \cap N_b \cap Z_c, N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c,\\
&\quad N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c, N_c \cap N_b \cap N_c\}
\end{aligned}$$

where $\cap = \min$. Using (16) and (31) respectively for calculating fuzzy lower approximations and the approximation quality, we obtain

$$\hat{\gamma}_A(Q) = \frac{2}{6}, \hat{\gamma}_B(Q) = \frac{2.4}{6}, \hat{\gamma}_C(Q) = \frac{1.6}{6}, \hat{\gamma}_{\{a,b\}}(Q) = \frac{3.4}{6}$$

$$\hat{\gamma}_{\{b,c\}}(Q) = \frac{3.2}{6}, \hat{\gamma}_{\{a,c\}}(Q) = \frac{3.2}{6}, \hat{\gamma}_{\{a,b,c\}}(Q) = \frac{3.4}{6}$$

**Table 7** Membership functions of corresponding fuzzy classes

| Object | $a$ | | $b$ | | $c$ | | $q$ | |
|---|---|---|---|---|---|---|---|---|
| | $N_a$ | $Z_a$ | $N_b$ | $Z_b$ | $N_c$ | $Z_c$ | $\{1,3,6\}$ | $\{2,4,5\}$ |
| 1 | 0.8 | 0.2 | 0.6 | 0.4 | 1.0 | 0.0 | 1.0 | 0.0 |
| 2 | 0.8 | 0.2 | 0.0 | 0.6 | 0.2 | 0.8 | 0.0 | 1.0 |
| 3 | 0.6 | 0.4 | 0.8 | 0.2 | 0.6 | 0.4 | 1.0 | 0.0 |
| 4 | 0.0 | 0.4 | 0.6 | 0.4 | 0.0 | 1.0 | 0.0 | 1.0 |
| 5 | 0.0 | 0.6 | 0.6 | 0.4 | 0.0 | 1.0 | 0.0 | 1.0 |
| 6 | 0.0 | 0.6 | 0.0 | 1.0 | 0.0 | 1.0 | 1.0 | 0.0 |

From these results it can be seen that the attribute $c$ is not significant at all because removing it from the set of conditional attributes does not cause any change in the approximation quality, i.e. $c$ is dispensable. Details on the fuzzy-rough QuickReduct algorithm as well as how it could be applied to generate the $Q$-reduct $\{a, b\}$ of $P = \{a, b, c\}$ for this example can be referred to [22, 24, 25].

In the study of fuzzy information systems, in which attribute values of object may be fuzzy (linguistic) values, Mieszkowicz-Rolka and Rolka [28] proposed to define a so-called compatibility relation over $\mathbb{U}$ induced from a set of attributes $P$ as follows

$$\mu_P(x, y) = \min_{a \in P} \sup_{v \in V_a} \min(\mu_{f(x,a)}(v), \mu_{f(y,a)}(v)) \tag{34}$$

where $V_a$ is the domain of attribute $a$; $f(x, a)$ and $f(y, a)$ are fuzzy values of $x$ and $y$ at attribute $a$, respectively. Using this definition of a fuzzy similarity relation, fuzzy lower approximations of fuzzy sets can be defined using (18) and then (31) can be also used to define the approximation quality in case of fuzzy information systems.

Similarly, as discussed in the preceding section, it is of interest to mention here that equivalent representation of the approximation quality $\gamma_P(Q)$ by (20) may also suggest another extension for $\hat{\gamma}_P(Q)$. However, due to overlapping of fuzzy lower approximations, in this case we may need to carry out some normalization. For example, we can normalize involved fuzzy similarity relations so that (12) is satisfied, then a fuzzy counterpart of (20) can be used to define an extension for $\hat{\gamma}_P(Q)$. Another possibility is that we can carry out a normalization after defining a fuzzy counterpart of (20), for instance, as follows

$$\hat{\gamma}_P(Q) = \frac{1}{|\mathbb{U}||\mathbb{U}/Q|} \sum_{X \in \mathbb{U}/Q} \sum_{x \in \mathbb{U}} \mu_{\underline{P}(X)}(x) \tag{35}$$

Intuitively, we may observe that if the fuzzy lower approximation of some (fuzzy) class in $\mathbb{U}/Q$ dominates all those of the others, it solely affects the approximation quality $\hat{\gamma}_P(Q)$ defined by (31), while others classes play no role. This situation does not occur in the crisp case because of the disjoint union. In such a situation, an extension for $\hat{\gamma}_P(Q)$ guided by (20) may be interesting to be considered since, in any case, it takes fuzzy lower approximations of all classes in $\mathbb{U}/Q$ into account. This, however, requires further research.

## 7 Conclusion and Future Work

The concepts of approximation quality essentially play an important role in practical applications of rough set theory. They supply numerical characterizations for measuring the dependency between attributes in databases and the accuracy of concept approximation using the given data alone and no additional information. At the same time, rough-fuzzy hybrids have emerged naturally due to the need of encapsulating the related but distinct concepts of vagueness and indiscernibility, both of which occur as a result of imperfection in knowledge. This review paper has focused on those extensions of approximation quality that make use of rough fuzzy and fuzzy rough sets. We have also discussed how different but equivalent representations of approximation quality in the (crisp) rough case may lead to various different extensions for rough-fuzzy cases. However, much research work should be done in the future to explore theoretical features as well as practical implications of these mentioned extensions.

Let us conclude here by pointing out some issues regarding the research topic discussed, which would be interestingly considered for further research:

- Exploiting practical applications of roughness measure for fuzzy sets, particularly in classification and image analysis problems as pointed out in [3, 42], as well as its generalization in a fuzzy approximation space.
- Apart from those having been well studied, formulating and investigating other extensions of the approximation quality, for example as mentioned in the preceding section, and conducting comparative experiments to verify their applicability in, for example, dimensionality reduction in comparison with known extensions as studied in [22, 23, 24, 25, 44].
- Using rough-fuzzy hybrids based extensions of the approximation quality in areas of decision analysis [16], case-based reasoning [32] and knowledge discovery [39].
- Studying extensions of approximation quality in variable precision fuzzy rough sets model [29, 54] and their applicability.

**Acknowledgment** The authors would like to thank two anonymous reviewers and Editors for their helpful comments and suggestions.

## References

1. A. Apostolico, E. R. Caianiello, E. Fischetti, S. Vitulano, C-Calculus: an elementary approach to some problems in pattern recognition, **Pattern Recognition 19** (1978) 375–387.
2. J. F. Baldwin, The management of fuzzy and probabilistic uncertainties for knowledge based systems, in: S.A. Shapiro (Ed.), *The Encyclopaedia of AI*, Wiley, New York, 1992, pp. 528–537.
3. M. Banerjee, S. K. Pal, Roughness of a fuzzy set, *Information Sciences* **93** (1996) 235–246.
4. R. B. Bhatt, M. Gopal, On fuzzy-rough sets approach to feature selection, *Pattern Recognition Letters* **26** (2005) 965–975.
5. E. R. Caianiello, A. Petrosino, Neural networks, fuzziness and image processing, in V. Cantoni (Ed.): **Machine and Human Perception: Analogies and Divergences**, Plenum Press, 1994, pp. 355–370.
6. M. Ceccarelli, A. Petrosino, A parallel fuzzy scale-space approach to the unsupervised texture separation, **Pattern Recognition Letters 23** (2002) 557–567.
7. M. Ceccarelli, F. Musacchia, A. Petrosino, A fuzzy scale-space approach to feature-based image representation and retrieval, in M. De Gregorio et al. (Eds.): **Brain, Vision, and Artificial**

**Intelligence**, BVAI'2005, Naples, Italy, 2005, LNCS **3704**, Springer-Verlag, pp. 377–385.
8. S. M. Chen, C. M. Huang, Generating weighted fuzzy rules from relational database systems for estimating null values using genetic algorithms, *IEEE Transactions on Fuzzy Systems* **11** (2003) 495–506.
9. A. Chouchoulas, Q. Shen, Rough set-aided keyword reduction for text categorisation, **Applied Artificial Intelligence** 15 (9) (2001) 843–873.
10. D. Dubois, H. Prade, Putting fuzzy sets and rough sets together, in R. Słowiński (Ed.), *Intelligent Decision Support*, Kluwer Academic, Dordrecht, 1992, pp. 203–232.
11. D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, *International Journal of General Systems* **17** (1990) 191–209.
12. D. Dubois, H. Prade, Fuzzy sets, possibility and measurement, *European Journal of Operational Research* **40** (1989) 135–154.
13. G. Gediga, I. Düntsch, Rough approximation quality revisited, *Artificial Intelligence* **132** (2001) 219–234.
14. G. Gediga, I. Düntsch, Approximation quality for sorting rules, **Computational Statistics & Data Analysis 40** (2002) 499–526.
15. S. Greco, B. Matarazzo, R. Słowiński, Rough set handling of ambiguity, in H.-J. Zimmermann (Ed.), **Proceedings of the Sixth European Congress on Intelligent Techniques and Soft Computing** (EUFIT98), Mainz Wissenschaftsverlag, Aachen, 1998, pp. 3–14.
16. S. Greco, B. Matarazzo, R. Słowiński, Rough sets theory for multicriteria decision analysis, *European Journal of Operational Research* **129** (2001) 1–47.
17. S. Greco, M. Inuiguchi, R. Slowinski, Fuzzy rough sets and multiple-premise gradual decision rules, *International Journal of Approximate Reasoning* **41** (2006) 179–211.
18. S. Greco, M. Inuiguchi, R. Slowinski, A new proposal for fuzzy rough approximations and gradual decision rule representation, in J. F. Peters et al. (Eds.), **Transactions on Rough Sets II: Rough Sets and Fuzzy Sets**, LNCS **3135**, 2004, pp. 319–342.
19. U. Höhle, Quotients with respect to similarity relations, *Fuzzy Sets and Systems* **27** (1988) 31–44.
20. V. N. Huynh, Y. Nakamori, A rouhgness measure for fuzzy sets, *Information Sciences* **173** (2005) 255–275.
21. V. N. Huynh, T. Murai, T. B. Ho, Y. Nakamori, An extension of rough approximation quality to fuzzy classification, in D. Slezak et al.(Eds.) *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, LNCS **3641**, 2005, pp. 512–521.
22. R. Jensen, **Combining Rough and Fuzzy Sets for Feature Selection**, PhD Thesis, University of Edinburgh, UK, 2005.
23. R. Jensen, Q. Shen, Fuzzy-rough data reduction with ant colony optimization, **Fuzzy Sets and Systems 149** (2005) 5–20.
24. R. Jensen, Q. Shen, Fuzzy-rough attribute reduction with application to web categorization, **Fuzzy Sets and Systems 141** (2004) 469–485.
25. R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches, **IEEE Transactions on Knowledge and Data Engineering 16** (2004) 1457–1471.
26. R. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall PTR, Upper Saddle River, NJ, 1995.
27. E. Marczewski, H. Steinhaus, On a certain distance of sets and the corresponding distance of functions, *Colloquium Mathematicum* **6** (1958) 319–327.
28. A. Mieszkowicz-Rolka, L. Rolka, Fuzziness in information systems, **Electronic Notes in Theoretical Computer Science 82** (4) (2003) 1–10.
29. A. Mieszkowicz-Rolka, L. Rolka, Remarks on approximation quality in variable precision fuzzy rough sets model, in S. Tsumoto et al. (Eds.), RSCTC'2004, LNAI **3066**, Springer-Verlag, 2004, pp. 402–411.
30. T. Murai, H. Kanemitsu, M. Shimbo, Fuzzy sets and binary-proximity-based rough sets, *Information Sciences* **104** (1–2) (1998) 49–80.
31. S. K. Pal, A. Skowron (Eds.), *Rough Fuzzy Hybridization: New Trends in Decision Making*, Springer Verlag, Singapore, 1999.

32. S. K. Pal, P. Mitra, Case generation using rough sets with fuzzy representation, **IEEE Transactions on Knowledge and Data Engineering 16** (2004) 292–300.
33. Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences* **11** (1982) 341–356.
34. Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, MA, 1991.
35. Z. Pawlak, Rough sets and fuzzy sets, *Fuzzy Sets and Systems* **17** (1985) 99–102.
36. A. Petrosino, Rough fuzzy sets and unsupervised neural learning: applications in computer vision. In: A. Bonarini et al. (Eds.), **New trends in Fuzzy Logic**. World Scientific, Singapore, 1996, pp. 166–176.
37. A. Petrosino, M. Ceccarelli, M., Unsupervised texture discrimination based on rough fuzzy sets and parallel hierarchical clustering, **Proceedings of IEEE International Conf. on Pattern Recognition**, IEEE-CS Press, Silver Spring 2000, pp. 1100–1103.
38. A. Petrosino, G. Salvi, Rough fuzzy set based scale space transforms and their use in image analysis, *International Journal of Approximate Reasoning* **41** (2006) 212–228.
39. L. Polkowski, A. Skowron (Eds.), *Rough Sets in Knowledge Discovery*, Parts 1 & 2, Physica-Verlag, Heidelberg-New York, 1998.
40. A. M. Radzikowska, E. E. Kerre, A comparative study of fuzzy rough sets, **Fuzzy Sets and Systems 126** (2002) 137–155.
41. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).
42. M. Sarkar, Rough–fuzzy functions in classification, *Fuzzy Sets and Systems* **132** (2002) 353–369.
43. Q. Shen, A. Chouchoulas, A rough-fuzzy approach for generating classification rules, **Pattern Recognition 35** (2002) 2425–2438.
44. Q. Shen, R. Jensen, Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring, **Pattern Recognition 37** (2004) 1351–1363.
45. R. Słowiński (Ed.), *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic, Dordrecht/Boston/London, 1992.
46. W. Z. Wu, W. X. Zhang, Constructive and axiomatic approaches of fuzzy approximation operators, *Information Sciences* **159** (2004) 233–254.
47. M. Wygralak, Rough sets and fuzzy sets: some remarks on interrelations, *Fuzzy Sets and Systems* **29** (1989) 241–243.
48. Y. Y. Yao, Information granulation and rough set approximation, *International Journal of Intelligent Systems* **16** (2001) 87–104.
49. Y. Y. Yao, Combination of rough and fuzzy sets based on alpha-level sets, in T. Y. Lin & N. Cercone (Eds.), *Rough Sets and Data Mining: Analysis of Imprecise Data* (Kluwer Academic Publishers, Boston/London/Dordrecht, 1997) 301–321.
50. Y. Y. Yao, A comparative study of fuzzy sets and rough sets, *Information Sciences* **109** (1998) 227–242.
51. L. A. Zadeh, Fuzzy sets, *Information and Control* **8** (1965) 338–353.
52. L. A. Zadeh, Similarity relations and fuzzy orderings, *Information Sciences* **3** (1971) 177–200.
53. H. Zhang, H. Liang, D. Liu, Two new operators in rough set theory with applications to fuzzy sets, *Information Sciences* **166** (2004) 147–165.
54. W. Ziarko, Variable precision rough set model, **Journal of Computer and System Sciences 46** (1993) 39–59.

# Fuzzy Sets in Information Retrieval: State of the Art and Research Trends

**Gabriella Pasi**

**Abstract** In this contribution some applications of Fuzzy Set Theory to Information Retrieval are described, as well as the more recent outcomes of this research field. Fuzzy Set Theory is applied to Information Retrieval to the main aim to define *flexible* systems, i.e. systems that can represent and manage the vagueness and subjectivity which characterizes the process of information representation and retrieval.

## 1 Introduction

The advent and rapid diffusion of the Internet and the birth of the World Wide Web have caused a strong resurgence of interest in Information Retrieval, a Computer Science discipline whose roots date to late 60ties. With the diffusion of the World Wide Web the information available on-line have been increasing, and consequently the need for effective systems that allow an easy and flexible access to information has become a urgent need [38]. By flexibility is here meant the capability of the system to both manage imperfect (vague and/or uncertain) information, and to *adapt* its behaviour to the user context. Moreover, more recently, the increasing interest in defining the so called Semantic Web requires the definition of a basic infrastructure more powerful and flexible than the existing one to organise and to give a meaning to the available information, and to allow a better communication between humans and machines.

Search engines represent the most recent outgrowth of IR [20]. However, despite of the above mentioned needs most search engines are based on retrieval models defined several years ago, and, more surprisingly, the query language on which these systems are based is the Boolean query language, defined as the first formal query language for IRSs,. The Boolean query language forces the user to precisely express her/his information needs as a set of un-weighted keywords, thus not allowing users to express vague requirements for specifying selection conditions tolerant to imprecision. Two distinct users formulating the same query will obtain the same results despite of the fact that their choices can be defined on different criteria and to different aims.

For the above mentioned reasons, in recent years a great deal of research has been devoted to the promising direction of improving the (semi) automatic access

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008

to information, by modelling the subjectivity, vagueness and imprecision intrinsic in the process of locating relevant information. To this aim the application of Soft Computing techniques has been experienced as a means to obtain a greater flexibility in designing systems for Information Access [24, 18]. The expression Soft Computing (SC) was introduced by Lotfi Zadeh as a synergy of methodologies useful to solve problems using some form of intelligence that divert from traditional computing. The principal constituents of SC are: fuzzy logic, neural networks, probabilistic reasoning, and evolutionary computing, which in turn subsume belief networks, genetic algorithms, parts of learning theory, and multi-valued logics. As each of these methodologies allows to independently represent imprecision, uncertainty and learning, it is frequently advantageous to employ them in combination, rather than exclusively. Because of these characteristic Soft Computing has provided very powerful tools for modelling flexibility in IRSs.

In particular, Fuzzy Set Theory has been applied to IR starting in the 70ties, and has allowed the definition of retrieval techniques capable of modelling, at some extent, the human subjectivity for estimating the partial relevance of documents to the user needs. The objective of this contribution is to provide an overview of how fuzzy set theory has been applied to the aim of designing flexible Information Retrieval Systems. The chapter is organized as follows: in the next section the Information Retrieval problem is introduced. In Sect. 3 an overview of the main approaches to apply fuzzy set theory to model flexible Information Retrieval Systems is presented. In Sect. 4 a description of the traditional fuzzy document representation is first sketched; then some more recent and promising approaches to fuzzy indexing are described. Section 5 is devoted to the description of flexible query languages for Information Retrieval Systems based on the specification of soft constraints expressed by linguistic selection conditions which capture the vagueness of the user needs and simplify the query formulation.

In Sect. 6 some approaches to the application of fuzzy set theory to distributed information retrieval are described. Finally, in Sect. 7 a description of fuzzy associative retrieval models based either on fuzzy pseudo-thesauri of terms or fuzzy clustering techniques are introduced.

## 2 Information Retrieval

Information Retrieval (IR) aims at defining systems able to provide a fast and effective content-based access to a large amount of stored information usually organized in documents (information items) [3, 40, 41, 42, 44]. Information can be multimedia: textual, visual, or auditory, although most actual IR systems (IRS) store and enable the retrieval of only textual information.

A user accesses the IRS by explicitly formulating a query through a set of constraints that the relevant information items must satisfy. The aim of the IRS is to evaluate the user query and to retrieve all documents which it estimates relevant to that query. This is achieved by comparing the formal representation of the documents with the formal user's query. The activity of IRSs is then based on the solution of a decision-making problem: how to identify the information items that correspond to the users' information preferences (i.e. *relevant* to their information needs)? What a user expects from an IRS is a list of the relevant documents ordered according to her/his preferences. The IRS acts then as an intermediary in this decision process: it "simulates" the decision process that the user would personally undertake. The documents constitute the alternatives on which the decision process has to be performed to the aim of identifying the relevant ones [45].

In order to estimate the relevance of each document with respect to a specific user query the IRS must be based on a formal model which provides a formal representation of both documents and user queries. The main components of IRSs are: a collection of documents, a query language which allows the expression of selection criteria synthesizing the users' needs, and a matching mechanism which estimates the relevance of documents to queries (see Fig. 1).

The input of these systems is constituted by a user query; their output is usually an ordered list of selected items, which have been estimated relevant to the information needs expressed in the user query.

Most of the existing IRSs and search engines offer a very simple modelling of IR, which privileges efficiency at the expenses of effectiveness. A crucial aspect affecting the effectiveness of an IRS is related to the characteristics of the query language, which should represent in the more accurate and faithful way the user's information needs. The available query languages are based on keywords specification, and do not allow to express uncertainty and vagueness in specifying the constraints that the relevant information items must satisfy. In real situations, however, the users would find much more natural to express their information needs in an uncertain and vague way.

Another important aspect which affects the effectiveness of IRSs is related to the way in which the documents' information content is represented; the documents' representations are extremely simple, based on keywords extraction and weighting. Moreover the IRSs generally produce a unique representation of documents for all users, not taking into account that each user looks at a document content in a

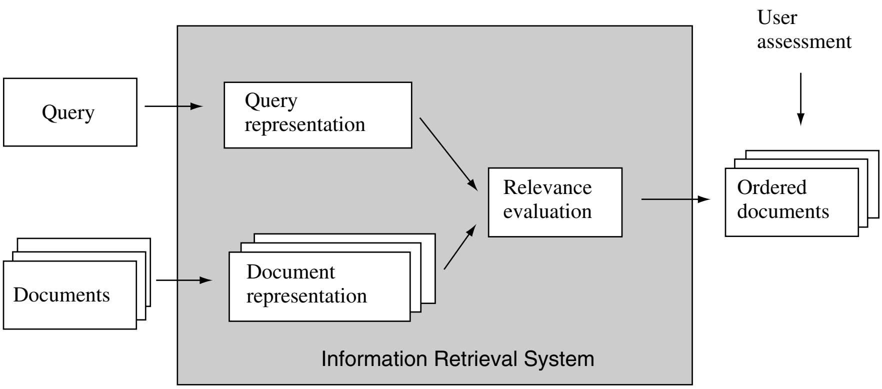



**Fig. 1** Scheme of a system for the storage and retrieval of information

subjective way, by possibly emphasizing some subparts with respect to others. This adaptive view of the document is not modelled. Another important aspect is related to the fact that on the WWW some standard for the representation of semi-structured information are becoming more and more employed (such as XML); for this reason it is important to exploit their structure in order to represent the information they contain.

In recent years a strong deal of research to improve IRSs was devoted to the modelling of the concept of partiality intrinsic in the IR process and to making the systems adaptive, i.e. able to "learn" the users' concept of relevance. In recent years big efforts have been devoted to the attempt to improve the performance of IR systems, and the research has explored several different directions to the aim of modelling the vagueness and uncertainty that invariably characterize the management of information. In particular, a set of approaches that has received a strong interest goes under the name of Soft Information Retrieval [8, 18, 19, 24]. These approaches apply some of the so called Soft Computing techniques, among which Fuzzy Set Theory.

In particular, fuzzy set theory has been extensively applied to extend IR to model some aspects of the vagueness and subjectivity characterizing the retrieval process. In the next section the main applications of fuzzy set theory to IR are synthesized.

## 3 Applications of Fuzzy Set Theory to Information Retrieval

To the aim of defining flexible IRS, fuzzy set theory has been successfully employed to the following aims:

1. to define new IR models;
2. to deal with the imprecision and subjectivity that characterize the document indexing process;
3. to manage the user's vagueness in query formulation;
4. to soften the associative mechanisms, such as thesauri and documents' clustering algorithms, which are often employed to extend the functionalities of the basic IR scheme;
5. to the definition of meta-search engines and to define flexible approaches to distributed IR;
6. to represent and inquiry semi-structured information (XML).

A survey on the definition of fuzzy IR models and of fuzzy generalizations of the Boolean IR model can be found in [8, 27]. Fuzzy generalizations of the Boolean model have been defined to the aim of designing IRSs able to produce discriminated answers in response to users' queries. In fact, Boolean IRSs apply an exact matching between a Boolean query and the representation of each document, defined as a set of index terms. They partition the archive of items into two sets: the relevant documents and those which are not relevant. As a consequence of this crisp behaviour, they are liable to reject relevant items as a result of too restrictive queries, and to retrieve useless material in reply to general queries [40].

Another approach to fuzzy modelling of IR is based on the use of linguistic information at various levels in the retrieval process [7, 22, 23]. More recently some interesting approaches have been defined to possibilistic based information retrieval [14, 16, 29].

At the level of document indexing some fuzzy techniques have been defined to the aim of providing more specific and personalized representations of documents than those generated by the existing indexing procedures. In Sect. 4, the basic fuzzy interpretation of the weighted document representation is introduced. As it happens with search engines, the incorporation of a weighted document representation in a Boolean IRS is a sufficient condition to improve the system with a document ranking ability. As a consequence of this extension the exact matching applied by a Boolean system can be softened to a partial matching mechanism, evaluating the degree of satisfaction of the user's query for each retrieved document. This value is called the Retrieval Status Value (RSV), and can be used for ranking documents. However, as it will be seen in Sect. 4, more flexible indexing functions can remarkably improve the systems' effectiveness. The main idea is to explicitly model an indexing strategy that adapts the formal document representation to the user personalized view of documents' information contents. In Sect. 4 a fuzzy and personalised indexing model of documents structured in logical sections (such as XML documents) is presented. This model can be tuned by users on the basis of their personal criteria for interpreting the content of documents [6, 12]. An indexing procedure for HTML documents is also shortly described [31, 35].

Fuzzy set theory has also been employed for defining flexible query languages, able to capture the vagueness of user needs as well as to simplify the user system interaction. This aim has been pursued at two levels: through the definition of soft selection criteria (soft constraints), which allow the specification of the distinct importance of the search terms, and by softening the way in which (weighted) search terms can be aggregated. Query languages based on numeric query term weights with different semantics have been first proposed as an aid to define more expressive selection criteria [17, 27]. Then, an evolution of these approaches has been defined, which introduces linguistic query weights, specified by fuzzy sets such as *important* or *very important*, in order to express the distinct importance of the query terms [4]. Another level of flexibility concerns the definition of soft aggregations of the selection criteria, by means of operators characterized by a parametric behaviour which can be set between the two extremes AND and OR adopted in the Boolean language. In [5, 30] the Boolean query language has been generalized by defining aggregation operators as linguistic quantifiers such *as at least k* or *most of.* In [13] an approach to extend the query languages for inquiring XML documents has been proposed. These extensions are presented in Sect. 5.

Fuzzy associative mechanisms based on thesauri or clustering techniques [33, 34] have been defined in order to cope with the incompleteness characterizing either the representation of documents or the users' queries. Fuzzy thesauri and pseudo-thesauri can be used to expand the set of index terms of documents with new terms by taking into account their varying significance in representing the topics dealt with in the documents; the degree of significance of the associated terms depends on the strength of the associations with the documents' descriptors. An alternative use of fuzzy thesauri and pseudo-thesauri is to expand the search terms in the query with associated terms, by taking into account their distinct importance in representing the concepts of interest; the varying importance is dependent on the associations' strength with the search terms. Fuzzy clustering can be used to expand the set of the documents retrieved by a query with associated documents; their degrees of association with respect to the documents originally retrieved influence their Retrieval Status Value. These approaches are more extensively explained in Sect. 7.

## 4 Fuzzy Approaches to Document Indexing

The production of effective retrieval results depends on both subjective factors, such as the users' capability to express their information needs through a formal query, and the characteristics of the Information Retrieval System. A component which plays a crucial role is the indexing mechanism, which has the aim of generating a formal representation of the contents of the information items (documents' surrogates). The most used automatic indexing procedures are based on term extraction and weighting: documents are represented by a collection of index terms with associated weights (the index term weights). An index term weight expresses the degree of significance of the index term as a descriptor of the document information content [40, 42]. The vector space model, the probabilistic models and fuzzy models adopt a weighted document representation. The automatic computation of the index term weights is based on the occurrences count of a term in the document and in the whole archive. In this case an indexing function F computes for each document d and each term t a numeric value. An example of definition of the function F is the following, in which the index term weight is proportional to the frequency of term t in document d, and inversely proportional to the frequency of the term in the documents of the archive:

$$\mathrm{F}(d,t) = tf_{dt} \times IDF_t \tag{1}$$

where:

- $tf_{dt}$ is a normalized term frequency which can be defined as: $tf_d = OCC_{dt}/MAXOCC_d$; where $OCC_{dt}$ is the number of occurrences of $t$ in $d$, and $MAXOCC_d$ is the number of occurrences of the most frequent term in $d$;
- $IDF_t$ is an inverse document frequency which can be defined as: $IDF_t = log(N/NDOC_t)$, where $N$ is the total number of documents in the archive and $NDOC_t$ is the number of documents indexed by $t$. The computation of $IDF_t$ is particularly costly in the case of large collections which are updated online.

The definition of such a function F is based on a quantitative analysis of the text which makes it possible to model the qualitative concept of significance of a term in describing the information carried by the text. The adoption of weighted indexes allows for an estimate of the relevance or of a probability of relevance of the documents to the considered query [3, 40, 44].

Based on such an indexing function and by maintaining the Boolean query language, the first fuzzy interpretation of an extended Boolean model has been to adopt a normalized weighted document representation and to interpret it as a fuzzy set of terms [8, 27]. From a mathematical point of view this is a quite natural extension: the concept of the significance of index terms in describing the information content of a document can be naturally described by adopting the function F (such as the one defined in (1) but normalized so as to obtain values in the range [0,1]) as the membership function of the fuzzy set representing a document. Formally, a document is represented as a fuzzy set of terms: $R_d = \Sigma_{t \in T} \mu_d(t) / t$ in which the membership function is defined as $\mu_d$: $D \times T \rightarrow [0,1]$. In this case $\mu_d(t) = F(d,t)$, i.e. the membership value is obtained by the indexing function F. Through this extension of the document representation, the evaluation of a Boolean query produces a numeric estimate of the relevance of each document to the query, expressed by a numeric score, called the Retrieval Status Value (RSV), which is interpreted as the degree of satisfaction of the constraints expressed in a query.

## *4.1 A Fuzzy Approach to Personalized Document Indexing*

The weighted representation of documents based on function (1) has the limitation of not taking into account that a term can play a different role within a text, according to the distribution of its occurrences. Let us think for example at an XML document organized in "logical" sections. For example scientific papers are usually organised into sections like *title*, *authors*, *abstract*, *introduction*, *references*, etc. An occurrence of a term in the *title* has a distinct informative role than an occurrence of the same term in the *references*. Moreover, indexing procedures based on the F function like the one defined in (1) behave as a black box producing the same document representation for all users; this enhances the system's efficiency but implies a severe loss of effectiveness. In fact, when examining a document structured in logical sections the users have their personal views of the document's information content; according to this view in the retrieval phase they would naturally privilege the search in some subparts of the documents' structure, depending on their preferences. This last consideration outlines the fact that the estimate of relevance of a given document could take advantage from an explicit user's indication of her/his interpretation of the document's structure, and supports the idea of *dynamic* and *adaptive* indexing [6, 12]. By adaptive indexing we intend personalized indexing procedures which take into account the users' indications to *interpret* the document contents and to "build" their synthesis on the basis of this interpretation. It follows that if an archive of semi-structured documents is considered (e.g. XML documents), flexible indexing procedures could be defined by means of which the users are allowed to direct the indexing process by explicitly specifying some constraints on the document structure (preference elicitation on the structure of a document).

This preference specification can be exploited by the matching mechanism to the aim of privileging the search within the most preferred sections of the document, according to the users' indications. The user/system interaction can then generate a personalized document representation, which is distinct for distinct users.

In [6] a user adaptive indexing model has been proposed, based on a weighted representation of semi-structured documents that can be tuned by users according to their search interests to generate their personal document representation in the retrieval phase. The considered documents may contain multimedia information with different structures. A document is represented as an entity composed of sections (such as *title*, *authors*, *introduction*, *references*, in the case of a scientific paper). The model is constituted by a static component and by an adaptive query-evaluation component; the static component provides an a priori computation of an index term weight for each logical section of the document. The formal representation of a document becomes then a fuzzy binary relation defined on the Cartesian product T × S (where T is the set of index terms and S is the set of identifiers of the documents' sections): with each pair <section, term¿, a significance degree in [0,1] is computed, expressing the significance of the term in the document section.

The adaptive component is activated by the user in the phase of query formulation and provides an aggregation strategy of the n index term weights (where n is the number of sections) into an overall index term weight. The aggregation function is defined on the basis of a two level interaction between the system and the user. At the first level the user expresses preferences on the document sections, outlining those that the system should more heavily take into account in evaluating the relevance of a document to a user query. This user preference on the document structure is exploited to enhance the computation of index term weights: the importance of index terms is strictly related to the importance for the user of the logical sections in which they appear.

At the second level, the user can decide which aggregation function has to be applied for producing the overall significance degree (see Fig. 2). This is done by the specification of a linguistic quantifier such as *at least k* and *most* [47]. In the fuzzy indexing model defined in [6, 12] linguistic quantifiers are formally defined as Ordered Weighted Averaging (OWA) operators [48].

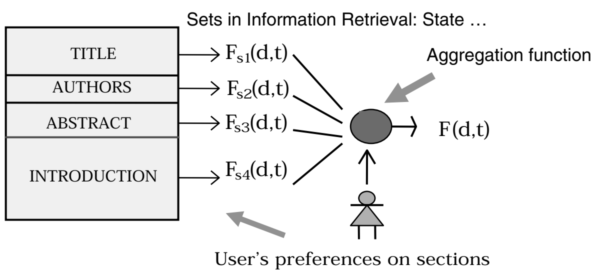



**Fig. 2** Sketch of the personalized indexing procedure

By adopting this document representation the same query can select documents in different relevance orders depending on the user indications.

It is very important to notice that the elicitation of users' preferences on the structure of a document is a quite new and recent research approach, which can remarkably improve the effectiveness of IRSs. In [31, 35] another representation of structured documents is proposed, which produces a weighted representation of documents written in HyperText Markup Language. An HTML document has a syntactic structure, in which its subparts have a given format specified by their delimiting tags. In this context tags are seen as syntactic elements carrying an indication of the importance of the associated text: when writing a document in HTML, one associates a distinct importance with distinct documents' subparts, by delimiting them by means of appropriate tags. On the basis of these considerations, an indexing function has been proposed, which computes the significance of a term in a document by taking into account the distinct role of term occurrences according to the importance of tags in which they appear.

### *4.2 A Fuzzy Approach to Concept-based Document Indexing and Retrieval*

Recently, an increasing number of approaches to IR have defined and designed IR models which are based on concepts rather than keywords, thus modeling document representations at a higher level of granularity, trying to describe the topical content and structure of documents [2]. These efforts gave raise to the so called concept-based Information Retrieval, which aims at retrieving relevant documents on the basis of their meaning rather than their keywords. The main idea at the basis of conceptual IR is that the meaning of a text depends on conceptual relationships to objects in the world rather than to linguistic relations found in text or dictionaries [43]. To this aim, sets of words, phrases, names are related to the concepts they encode.

In [15] a fuzzy set approach to concept-based Information Retrieval has been proposed. Based on the existence of a conceptual hierarchical structure which encodes the contents of the domain to which the considered collection of documents belongs, both documents and queries are represented as weighted trees. The evaluation of a conjunctive query is then interpreted as computing a degree of inclusion between sub-trees. The ontology-based description of the contents of the documents takes into account the semantic equivalences between expressions, as well as the basic principle stating that if a document explicitly heavily includes some terms, it also concerns to some extent more general concepts. This latter point is handled at the technical level by a completion procedure which assesses positive weights also to terms which do not appear directly in the documents. The possible completion of queries is also discussed in [16].

## 5 Fuzzy Approaches to the Definition of Flexible Query Languages

By flexible query language is intended a language that makes possible a simple and natural expression of subjective information needs. By means of fuzzy set theory some flexible query languages have been defined as generalizations of the Boolean query language. In this context a flexible query may consist of either both of the two following soft components or just one: the first component is constituted by weighted terms that are interpreted as flexible constraints on the significance of the index terms in each document representation. The second component is constituted by linguistic aggregation operators which can be applied to the flexible constraints in order to specify compound selection conditions. The atomic selection conditions are expressed by weighted terms expressed by pairs <term, weight>, in which weight can be either a numeric value in [0,1] (which identifies a soft constraint) or a linguistic value of the linguistic variable *Importance,* and the compound conditions are expressed by means of linguistic quantifiers used as aggregation operators. The notion of linguistic variable is suitable to represent and manage linguistic concepts and for this reasons it has been used to formalize the semantics of linguistic terms introduced in the generalized Boolean query language [46]. When flexible constraints are specified, the query evaluation mechanism is regarded as performing a fuzzy decision process that evaluates the degree of satisfaction of the query constraints by each document representation by applying a partial matching function. This degree (the Retrieval Status Value) is interpreted as the degree of relevance of the document to the query and is used to rank the documents. Then, as a result of a query evaluation, a fuzzy set of documents is retrieved in which the RSV is the membership value. The definition of the partial matching function is strictly dependent on the query language definition and specifically on the semantics of the flexible constraints, and is defined as a bottom-up evaluation procedure: first, each atomic selection condition (flexible constraint) in the query is evaluated for a given document, and then the aggregation operators are applied to the results starting from the inmost operator in the query to the outermost operator. Flexible constraints are defined as fuzzy subsets of the set [0,1] of the index term weights; the membership value $\mu_{weight}(F(d,t))$ is the degree of satisfaction of the flexible constraint imposed by the *weight* associated with query term t by the index term weight of t in document d. The result of the evaluation is a fuzzy set: $\Sigma_{d \in D}\mu_{weight}(F(d,t))/d$.

A first proposal to specify flexible constraints was by means of numeric weights associated with terms. A numeric weight identifies a constraint on the weighted document representation, which depends on the considered semantics. Distinct semantics have been associated with query weights [17, 27]. However, the association of a numeric value forces the user to quantify the qualitative concept of importance of query weights, also if at the level of query evaluation this constraint is evaluated in a gradual way. To overcome this limitation and to make the query language more user friendly, in [4] a linguistic extension of the Boolean query language was defined, based on the concept of linguistic variable [46]. By this language the user can associate with query terms either the primary term "*important*", or some compound terms, such as "*very important*" or "*fairly important"* to qualify the desired importance of the search terms in the query. A pair <t, *important*>, expresses a flexible constraint evaluated by the function $\mu_{important}$ on the term significance values (the F(d,t) values). The evaluation of the relevance of a given document d to a query consisting of the pair <t,*important*> is then computed by applying the function $\mu_{important}$ to the value F(d,t).

A second approach to make the Boolean query language more flexible has concerned the specification of aggregation operators. In the Boolean query language, the AND and OR aggregation operators are used. When the AND is used for aggregating M selection conditions, the satisfaction of all conditions but one is not tolerated, with the consequence that this may cause the rejection of useful items. To face this problem, within the framework of fuzzy set theory a generalization of the Boolean query language has been defined, based on the concept of linguistic quantifiers: they are employed to specify both crisp and vague aggregation criteria of the selection conditions [5]. New aggregation operators can be specified by linguistic expressions, with a self-expressive meaning such as *at least k* and *most of*. They are defined with a behaviour between the two extremes corresponding to the AND and the OR connectives, which allow, respectively, requests for *all* and *at least one of* the selection conditions. The linguistic quantifiers used as aggregation operators, have been defined by Ordered Weighted Averaging (OWA) operators [48]. An alternative approach is proposed in [30].

By adopting linguistic quantifiers, the requirements of a complex Boolean query can be more easily and intuitively formulated. For example when desiring that *at least 2* out of the three selection conditions "politics", "economy", "inflation" be satisfied, one should formulate the following Boolean query:

(politics AND economy) OR (politics AND inflation) OR
(economy AND inflation)

which can be replaced by the simpler one:

*at least 2*(politics, economy, inflation)

The expression of any Boolean query is supported by the new language via the nesting of linguistic quantifiers. For example a query such as:

<image> AND (<processing> OR <analysis>) AND <digital>

can be translated into the following new formulation:

*all* (<image>, *at least 1 of* (<processing>, <analysis>), <digital>)

A quantified aggregation function can thus be applied not only to single selection conditions, but also to other quantified expressions.

In [12] a generalisation of the Boolean query language that allows to personalize the search in structured documents (as showed in Sect. 4) was proposed; both content-based selection constraint, and soft constraints on the document structure can be expressed. The atomic component of the query (basic selection criterion) is defined as follows:

$$aq = t \text{ in } Q \text{ preferred sections}$$

in which $t$ is a search term expressing a content-based selection constraint, and $Q$ is a linguistic quantifier such as *all, most*, or *at least k%*. $Q$ expresses a part of the structure-based selection constraint. It is assumed that the quantification refers to the sections that are semantically meaningful to the user. $Q$ is used to aggregate the significance degrees of $t$ in the desired sections and then to compute the global Retrieval Status Value RSV(d,*aq*) of the document d with respect to the atomic query condition *aq*.

## *5.1 An Approach to Extend the XPath Query Language*

With the development of the World Wide Web the diffusion of the de-facto standards for the definition of structured documents such XML, witnesses the tendency of producing documents in which the information is organized into (often hierarchical) components. In particular, XML is increasingly gaining importance as a standard format for information interchange on the WWW. XML has been employed as a basic model for describing semi-structured data, and it constitutes the basic standard for representing structured documents in IR.

In order to inquiry semi-structured information the need for flexible query languages has soon emerged. In the context of semi-structured databases, by flexible query languages it is substantially meant languages that take into account the lack of a rigid schema of the database, thus allowing to enquiry both data and the type-/schema [1, 12]. In the context of IRSs, modelling flexibility means to take into account the possibility to make explicit a non-uniform structure of the documents when formulating queries.

In [13], fuzzy set theory has been applied to define a flexible extension of the XPath query language to the aim of expressing soft selection conditions on both the documents' structure and contents. XPath is a standard language (www.w3.org/TR/xpath) that allows to write "tree traversal expressions" for selecting XML tree nodes. XPath expressions are also used as selection conditions in the framework of fully-fledged XML query languages. In the last years, much work has been done towards a standard for XML querying and recently the W3C endorsed the XQuery language (www.w3.org/TR/xquery) as a candidate recommendation. Both in XQuery and in XPath a retrieved information item is usually a "node set".

The extensions of XPath proposed in [13] are finalized at:

- fuzzy sub-tree matching to the aim of providing a ranked list of retrieved information items rather than the usual set oriented one;
- use of fuzzy predicates, to the aim of specifying flexible selection conditions;

- fuzzy quantification, to the aim of allowing the specification of linguistic quantifiers as aggregation operators.

The research work presented in this paper constitutes a step towards the more and more increasingly studied problem of inquiring XML documents not only from a structural point of view, but also from a content-based point of view [25].

## 6 Fuzzy Approaches to Distributed Information Retrieval

With the increasing use of the network technologies, the need of defining distributed applications has emerged. In distributed Information Retrieval, there are two main models: in the first model the information is considered as belonging to a unique, huge database which is distributed but "centrally" indexed for retrieval purposes. This is the model adopted by search engines on the WWW. A second model is based on the distribution of the information on distinct databases, independently indexed, and thus constituting distinct sources of information. This last model gives rise to the so called distributed or multi-source information retrieval problem. In this second case the databases reside on distinct servers each of which can be provided with its own search engine (IRS). The multi-source information retrieval paradigm is more complex than the centralized model as it presents additional problems, such as the selection of an appropriate information source for a given information need. A common problem which can be identified with both models is the problem of list fusion. In the case in which we have a unique, huge and distributed information repository (like in the WWW), and distinct IRSs (search engines), which can be used to inquiry overlapping collections, meta-search engines have been defined to improve the effectiveness of the individual search engines. The main aim of a meta-search engine is to submit the same query to distinct search engines and to fuse the individual resulting lists into an overall ranked list of documents that is presented to the user. In this case we typically have overlapping individual lists since a document may be retrieved by more than a single search engine. The fusion method must then be able to handle situations in which a document may appear in more than one list and in different positions within them. In the case of multi-source information retrieval the problem is to merge the lists resulting from the processing of the same query by (generally distinct) search engines on the distinct databases residing on distinct servers. However, in this second case we generally do not have overlapping lists as a result of the same query evaluation. Typically a document will be retrieved by just one single search engine, and thus the fusion problem is simplified with respect to the previous case. Recently in the literature several papers have addressed the problem of defining effective solutions to the problem of retrieving information on a network. In [10] some approaches to the definition of meta-search engines are presented, while in [9] some solutions to the problem of multi-source information retrieval are described. In this last paper both previous models have been considered, and some fuzzy approaches to the solution of the two above mentioned problems have been proposed. The uniqueness of these approaches is that they are based on soft computing techniques to more flexibly model the resource selection problem (in distributed information retrieval), and the list fusion problem [9, 10].

In [10] a meta-search model has been proposed where the soft fusion of overlapping ordered lists into an overall ordered list is regarded as a Group Decision Making activity in which the search engines play the role of the experts, the documents are the alternatives that are evaluated based on a set of criteria expressed in a user query, and the decision function is a soft aggregation operator modelling a specific user retrieval attitude.

# 7 Fuzzy Associative Mechanisms

Associative retrieval mechanisms are defined to enhance the retrieval of traditional IRSs. They work by retrieving additional documents that are not directly indexed by the terms in a given query but are indexed by other terms, associated descriptors. The most common type of associative retrieval mechanism is based on the use of a thesaurus to associate entry terms with related terms. In traditional associative retrieval the associations are crisp.

The fuzzy associative retrieval mechanisms are based on the concept of fuzzy associations [33]. A fuzzy association between two sets $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$ is formally defined as a fuzzy relation $f : X \times Y \rightarrow [0,1]$: the value f(x,y) represents the degree of strength of the association existing between the values $x \in X$ and $y \in Y$.

In information retrieval, different kinds of fuzzy associations can be derived depending on the semantics of the sets X and Y.

Fuzzy associative mechanisms employ fuzzy thesauri, fuzzy pseudothesauri, and fuzzy clustering techniques to serve three alternative, but compatible purposes:

- to expand the set of index terms of documents with new terms
- to expand each of the search terms in the query with associated terms,
- to expand the set of the documents retrieved by a query with associated documents.

A thesaurus is an associative mechanism that can be used to improve both indexing and querying. It is well known that the development of thesauri is very costly, as it requires a large amount of human resources. Moreover, in highly dynamic situations, where terms are added and new meanings derived for old terms quite rapidly, the thesaurus needs frequents updates. For this reason, methods for the automatic construction of thesauri have been proposed, based on statistical criteria such as the terms' co-occurrences, i.e., the simultaneous appearance of pairs (or triplets, or even larger subsets) of terms in the same documents.

In a thesaurus the relations defined between terms are of different type: if the associated descriptor has a more general meaning than the entry term, the relation is classified as broader term (BT), while a narrower term (NT) is the inverse relation; synonyms or near-synonyms are associated by a related term (RT) relation.

Some authors have proposed the definition of fuzzy thesauri, see [34, 36, 37], where the links between terms are weighted to indicate the strength of the association. Moreover, this notion includes generalizations such as fuzzy pseudo-thesauri [34], and fuzzy associations based on a citation index [36].

## 7.1 Fuzzy Clustering

Clustering in information retrieval is applied for partitioning a given set of documents D into groups using a measure of similarity (or distance) which is defined on every pairs of documents. The similarity between documents in the same group should be large, while it should be small for documents in different groups. A common method to perform clustering of documents is based on the simultaneous occurrences of citations in pairs of documents. Documents are so clustered using a measure defined on the space of the citations. Generated clusters can then be used as an index for information retrieval; that is, documents which belong to the same clusters as the documents directly indexed by the terms in the query are retrieved. Often, similarity measures are suggested empirically or heuristically [41, 42]. When adopting the fuzzy set model, clustering can be formalized as a kind of fuzzy association. In this case, the fuzzy association is defined on the domain $D \times D$, where D is the set of documents. By assuming R(d) to be the fuzzy set of terms representing a document d with membership function values $\mu_d(t)=F(d,t)$ being the index term weights of term t in document d, the symmetric fuzzy relation s is taken to be the similarity measure for clustering documents:

$$
\begin{aligned}
s(d_1, d_2) &= \sum_{k=1}^{M} \min[\mu_{d1}(t_k), \mu_{d2}(t_k)] / \sum_{k=1}^{M} \max[\mu_{d1}(t_k), \mu_{d2}(t_k)] \\
&= \sum_{k=1}^{M} \min[F(t_k, d_1), F(t_k, d_2) / \sum_{k=1}^{M} \max[F(t_k, d_1), F(t_k, d_2)] \qquad (2)
\end{aligned}
$$

in which M is the cardinality of the set of index terms T.

In fuzzy clustering, documents can belong to more than one cluster with varying degree of membership [21, 26, 28, 39]. Each document is assigned a membership value to each cluster. In a pure fuzzy clustering, a complete overlap of clusters is allowed. Modified fuzzy clustering, or soft clustering, approaches use threshold mechanisms to limit the number of documents belonging to each cluster. The main advantage of using modified fuzzy clustering is the fact that the degree of fuzziness is controlled.

In [11] a new unsupervised hierarchical fuzzy clustering algorithm has been defined to the aim of identifying the main categories of news in a new-stream information filtering system. In the following we list the distinguished characteristics of the proposed approach to support category based news filtering.

The output of the proposed fuzzy algorithm is fuzzy hierarchy of the news given as input; this reflects the very nature of a news, which may deal with multiple topics. The algorithm computes a membership degree in [0,1] for each item (news) to each generated fuzzy cluster. This allows to rank the news within a cluster and thus easily support flexible filtering strategies such as the selection of the top ranked news within a cluster of interest. The generated fuzzy hierarchy represents the topics at different levels of granularity, from the most specific ones corresponding to the clusters of the lowest hierarchical level (the deepest level in the tree structure representing the hierarchy), to the most general ones, corresponding with the clusters of the top level. Since topics may overlap one another, the hierarchy is fuzzy, thus allowing each cluster of a level to belong with distinct degrees to each cluster in the next upper level. The proposed algorithm works bottom up in building the levels of the fuzzy hierarchy. Once the centroids of the clusters in a level of the hierarchy are generated, the fuzzy clustering algorithm is re-applied to group the newly identified centroids into new fuzzy clusters of the next upper level. In this way, each level contains fuzzy clusters that reflect topics homogeneous with respect to their specificity (or granularity), so that, in going up the hierarchy, more general topics are identified.

The clusters hierarchy can be easily and efficiently updated on-line when recent news arrive on the stream. This may possibly increase the number of the clusters already identified, and thus may require to compute the association of the old news to the new clusters.

Since the optimal number of clusters to generate is unknown, the proposed algorithm automatically determines this number. The procedure is based on the analysis of the shape of the cumulative histogram curve of overlapping degrees between pairs of news vectors. It identifies the number of clusters of news sharing a minimum overlapping degree corresponding with the point on the curve of highest trend's variation.

## 8 Conclusions

In this contribution some approaches to the definition of flexible Information Retrieval Systems by applying Fuzzy Set Theory have been presented. In particular some promising research directions that could guarantee the development of more effective IRSs have been outlined. Among these, the research efforts aimed at defining new indexing techniques of semi-structured documents (such as XML documents) are very important: the possibility of creating in a user-driven way the documents' surrogates would ensure a modeling of the users' interests also at the indexing level (usually this is limited to the query formulation level). Other promising directions are constituted by conceptual document indexing, and flexible distributed Information Retrieval.

## References


1. Abiteboul S., *Querying Semi-Structured Data*, Lecture Notes In Computer Science, Proceedings of the 6th International Conference on Database Theory, pp. 1–18, 199.
2. Azzopardi, L., Girolami M. L., and van Rijsbergen C.J., *Topic Based Language Models for ad hoc Information Retrieval*, in: Proceedings of the International Joint Conference on Neural Networks, Budapest, Hungary, 2004.
3. Baeza-Yates R., Ribeiro-Neto B., *Modern Information Retrieval*. Addison-Wesley, Wokingham, UK, 1999.
4. Bordogna G.and Pasi G., *A fuzzy linguistic approach generalizing Boolean information retrieval: a model and its evaluation*, Journal of the American Society for Information Science, 44(2), pp. 70–82, 1993.
5. Bordogna G.and Pasi G., *Linguistic aggregation operators in fuzzy information retrieva,*. International Journal of Intelligent systems, 10(2), pp. 233–248, 1995.
6. Bordogna G.and Pasi G., *Controlling retrieval trough a user-adaptive representation of documents*, International Journal of Approximate Reasoning, 12, 317–339, 1995.
7. Bordogna G.and Pasi G., *An Ordinal Information Retrieval Model*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 9, 2001.
8. Bordogna G.and Pasi G., *Modelling Vagueness in Information Retrieval*, in *Lectures in Information Retrieval*, M. Agosti, F. Crestani and G. Pasi eds., Springer Verlag., 2001.
9. Bordogna G., Pasi G., and Yager R.R., *Soft approaches to distributed information retrieval*, International Journal of Intelligent Systems, Vol. 34, pp. 105–120, 2003.
10. Bordogna G.and Pasi G., *Soft fusion of Infomation Accesses*, Fuzzy Sets and Systems, 148, pp. 205–218, 2004.
11. Bordogna G.., Pagani M., and Pasi G., *A dynamical Hierarchical fuzzy clustering algorithm for document filtering*, in "Soft Computing for Information Retrieval on the Web", Springer Verlag, 2006.
12. Bordogna G. and Pasi G., *Personalized Indexing and Retrieval of Heterogeneous Structured Documents*, Information Retrieval, Kluwer, Vol. 8, Issue 2, pp. 301–318, 2005.
13. Braga D., Campi A.. Damiani E.,.Pasi G, Lanzi PL, *FXPath: flexible querying of XML documents*, in Proceedings of EUROFUSE 2002, Varenna, Italy, 2002.
14. Boughanem M., Loiseau Y., Prade H., *Improving document ranking in information retrieval using ordered weighted aggregation and leximin refinement.*, in: EUSFLAT-LFA 2005, 4th Conference of the European Society for Fuzzy Logic and Technology and 11me Rencontres Francophones sur la Logique Floue et ses Applications, pp. 1269–1274, 2005.
15. Boughanem M., Pasi G., Prade H., Baziz M., *A fuzzy logic approach to information retrieval using an ontology-based representation of documents*, in "Fuzzy Logic and the Semantic Web" (E. Sanchez, Ed.), Elsevier Science, 2006.
16. Brini A., Boughanem M., Dubois D., *A Model for Information Retrieval Based on Possibilistic Networks*, in: String Processing and Information Retrieval (SPIRE 2005), LNCS, Springer Verlag, pp. 271–282, 2005.
17. Buell D.A., and Kraft D.H., *Threshold values and Boolean retrieval systems*, Information Processing & Management 17, pp. 127–136, 1981.
18. Crestani F. and Pasi G. eds., *Soft Computing in Information Retrieval: Techniques and Applications,* Physica Verlag, *series* Studies in Fuzziness, 2000.
19. Crestani F. and Pasi G., *Soft Information Retrieval: Applications of Fuzzy Set Theory and Neural Networks*, in: "Neuro-fuzzy Techniques for Intelligent Information Systems", N.Kasabov and Robert Kozma Editors, Physica-Verlag , Springer-Verlag Group , pp. 287–313, 1999.
20. Glover E. J., Lawrence S., Gordon M. D., Birmingham W. P., and Lee Giles C., *Web Search – YourWay*, Communications of the ACM, 1999.
21. Hathaway R.J., Bezdek J.C., and Hu Y., *Generalized Fuzzy C-Means Clustering Strategies Using Lp Norm Distances*, IEEE Transactions on Fuzzy Systems, 8(5), pp. 576–582, 2000.

22. Herrera-Viedma E., *Modeling the Retrieval Process of an Information Retrieval System Using an Ordinal Fuzzy Linguistic Approach*, Journal of the American Society for Information Science and Technology (JASIST), Vol. 52 N. 6, pp. 60–475, 2001.
23. Herrera-Viedma E., Cordon O., Luque M., Lopez A.G., Muñoz A.N., *A Model of Fuzzy Linguistic IRS Based on Multi-Granular Linguistic Information*, Int. Journal of Approximate Reasoning, 34(3), pp. 221–239, 2003.
24. Herrera-Viedma E., Pasi G. and Crestani F. eds., *Soft Computing in Web Information Retrieval: Models and Applications*, Series of Studies in Fuzziness and Soft Computing, Springer Verlag, 2006.
25. Fuhr N., Lalmas M eds., *Introduction to the Special Issue on INEX*, Information Retrieval, Kluwer, 8(4), pp. 515–519, 2005.
26. Kraft D., Chen J., Martin–Bautista M.J., Vila M.A., *Textual Information Retrieval with User Profiles using Fuzzy Clustering and Inferencing*, in: Intelligent Exploration of the Web, Szczepaniak P., Segovia J., Kacprzyk J., Zadeh L.A. eds., Studies in Fuzziness and Soft Comp. Series, 111, Physica Verlag, 2003.
27. Kraft D., Bordogna G., Pasi G., *Fuzzy Set Techniques in Information Retrieval*, in: "Fuzzy Sets in Approximate Reasoning and Information Systems", J. C. Bezdek, D. Dubois and H. Prade eds, volume of the series "The Handbooks of Fuzzy Sets Series", Kluwer Academic Publishers, pp. 469–510, 1999.
28. Lin K., Ravikuma K., *A Similarity-Based Soft Clustering Algorithm for Documents,* in: Proceedings of the 7th International Conference on Database Systems for Advanced Applications, pp. 40–47, 2001.
29. Loiseau Y., Boughanem M., Prade H., *Evaluation of term-based queries using possibilistic ontologies*, in: Soft Computing for Information Retrieval on the Web, Herrera-Viedma E., Pasi G., Crestani F. Eds., Springer-Verlag, 2006.
30. Losada D., Diaz-Hermida F. and Bugarin A., *Semi-fuzzy quantifiers for information retrieval*, in: "Soft Computing in Web Information Retrieval: Models and Applications", Series of Studies in Fuzziness and Soft Computing, Springer Verlag. Edited by E .Herrera-Viedma, G. Pasi and F. Crestani, volume 197/2006.
31. Marques Pereira R.A., Molinari A., and Pasi G., *Contextual weighted representations and indexing models for the retrieval of HTML documents*, Soft Computing, Vol. 9, Issue 7, pp. 481–492, July 2005.
32. Mendes Rodrigues M.E.S. and Sacks L., *A Scalable Hierarchical Fuzzy Clustering Algorithm for Text Mining*, in: Proceedings of the 4th International Conference on Recent Advances in Soft Computing, RASC'2004, pp. 269–274, Nottingham, UK, 2004.
33. Miyamoto S., *Fuzzy sets in Information Retrieval and Cluster Analysis.* Kluwer Academic Publishers, 1990.
34. Miyamoto S., *Information retrieval based on fuzzy associations*, Fuzzy Sets and Systems, 38(2), pp. 191–205, 1990.
35. Molinari A., and Pasi G., *A Fuzzy Representation of HTML Documents for Information Retrieval Systems*, in: Procedings of the IEEE International Conference on Fuzzy Systems, 8–12 September, New Orleans, U.S.A., Vol. 1, pp. 107–112, 1996.
36. Nomoto, K., Wakayama, S., Kirimoto, T., and Kondo, M., *A fuzzy retrieval system based on citation*, Systems and Control, 31(10), pp. 748–755, 1987.
37. Ogawa, Y., Morita, T., and Kobayashi, K., *A fuzzy document retrieval system using the keyword connection matrix and a learning method*, Fuzzy Sets and Systems, 39(2), pp. 163–179, 1991.
38. Pasi G., *Modelling Users' Preferences in Systems for Information Access*, International Journal of Intelligent Systems, Vol. 18, pp. 793–808, 2003.
39. Pedrycz W., *Clustering and Fuzzy Clustering*, Chap. 1, in: Knowledge-based clustering, J. Wiley and Son, 2005.
40. Salton G., *Automatic Text Processing - The Transformation, Analysis and Retrieval of Information by Computer*, Addison Wesley Publishing Company, 1989.
41. Salton G., and McGill M.J., *Introduction to modern information retrieval.* New York, NY: McGraw-Hill, 1983.

42. Sparck Jones K. A., *A statistical interpretation of term specificity and its application in retrieval*, Journal of Documentation, 28(1), pp. 11–20, 1972.
43. Thomopoulos R., Buche P., Haemmerlé O., *Representation of weakly structured imprecise data for fuzzy querying*. Fuzzy Sets and Systems, 140, 111–128, 2003.
44. van Rijsbergen C.J., *Information Retrieval. London,* England, Butterworths & Co., Ltd., 1979.
45. Vincke P., *Multicriteria Decision Aid*, John Wiley & Sons, 1992.
46. Zadeh L. A., *The concept of a linguistic variable and its application to approximate reasoning, parts I, II*, Information Science, 8, pp. 199–249, pp. 301–357, 1975.
47. Zadeh L.A., *A computational Approach to Fuzzy Quantifiers in Natural Languages*, Computing and Mathematics with Applications. 9, 149–184, 1983.
48. Yager, R.R., *On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making*, IEEE Transactions on Systems Man and Cybernetics, 18(1), pp. 183–190, 1988.

# Fuzzy Sets and Web Meta-search Engines

**José A. Olivas**

**Abstract** In this work, a description of what is a Web Meta-search engine and the roles that fuzzy logic as a Soft Computing technique, can play to improve the search with this kind of artefacts are described. Fuzzy logic can provide tools for extracting and use knowledge from thesaurus and ontologies, formalize sentences and implement deduction capabilities in Question Answering systems, combine fuzzy values and different logics, design clustering algorithms and manage Meta-search engines architectures.

## 1 Introduction

The Web search process has become complicated and ineffective due to the big size of Internet and the exponential growth in the information. In every web search, users retrieve a big amount of irrelevant information. The reason is that most of the major Search engines are based on lexical and sometimes syntactic aspects [43]. Soft computing techniques have taken an important role [22, 40] in order to provide a way for improving web search results and lots of approaches have been proposed in the recent last years. Some of them are leaded to the construction of flexible adaptive sites (based on web patterns, user profiles, access patterns, user behaviour patterns . . .) using Data Mining techniques [8, 35, 41, 49]. Some others are focused on the organization of the retrieved documents into groups, being important to point out the ones based on clustering algorithms [58] in contrast to those supported by predefined thematic groups. Other approaches include systems based on flexible query languages [7, 21], or systems based on fuzzy association rules that help the user to find new terms to be used in the query [9].

Other approaches are focused in the documents representation, most of them based on extensions of the standard vector space model [45]. It is frequent to find systems based on term interrelations stored in Thesaurus and ontologies such as WordNet, a semantic net of word groups [37]. Meta-search engines appears as a promising new tool for improve Web search results, based on the use of several major Search engines such as Google or Yahoo and selecting the results from these very good sources. These kinds of systems are very different to the one proposed by

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008 

Zadeh [56] which is focused on the development of Question-Answering Systems, a very interesting point of view in information retrieval problems.

In this work, first is described what is a Meta-searcher and later some topics where fuzzy logic can contribute very interesting approaches are explained.

## 2 What is a Meta-searcher?

The Meta-searchers are search engines that have not an own database that contains the index of the documents. They usually are slower than search engines due to the fact that they follow a more complex and elaborated process. They provide a unified interface for consulting different search engines. Therefore, they limit to receive the requests of the users and them to send it to other searchers. The results must be classified to bring together in a unique list the documents returned by all the search engines. The main problem of the meta-searchers consists of combining the lists returned by the other searchers so that the performance is optimized.

Nevertheless, this type of systems improves some of the present problems in the traditional searchers, as the *Recall* problem, though they still suffer the *Precision* problem. According to Kerschberg [24], the way of solving this problem is approached using mainly four mechanisms: based on the content, collaborative, of knowledge of the domain and based on ontology methods. The methods based on the content try to obtain a representation of the as concrete as possible preferences of the user to later improve the evaluation of the returned pages, based on the content of the document and the preferences of the user. Inside this category there can be *WebWatcher* [1], *WAWA* [48] and *WebSail* [6]. The collaborative method is based on the similarity among the users to determine the relevancy of the information, *Phoaks* [50] and *Site-seer* [4]. The method based on the knowledge of the domain uses the help of the user and of the knowledge of the domain of the search to provide a major relevancy. Finally, the method based on ontology establishes a hierarchy among concepts that allows to make concrete and to improve the search. Of this type are *WebSifter II* [24], that uses a tree representation named *WSTT* (*Weighted Semantic Taxonomy-Tree*) to represent the intentions of search of the users, *OntoSeek* [20], *On2Broker* [12] and *WebKB* [34].

The main advantages of a Meta-searcher [36] are:

- *It facilitates the invocation of multiple searchers.* It allows, by means of a unique query, to obtain the most relevant documents index-linked by multiple searchers, which avoids the user to search in each of them.
- *It improves the efficiency of the recovery.* Since it is possible to consult specialized searchers, it allows obtaining of them a set more relevant of documents, without suffering the standard deviation that produces the high number of documents that there index the searchers of general intention.
- *It solves the scalability of the search in the Web.*
- *It increases the coverage of the search in the Web.* Due to the enormous quantity of documents that Internet contains, it is impossible that an alone search engine

index links the totality of the Web [29]. Therefore, by means of the combination of different searchers, it is possible to cover a bigger number of documents in the searches.

Likewise, according to Aslam et al. [2], a Meta-search engine also presents the following potential advantages:

- *Modular architecture*: the technologies used in the Meta-searchers can be split in small and more specialized modules that can be parallel and collaboratively executed.
- *Consistency*: current searchers often answer very different the same query once passed a time [47]. If different sources are used for obtaining the results, the variability would be small, favoured by the searchers that provide more stable results.
- *It improves the Recall factor:* on having obtained the results of multiple searchers it can improve the number of relevant documents recovered (the *Recall* factor).
- *It improves the Precision:* different algorithms recover many relevant equal documents, but different irrelevant documents [31]. Based on this phenomenon, in case of being true this theory, any algorithm that gives priority to the documents that appear in the first positions in results of different searchers will obtain an improvement in the recovery. This phenomenon is usually called "choir effect" [52].

But there are some disadvantages that Meta-search engines could have:

- *The selection of the database*: this problem is associated with the selection of the searcher that will receive the query: to select the searchers that return good results for a concrete query. For example, it makes no sense the query "electric guitar" on a specialized in scientific literature searcher. To try to solve this disadvantage, Meng proposes the use of measures that indicate the utility of every database with regard to a given query. He classifies these mechanisms under three categories: wide representation methods, statistical representation methods and methods based on learning.
- *The selection of documents:* once selected the origin of the documents the problem consists of determining the appropriate number of documents that it is necessary to request. If too many documents are considered, the computational cost to determine the best documents and the cost of communication to obtain them can be excessive. Meng also establishes a series of mechanisms that try to solve this problem divided in four categories: decision of the user, weights (major number of documents are obtained from the searcher that is considered to be the best), methods based on learning (it is based on the past to determine the number of documents of every searcher) and the guaranteed retrieval (it tries to guarantee the recovery of all the potentially useful documents).
- *Mix up of the results:* the problem consists of merging the results of different searchers with its own characteristics and forms of evaluation in an ordered by relevancy list. In addition, the possibility exists of finding repeated documents returned by different searchers. The technologies used to solve this problem can

be classified in both local similarity adjustment (it is based on the characteristics of the searcher or the similarity returned) and estimation for global similarity (the similarity of every recovered document with the original query is evaluated or estimated).

The translation of a query into each specific searcher language can play an important role in a Meta-searcher due to the fact that every search engine has a typical query language. Then, to adapt every query to the language of every searcher seems to be necessary.

There are a lot of Meta-search engine architectures, such as the ones proposed by Li [32], Kerschberg [24] and Glover [18]. Usually, the structure is decomposed in a series of more or less specific modules. Meng describes an architecture of reference with 5 components (Fig. 1).

- *User's Interface:* For obtaining the query of the user. In some cases it can provide an interactive query refinement system, based on the use of some knowledge structures. In addition, it is the manager of showing the results of the search.
- *Database selector:* It tries to select the searchers that better answers will give to the query of the user. It tries to avoid a massive sending of queries to all the searchers which causes a low performance and a high cost in time.
- *Documents selector:* The aim is to recover the major number of relevant documents, avoiding retrieving not relevant ones. If an excessive number of not relevant documents are retrieved, the efficiency of the search will be negatively affected.
- *Query Dispatcher:* It is in charge of establishing the connection with the searcher and to send it the query (or queries), as well as to get the results. It is usually used *http* (HyperText Transfer Protocol) by means of the use of the methods GET and POST. Nevertheless, there exist searchers which facilitate an interface of programming (API) to send queries and they use different protocols (Google uses the SOAP protocol in its API).

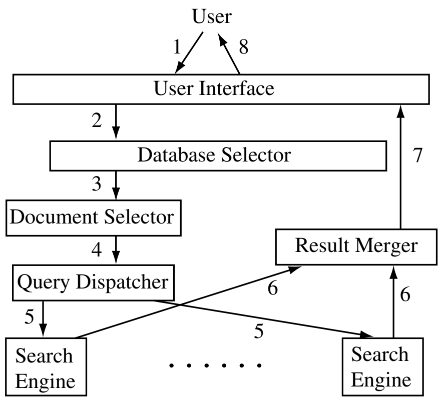



**Fig. 1** Meta-searcher software components architecture [36]

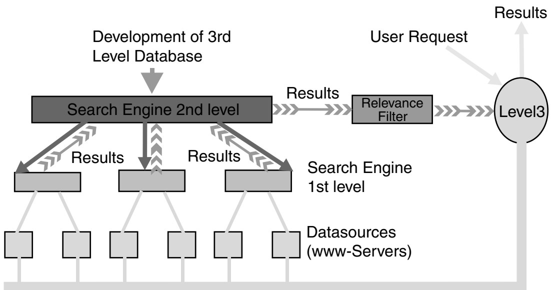



**Fig. 2** Search engines of the 3rd generation [23]

- *Result Merger:* Its principal function is to combine the results of the different searchers in a list. It is essential the use of some evaluation criteria to establish an order in the list that shows to the user.

Nowadays it is usual to find in literature references about Meta-Search engines of the 3rd generation (or level-3 search engines). They work as it is shown on Fig. 2: The user gives order to create a level-3 database. Meta-search engines (representing level 2 search engines) use standard search engines (level 1) to find corresponding results. Then a relevance-feedback analysis is done on these results. With that collection of addresses and text documents a level 3 database is developed. It contains only documents which are relevant for that user, who can get more query concerned information by querying this database. In other words, a database is created and is tailor-made for user's personal fields of interest. A query on this database therefore provides high quality results.

Then, it is necessary to think about including user profiles and customization in future Meta-searchers.

The "real" Meta-search engines simultaneously search the major search engines, aggregate the results, eliminate the duplicates and return the most relevant matches, according to the engine's algorithm. Some of the most popular Meta-search engines in the Web are:

- **Vivisimo** [http://vivisimo.com] uses clustering algorithms, meaning matches are organized in folders. It was created by researchers at Carnegie-Mellon University. There are some advanced searching options available: exact phrase, Boolean operators, fields searching (domain, host, title, URL, etc.) . . .
- **Pro Fusion** [http://www.profusion.com/], from the University of Texas [11, 16].
- **Mamma** [http://mamma.com] displays the results in a uniform manner according to relevance.
- **MetaCrawler**. [http://www.metacrawler.com].
- **Dogpile** [http://www.dogpile.com/].
- **Ixquick** [http://www.ixquick.com/] Multilanguage.
- **Ez2Find** [http://Ez2Find.com/] searches several Search engines and directories.

Through its "Advanced Search" function it also searches a small part of the invisible (Deep) Web.

- Some other interesting Meta-search engines are the following: **InfoGrid** [http://www.infogrid.com/]. **Infonetware** [http://www.infonetware.com/] sorts the results into topics. **IBoogie** [http://iboogie.com/] has a minimalist design and uses clustering algorithms. **Fazzle** [http://www.fazzle.com/] uses an interesting ranking algorithm. **Query Server** [http://www.queryserver.com/web.htm] searches a list of 11 Search engines (except Google). This is another example of the use of clustering algorithms. **Meta Bear** [http://www.metabear.com/] provides relevant results from both international and Russian sites. **Web Scout** [http://www.webscout.com/] uses the major Search engines (except Google) and provides relevant matches in a clean results list. **Experts Avenue** [http://www.expertsavenue.com/] enables online language translation of Web pages. **EmailPinoy** [http://www.emailpinoy.com/] from the Philippines. **Search 66** [http://www.search66.com/] groups together pages from the same domain. **Internav** [http://www.internav.com/]. **Metengine** [http://www.metengine.com/] (Antigua). **One2Seek** [http://www.one2seek.com/]. **Ithaki** [http://www.ithaki.net/]. **meta EUREKA** [http://www.metaeureka.com/] (Netherlands). **Widow** [http://www.widow.com/]. **7 Meta Search** [http://www.7metasearch.com/]. **Bytedog** [http://www.bytedog.com/] (Canada). **il motore** [http://www.ilmotore.com/] (Italy). **ApocalX** [http://www.search.apocalx.com/] (France) . . .

There can be taken into account, the engines that could be considered *pseudo* Meta-search engines: Those which exclusively group the results by search engine in one long list (such as **qb Search** [http://www.qbsearch.com/], **Better Brain** [http://www.betterbrain.com/], **My Net Crawler** [http://www.mynetcrawler.com/], **NBCi** [http://www.nbci.msnbc.com/], **Planet Search** [http://www.planetsearch.com/], **Rede Search** [http://www.redesearch.com/] and **Search Wiz** [http://www.searchwiz.com/]), and those which open a separate browser window for each search engine used (such as for example **Multi-Search-Engine** [http://www.multi-search-engine.com/] that opens 36 windows, **The Info** [http://www.theinfo.com/], **Net Depot** [http://www.netdepot.org/], **Alpha Seek** [http://www.alfaseek.com/] or **Express Find** [http://www.expressfind.com/]).

## *2.1 Web Meta-search Engines*

This work is focused on Web Meta-search engines because Web pages are the main point of interest of most of the general Search engines, and most of the Meta-searchers described and developed are focused on the Web. The design and implementation of these kinds of systems is usually supported by an index process that consists of creating an index of the terms and some other features that contains the web pages that the searcher can or want to access. The index process is usually implemented as an offline task, and the index is stored in a big database (usually distributed) that is consulted when a user send a query.

## 3 Meta Search Engines and Fuzzy Logic

Nowadays there are not commercial Fuzzy Searchers. Soft computing and Fuzzy Logic could play an interesting role in Web Search and Meta-search engines. Following some relevant topics which could be improved with Soft computing techniques are presented.

### *3.1 The Role that Nowadays Fuzzy Logic Plays in Search and Meta-search Engines*

If it is done a search in Google with the words "*fuzzy searcher*" many results appear. Selected those that we consider to be relevant, we observe that the fuzzy character that some commercial searchers assume is based exclusively on the use of a syntactic fuzzy matching; this is, in a fitting of the word included and possibly badly messed with another from a dictionary that the searcher contains or the searcher accedes and which is correctly written. The result is that the searcher sends a sign of notice (written text) that says: *Did you mean?*... Evidently, though the utility of this function is, to name a searcher as "*fuzzy*" for implementing it is excessive.

More detailed, in popular search engines the fuzzy search operator makes it possible to search with a segment of a word when the user does not know how to spell the entire word. For example, many medical terms and pharmaceutical products have difficult spellings. You may know how to say the word, but searching for it requires you to know how to spell as well. *Fuzzy searching* and *wild card operators* allow you to get around this problem. For example, Netscape Search supports the use of fuzzy searching. To search for the popular antibiotic "amoxicillin" the user can use the fuzzy search operator. To do this, he has to type as much of the word as he knows how to spell followed by the tilde character (∼) followed by your best guess at how the rest of the word is spelled. For example, the search term amoxi∼cilan will successfully return any sites that contain rough matches of this search term including the correct one.

It is very similar the search based in Wild Card Characters. It is possible to accomplish searches for different variations and unknown spellings in other ways. For example, many search engines supports three different well known wild card characters: the dollar sign ($), the question mark (?), and the asterisk (*).

### *3.2 The Role that Fuzzy Logic can Play in Meta-search Engines*

The proposal is that a searcher will be fuzzy when it implements approximate semantic searches; this is, when it includes in the searches semantic approximate criteria, not only syntactic ones. Some aspects that there would be considered are presented.

### The Use of a Dictionary of Synonymous and Thesaurus (Ontology)

When a user searches for a single word, the search can be facilitated by the use of a dictionary of synonymous. The dictionary will allow searches not only for the source word, but for its synonymous and will make possible to calculate the synonymy degree, having to contemplate this degree in the relevancy of the retrieved pages as response to the source terms.

The search can also be improved using thesaurus and ontologies. Nowadays, there are many ontologies referring to different domains, that improve several aspects in some applications, but they all have been hand-made following different methodologies such as Methontology [13] or those by Gruninger [19]. On the other hand, automatic ontology building is a focus in current research where the results hitherto have not been too satisfactory.

Nowadays, the online more used thesaurus is probably WordNet [38] that is organized on the basis of the semantic relations among different words. The bases of WordNet are the synonymy and hyponymy relations and other similar ones. In this way, groups of synonymous of a word can be found grouped in sets called *synsets*, and a polysemic word can belong to different *synsets*. The hyponymy relation (and the hypernymy one) is established among different terms, making a hierarchy based on the hyponymy. So, if dog "is a type of" *canine*, then *canine* is a hyponymy of *dog*. These relations provide important information that allows the expansion of the query terms to incorporate semantic information into the search process, as well as a mechanism to identify the correct meanings of the terms involved in a query. This kind of systems usually require a special matching mechanism, as the ontomatching algorithm proposed by Kiryakov and Simov [26] when comparing the concepts associated to the words. Whaley [53] propose other corpus-based search system that uses the probability that certain concepts co-occur together to disambiguate meanings. Leacock and Chodorow [30] approach the sense disambiguation problem by studying the local context of the words and comparing them to the habitual context words of each one of the word senses. This system requires the usual context words to be stored in a repository. Loupy and El-Bèze [33] also uses WordNet and propose a disambiguation system based on training the system with corpus of documents. Ramakrishnan et al. [42] study the disambiguation from a soft point of view. In this case, words are not disambiguated to an only sense, but rather to a set of senses with their corresponding relevance degrees (obtained by Bayesian Belief Networks). Lafourcade et al. [28] introduce the concept of relative synonymy to define a model of concept-based vectors. In this model, a term can be represented by a conceptual vector which is obtained by the linear combination of the definitions of the whole set of concepts. This system requires a concept repository.

As an example, FIS-CRM [15] is a model for representing the concepts contained in any kind of document. It can be considered an extension of the vector space model (VSM) [45, 46]. Its main characteristic is that it is fed on the information stored in a fuzzy synonymy dictionary [14] and several fuzzy thematic ontologies. The dictionary stores the synonymy degree between every pair of recognized synonyms. The ontology stores the generality degree between every word and its more general words. The way of calculating this value is the one proposed by

Widyantoro and Yen [54]. The key of the FIS-CRM model is first to construct the base vectors of the documents considering the number of occurrences of the terms (what we call VSM vectors) and afterwards readjust the vector weights in order to represent concept occurrences, using for this purpose the information stored in the dictionary and the ontologies. The readjusting process involves sharing the occurrences of a concept among the synonyms which converge to the concept and give a weight to the words that represent a more general concept than the contained ones.

## Sentences Search and Deduction Capabilities

If the search includes sentences, besides the dictionary of synonymous, thesaurus and ontologies, suitable fuzzy connectives should be used, to discriminate for example between a search "A and B" where A and B have common information, of the search "A and B" where A and B are completely independent. Something similar can happen with the relation "A or B". Another desirable aspect is that the searcher keeps the meaning of the words in mind under the synonymy relation, to choose the best similarity function.

But the problem can be bigger in the case of "causal" relations. First, it is very difficult to detect a causal relation in a written sentence (a query or a text). For example, the text could be: "stormy and dark", that could be understood by a person as: "If the weather is stormy, the sky gets dark". How can a search engine distinguishes the conjunctive "and" and the causal one? Nowadays is rather impossible, even if there is some knowledge about the context. Second, it is very difficult to find the most adequate implication function to represent the sentence (it is well known that there is a huge variety of fuzzy implications). The detection and management of causal relations could be very important for developing *Question Answering Systems*.

To detect the causal relationships that exist in a collection of documents, a starting point could be to detect conditional phrases. Nevertheless, this is not an easy task. Descartes could not have possibly imagined that to propose his famous phrase "I think, therefore I am", would have given birth to so many conjectures and interpretations for centuries after. In reality, what did he want to say, "First I think and after I am a person", or "As I am capable of thought, I am a person". To sum up, even on this occasion the intention of Descartes seems clear when he expressed his maxim, it is not easy to interpret and format the information expressed in natural language, especially when it involves complex sentences with complicated turns.

With the aim of detecting conditional phrases, some basic systems of detecting structures and a classification of sentences have been developed (i.e. [39]) which allow to locate, in terms of basic components (verb tenses, adverbs, linguistic turns, etc.), some causal forms. To accomplish the grammatical analysis, it is observed on the one hand, that it is possible to separate certain causal relationships based on the verb form used, while on the other hand it is possible to separate others based on the adverbs used in the sentences. Both analyses give rise to some causal rules that can be used to make an automatic extraction of knowledge. In the same way, every structure is subdivided into two structures which correspond to the antecedent and consequence of the causal relationship, and a parameter that measures the degree of certainty, conjecture, or compliance of the said causal relationship. In other words, it is not the same to form a sentence such as: "If I win the lottery, I will buy a car", in which there is no doubt that if the antecedent comes true the consequence will come true, as to form the sentence "If we had bought a ticket in Sacramento, we could have won the lottery" which leaves many more doubts and conjectures, in which you cannot be sure that the completion of the antecedent guarantees the consequences.

But this is still a *Natural Language Processing* complex problem. There are some other very interesting approximations, such as the one of Trillas [51] for representing conditional sentences with fuzzy implications. On the other hand, an approach based on *PNL* and *protoforms* could be a promising work line, such as Prof. Zadeh proposes[1]:

*"Existing search engines—with Google at the top—have many remarkable capabilities; but what is not among them is deduction capability—the capability to synthesize an answer to a query from bodies of information which reside in various parts of the knowledge base.*

*In recent years, impressive progress has been made in enhancing performance of search engines through the use of methods based on bivalent logic and bivalent-logic-based probability theory. But can such methods be used to add nontrivial deduction capability to search engines, that is, to upgrade search engines to question-answering systems? A view which is articulated in this note is that the answer is "No".*

*The problem is rooted in the nature of world knowledge, the kind of knowledge that humans acquire through experience and education.*

*It is widely recognized that world knowledge plays an essential role in assessment of relevance, summarization, search and deduction. But a basic issue which is not addressed is that much of world knowledge is perception-based, e.g., "it is hard to find parking in Paris," "most professors are not rich," and "it is unlikely to rain in midsummer in San Francisco." The problem is that (a) perception-based information is intrinsically fuzzy; and (b) bivalent logic is intrinsically unsuited to deal with fuzziness and partial truth.*

*To come to grips with the fuzziness of world knowledge, new tools are needed. The principal new tool—a tool which is briefly described in their note—is Precisiated Natural Language (PNL). PNL is based on fuzzy logic and has the capability to deal with partiality of certainty, partiality of possibility and partiality of truth. These are the capabilities that are needed to be able to draw on world knowledge for assessment of relevance, and for summarization, search and deduction."*

---

[1] Seminar: Web Intelligence, World Knowledge and Fuzzy Logic. Lotfi A. Zadeh, September 14; 2004, University of California, Berlkeley.

### Combination of Fuzzy Values

A Meta-searcher has to carry out a combination of logics (the algorithms that every searcher uses) in order to combine the local similarities in a global similarity or final order. But the local similarities are not based on fuzzy criteria. Therefore, the orders of relevant pages are not approximate. Usually Meta-searchers consider the searchers according to the prestige that they grant for the question that the users do and, depending on it, they qualify its results to incorporate them into the final list. In this case, the approximation comes for a criterion of market, but, again, not for a linguistic criterion. It would be interesting to apply this criterion, first, as it is previously indicated, doing that searchers make fuzzy semantic searches. Later, achieving that Meta-searchers arrange the pages in a final list combining the relevancy of the searchers with the confidence degrees associated with every result of its local obtained list, not only with the word got in the search box, but also using its related (synonyms...) words, the measures of similarity used in the calculation of the degree of linguistic relations and the fuzzy Boolean operators used in searches with sentences. Each of these searches would answer, therefore, to a fuzzy logic used by the searchers, which the Meta-searcher would have to combine to provide the final order. The use of the links that provides the order that the Meta-searcher gives might be useful as a test bench to check hypothesis on the combination of fuzzy logics.

Another important problem appears when it is necessary to aggregate several different fuzzy values from various sources. Two words (concepts) can have more than one linguistic relations (each one with its fuzzy value), such as hyponymy and synonymy. For example "football" and "soccer" are synonyms but the first is also more general than the last. A causal relation can also exist between both words (concepts). Moreover, a fuzzy relation based on the physical distance (same sentence, paragraph, chapter...) could be considered. Then, it is necessary to join all these different fuzzy values in only one, to be applied in representation and search tasks. How to aggregate these fuzzy values is still an open problem. Nowadays it is usually done with standard OWA operators [55] (or similar such as LOWA or WOWA ones). Another solution could be to use other operators, such as the ones presented by Castro and Trillas [5].

### Results Fuzzy Clustering

Document classification or text categorization (as used in information retrieval context) is the process of assigning a document to a predefined set of categories based on the document content. However, the predefined categories are unknown in a real repository of documents. Text clustering methods can be applied to structure the resulting set of documents, so they can be interactively browsed by the user. Therefore, using a clustering process, it is possible to achieve the splitting up of the collection of documents in a reduced number of groups made up of documents with enough conceptual similarity.

The amount of documents on a repository can range from tens to thousands and the system should manage their contents efficiently. Thus, the most important features of the organization of the repository and its components (classifiers, etc.) must be the following:

- *Dimensionality:* The classifier can handle feature spaces of tens of thousands dimensions, this requires the ability to deal with sparse data spaces or a method of dimensionality reduction.
- *Efficiency:* The documental clustering algorithms must be very efficient and scalable. The method should be also accurate in the task of classifying an incoming document.
- *Understandability:* The method must provide an understandable description of the discovered clusters.
- *Updatability:* The classifier update promptly itself as each new document is filed in the repository.

There are a lot of Soft-computing based clustering algorithms described in the literature, such as *fuzzy C-means* [3], *self organized maps* based on Neural Networks architectures (i.e. Kohonen Maps) [27], etc. But nowadays *soft-clustering* algorithms (i.e. the one presented in [25]) and *dynamic clustering interfaces* (i.e. Grouper [57]) are frequently used for Meta-search engines classification tasks.

### Meta-search Engines Architectures

Some typical tendencies in the use of search engines limit the effectiveness of the search process. Added to search problems based on key-words, there is a usual lack of experience in the users when using search engines. Meta-search engines appear as a promising alternative for trying to relieve this low precision of major search engines. An increase in the semantic capability of user queries is necessary to improve the relevance of the search results. For this purpose, it is usually used the *query expansion technique* [10] with terms semantically related with the terms introduced by the user. There are different strategies that can be adopted to expand a query, each one with its own characteristics. Basically, these mechanisms can be classified in *manual*, *automatic* and *interactive* query expansions. The *automatic* algorithms try to incorporate automatically new terms semantically related with those ones introduced by the user, to obtain a set of documents closer to the user search intention. But these approximations often are not easy to do, due to the different problems that vocabulary presents. Different approaches have been developed, for trying to get the correct meaning of the terms of a query and to focus the search for obtaining better results, in an automatic fashion. These algorithms are usually named *Word Sense Disambiguation* (WSD) algorithms. The *interactive query expansion* method consists of requesting the help of the user. By means of this mechanism the system suggests a series of terms among which the user has to choose. This type of systems frequently uses a tree structure from more general concepts towards more specific terms. It can make the process slow and uncomfortable because usually the user needs to answer several questions. On the other hand, the query expansion is usually done using knowledge structures (which initially could be WordNet but it could also be based on other ontologies or thesaurus).

There are many Meta-search engine architectures proposed [17, 24, 32] in which there are several common components. This is the case for the components that make queries to be sent to different major search engines, as well as those components which calculate the relevance of the retrieved documents. Most of the proposed architectures are based on the use of *agents* with different functions which communicate with each other through a net. Each agent develops a specific task, working cooperatively with other agents and reducing the system complexity. However, a communication language known by all the system agents is necessary to work cooperatively. This common language is known as *Agent Communication Language* (ACL).

Fuzzy logic could play a fundamental role in this agent based architecture, mainly in the task of joining the information from different sources (agents) and managing the results in an efficient and satisfactory way.

## 4 Conclusions and Guidelines

Taking into account these presented criteria, among others, would make possible to have really fuzzy searchers, or what is the same, searchers that do searches in terms of approximate meanings. The main focus of these engines must be the Web, not for general search artefacts but for Meta-search tools, because they use General Web Search engines as a basis.

Having fuzzy searchers would offer the possibility to do interesting tests and experiments. The Artificial Intelligence is an area of mixture of logics, because the approaches in the formal analysis of a sentence can be very different. Then, the logical form of the following phrase, a bit long, but not strange: "I suppose that you believe that I will pick you up a little bit earlier", implies using different logics: belief, non monotonic, fuzzy, temporal, ... But the problem is more complex yet, because, for the words with vague meaning, there can be also several modalities of fuzzy logics. The election has not been studied too much. Meta-searchers could provide a useful frame, restricted by the language that it lets, to research on the variety of formalisms that fuzzy logic provides.

Using user profiles in Web Meta-search engines could provide some advantages to improve the search. The user profile can be another parameter to take into account for expanding the query (with profile-related concepts: synonyms, broader than ...), for selecting the search engines and adapting the queries to them and for choosing and ranking the results of the search. Soft computing techniques can help in learning and representation tasks.

Meta Question-answering Systems?, perhaps the next goal to achieve would be Meta Web Question-answering Systems, which analyze the user question and generate a set of precise queries (expanded queries) to the more suitable major Search engines and Directories, to get the correct answer to the query. Soft computing and mainly fuzzy logic, as tools closer to human expression nature, can play an essential role for detecting the human user correct meaning and intention.

## References

1. Armstrong, R., Freitag, D., Joachims, T., Mitchell, T. (1995): WebWatcher: A Learning Apprentice for the World Wide Web, in Proceedings of the 1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed environments.
2. Aslam, J. A., Montague, M. (2001): Models for Metasearch, in Proceedings of 24th Internationsl ACM SIGIR, New Orleans, Louisiana, USA, 276–283.
3. Bezdek, J. C., Pal, S. K. (eds.) (1991): Fuzzy Models for Pattern Recognition (Methods that search for structures in data). IEEE Press.
4. Bollacker, K. D. (2000): Discovering Relevant Scientific Literature on the Web, IEEE Intelligent Systems, Vol. 15, No. 2, 42–47.
5. Castro, J.L., Trillas, E., Zurita, J. M. (1998). Non-monotonic fuzzy reasoning. Fuzzy Sets and Systems 94, North Holland, 217–225.
6. Chen, Z. (2000): WebSail: from on-line learning to Web search, in Proceedings of the First International Confererence on Web Information Systems Engineering, Vol. 1, 206–213.
7. Choi, D. (2001). Integration of document index with perception index and its application to fuzzy query on the Internet, Proc. of the 1st BISC Int. Workshop on Fuzzy Logic and the Internet, UC Berkeley, 68–72.
8. Cooley, R., Mobasher, B., Srivastaba, J. (1997). Grouping web page references into transactions for mining world wide web browsing patterns, Technical report TR 97–021, University of Minnesota, Minneapolis.
9. Delgado, M., Martin-Bautista, M.J., Sánchez, D., Serrano, J.M., Vila, M.A. (2003). Association rules and fuzzy associations rules to find new query terms, Proc. of the EUSFLAT'03 Conference, 49–53.
10. Efthimiadis, E. N. (1996): Query Expansion. ARIST, Vol. 31, 121–187.
11. Fan, Y., Gauch, S. (1997): An Adaptive Multi-Agent Architecture for the ProFusion Meta Search System, Proc. of WebNet '97: The Second World Conference of the Web Society, Toronto, ON.
12. Fensel, D. (1999): On2Broker: Semantic-Based Access to Information Sources at the WWW, in Proceedings of the World Conference on the WWW and Internet (WebNet 99), Honolulu, Hawaii, USA, 25–30.
13. Fernández López, M., Gómez-Pérez, A., Pazos, J., Pazos, A. (1999): Building a Chemical Ontology Using Methontology and the Ontology Design Environment, IEEE Intelligent Systems 14 No.1, 37–46.
14. Fernandez, S. (2001). A contribution to the automatic processing of the synonymy using Prolog, PhD Thesis, University of Santiago de Compostela, Spain.
15. Garcés, P., Olivas, J. A., Romero, F. P. (2006): Concept-matching IR systems versus Word-matching IR systems: Considering fuzzy interrelations for indexing web pages. Journal of the American Society for Information Science and Technology JASIST, 57 (4): 564–576.
16. Gauch, S., Wang, G., Gomez, M. (1996): ProFusion: Intelligent Fusion from Multiple, Distributed Search Engines, Journal of Universal Computing, Springer-Verlag, Vol. 2 (9).
17. Glover, E. J., Lawrence, S., Birming- ham, W. P., Giles C. L. (1999): Architecture of a metasearch engine that supports user information needs. In Eighth International Conference on Information and Knowledge Management (CIKM'99) ACM, 210 216.
18. Glover, E. J., Lawrence, S., Virmingham, W. P., Giles, C. L. (1999): Architecture of a Metasearch Engine that Supports User Information Needs, in Proc. of the Eighth International Conference on Information Knowledge Management, (CIKM-99), 210–216.

19. Gruninger, M., Fox, M. S. (1995): Methodology for the Design and Evaluation of Ontologies, Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.
20. Guarino, N. (1999): OntoSeek: content-based access to the Web, IEEE Intelligent Systems, Vol. 14, No. 3, 70–80.
21. Herrera-Viedma, E. (2001). An information retrieval system with ordinal linguistic weighted queries based on two weighting elements, Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems, 9, 77–88.
22. Herrera-Viedma, E., Pasi, G. (2003). Fuzzy approaches to access information on the Web: recent developments and research trends, Proc. of the Third Conference of the EUSFLAT, 25–31.
23. http://www2.iicm.edu/cguetl/education/projects/seminar/seminarhtml/seminar.htm last visit September 2006.
24. Kerschberg, L., Kim, W., Scime, A. (2001): A Semantic Taxonomy-Based Personalizable Meta-Search Agent. Second International Conference on Web Information Systems Engineering (WISE'01).
25. King-Ip Lin, Kondadadi, R.: (2001): A similarity-based soft clustering algorithm for documents. Accessible at www. msci.memphis. edu/∼linki/_ mypaper/ dasfa 01clus .doc. Last visit September 2006.
26. Kiryakov, A.K., Simov, K.I. (1999): Ontologically supported semantic matching, Proc. of NODALIDA'99: Nordic Conference on Computational Linguistics, Trondheim.
27. Kohonen, T. (1984): Self-Organization and Associative Memory. Springer-Verlag, Berlín.
28. Lafourcade, M., Prince, V. (2001): Relative Synonymy and conceptual vectors, Proc. the Sixth Natural Language Processing Pacific Rim Symposium, Japan, (202), 127–134.
29. Lawrence, S., Giles, C. L. (1999): Accessibility of Information on the Web, Nature, Vol. 400, 107–109.
30. Leacock, C., Chodorow, M. (1998): Combining local context and Wordnet similarity for word sense disambiguation, In WordNet, an Electronic Lexical Database, MIT Press, Cambridge Ma, 285–303.
31. Lee, J. H. (1997): Analyses of multiple evidence combination, in Proc. of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, PA, USA, 267–275.
32. Li, Z., Wang, Y., Oria, V. (2001): A New Architecture for Web Meta-Search Engines, in Proc. of Seventh Americas Conference on Information Systems, 415–422.
33. Loupy, C., El-Bèze, M. (2002): Managing synonymy and polysemy in a document retrieval system using WordNet, Proc. of the LREC2002: Workshop on Linguistic Knowledge Acquisition and Representation.
34. Martin, P., Eklund, P. W. (2000): Knowledge retrieval and the World Wide Web, IEEE Intelligent Systems, Vol. 14, No. 3, 18–25.
35. Martin-Bautista, M.J., Vila, M., Kraft, D., Chen, J. (2001). User profiles and fuzzy logic in web retrieval, Proc. of the 1st BISC Int. Workshop on Fuzzy Logic and the Internet, UC Berkeley, USA, 19–24.
36. Meng, W., Yu, C., Liu, K-L. (2002): Building Efficient and Effective Metasearch Engines, in Proceedings of ACM Computing Surveys, Vol. 34, No. 1, 48–89.
37. Miller, G. (ed.) (1990). WORDNET: An online lexical database. International Journal of Lexicography, 3(4). 235–312.
38. Miller, G.A. (1995). WordNet: A lexical database for English, Communications of the ACM 11, 39–41.
39. Olivas, J.A., Puente, C., Tejado, A. (2005): Searching for causal relations in text documents for ontological application. Proc. of the Int. Conf. on Artificial Intelligence IC-AI'05. CSREA Press, vol. II. 463–468.
40. Pasi, G. (2002). Flexible information retrieval: some research trends, Mathware and Soft Computing 9, 107–121.
41. Perkovitz, M., Etzioni, O. (2000). Towards adaptive web sites: Conceptual framework and case study, Artificial Intelligence 118, 245–275.

42. Ramakrishnan, G., Prithviraj, B.P., Deepa E. (2004): Soft word disambiguation, Second Global WordNet Conference, 2004.
43. Ricarte, I., Gomide, F. (2001). A reference model for intelligent information search, Proc. of the 1st BISC Int. Workshop on Fuzzy Logic and the Internet, UC Berkeley, USA, 80–85.
44. Salton, G., Wang, A., Yang, C.S.A. (1975). Vector space model for automatic indexing, Communications of the ACM 18, 613–620.
45. Salton, G., Fox, E. A., Wu, H. (1983): Extended Boolean information retrieval. Communications of the ACM, 26(11), 1022–1036.
46. Salton, G., McGil, M. J. (1983): Introduction to Modern Information Retrieval, McGrawHill, Book Company, New York.
47. Selberg, E., Etzioni, O. (2000): On the instability of web search engines, in Proc. of Content-Based Multimedia Information Access (RIAO), Paris, France, 223–235.
48. Shavlik, J., Eliassi-Rad, T. (1998): Building intelligent agents for web-based tasks: A theory-Refinement approach, in Proceedings of the Conference on Automated Learning and Discovery: Workshop on Learning from Text and the Web, Pittsburgh, PA.
49. Tang, Y., Zhang, Y. (2001). Personalized library search agents using data mining techniques, Proc. of the 1st BISC Int. Workshop on Fuzzy Logic and the Internet, 119–124.
50. Terveen, L., Hill, W.; Amento, B., McDonald, D., Creter, J. (1997): Phoaks: a system for sharing recommendations, Communications of the ACM, Vol. 40, No. 3, 59–62.
51. Trillas, E. (2004): Lógica borrosa y narrativa: un párrafo de Vila-Matas, Actas del XII Congreso Español sobre Tecnologías y Lógica Fuzzy ESTYLF'04, Universidad de Jaen, España, 15–20.
52. Vogt, C. C. (1999): Adaptive Combination of Evidence for Information Retrieval, PhD thesis, University of California, San Diego.
53. Whaley, J.M. (1999): An application of word sense disambiguation to information retrieval, Dartmouth College Computer Science Technical Report PCS-TR99-352.
54. Widyantoro, D., Yen, J. (2001): Incorporating fuzzy ontology of term relations in a search engine, Proc. of the 1st BISC Int. Workshop on Fuzzy Logic and the Internet, UC Berkeley, USA, 155–160.
55. Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man and Cybernetics, Vol. 18, No. 1, 183–190.
56. Zadeh, L. A. (2003). From search engines to Question-Answering system: The need for new tools, E. Menasalvas, J. Segovia, P.S. Szczepaniak (Eds.): Proc. of the Atlantic Web Intelligence Conference - AWIC'2003, LNCS, Springer.
57. Zamir, O., Etzioni, O. (1999). Grouper: A dynamic clustering interface to web search results, Proceedings of the WWW8. http:// ai.rightnow.com/ colloquium/papers.php#9-1-00. Last visit September 2006.
58. Zamir, O., Etzioni, O.: (2001): Grouper: A dynamic clustering interface to Web search results. Accessible at http:// ai.rightnow.com/ colloquium/papers. php. Last visit September 2006.

# Fuzzy Set Techniques in E-Service Applications

**Jie Lu, Da Ruan and Guangquan Zhang**

**Abstract** E-services involve various types, delivery systems, advanced information technologies, methodologies and applications of online services that are provided by e-government, e-business, e-commerce, e-market, e-finance, e-learning systems, to name a few. They offer great opportunities and challenges for many areas, such as government, business, commerce, marketing, finance and education. E-service intelligence is a new research field that deals with fundamental roles, social impacts and practical applications of various intelligent technologies on the Internet based e-services. This chapter aims to offer a thorough introduction and systematic overview of the new field e-service intelligence mainly based on fuzzy set related techniques. It covers the state-of-the-art of the research and development in various aspects including both theorems and applications, of e-service intelligence by applying fuzzy set theory. Moreover, it demonstrates how adaptations of existing intelligent technologies benefit from the development of e-service applications in online customer decision, personalised services, web mining, online searching/data retrieval, online pattern recognition/image processing, and web-based e-logistics/planning.

## 1 Introduction

Electronic-service (e-service) intelligence is a new research field that deals with fundamental roles, social impacts and practical applications of various intelligent technologies and methodologies on the Internet based e-services. Over the last decade, many government and business online services have mainly gone through four stages in most industrialized countries: (a) online information presentation, (b) online transaction, (c) online information integration, and (d) intelligent e-services. Clearly, the keyword *intelligence* will be the next paradigm shift in the e-services thanks to web technological advances (Lu et al. 2006b). To provide intelligence for e-services, various intelligent technologies including fuzzy logic, expert systems, machine learning, and neural network etc. are being applied in various e-service approaches, systems and applications. In the framework of intelligent technologies, government and business e-services will provide with a much higher quality for online information presentation, online information searching, personalized recommendation, website evaluation, and customer decision support

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

(Lu et al. 2006a). We begin to see some successful developments recently in applying intelligent techniques to build intelligent e-service systems, such as intelligent recommender systems, intelligent e-shopping systems, intelligent online customer management systems and intelligent online decision support systems. We have also found some successful investigations based on intelligent systems to evaluate e-service systems, help users' online trading and support users' online decision making. In the following we describe the application of fuzzy set based intelligent e-service methods and systems to the World Wide Web.

This chapter presents a thorough introduction and systematic overview of the new field e-service intelligence mainly based on fuzzy set related techniques. The chapter is organised as follows. Section 2 summarises the role of fuzzy set techniques in e-services. Section 3 highlights applications of fuzzy set techniques in web information presentation, online search, and web mining. Section 4 discusses the implementation of personalization and trust in e-services supported by fuzzy logic approaches. Section 5 shows how fuzzy set techniques can help e-service evaluation and knowledge discovery. Section 6 displays intelligent e-service systems with the help of fuzzy set techniques, and finally Sect. 7 concludes the chapter and its related future research direction.

## 2 The Role of Fuzzy Set Techniques in E-Services

Soft computing (SC) aims to use learning, adaptive, or evolutionary computation to create programs. Expert systems, artificial neural networks, fuzzy logic systems and evolutionary computation are major technologies used in SC and/or intelligent systems. The power of each methodology as a design tool is limited only by the designer's imagination. Two features, in particular, stand out: (1) many of them are biologically inspired, and (2) they are all capable of solving non-linear problems (Ruan 1997). The methodologies comprising SC techniques are for the most part complementary and synergistic rather than competitive. SC has already enjoyed considerable success in e-services, which has proven to be instructive and vital (Lu et al. 2006a).

Fuzzy logic (FL) (Zadeh 1965, 1996) is designed to handle imprecise linguistic concepts such as *small, big, low, high, young*, or *old.* Systems based on FL exhibit an inherent flexibility and have proven to be successful in a variety of industrial control and pattern-recognition tasks ranging from handwriting recognition to traffic control. Central to the flexibility that FL provides is the notion of fuzzy sets. Fuzzy sets are the basic concept supporting fuzzy theory. The main research fields in fuzzy theory are fuzzy sets, fuzzy logic, and fuzzy measure. Fuzzy reasoning or approximate reasoning is an application of fuzzy logic to knowledge processing. Fuzzy control is an application of fuzzy reasoning to control. One of the main strengths of FL compared with other schemes to deal with imprecise data is that their knowledge bases, which are in a rule format, are easy to examine and understand. This rule format also makes it easy to update and maintain the knowledge base. Experts think in imprecise terms, such as *very often* and *almost never, usually* and *hardly ever*, *frequently* and *occasionally*, and use linguistic variables such as the above-mentioned *small, big, low* and *high* etc. FL provides a means to compute with words. It concentrates on the use of fuzzy values that capture the meaning of words, human reasoning and decision making, and provides a way of breaking through the computational burden of traditional expert systems. As for the limitations of FL, the main shortcoming is that the membership functions and rules have to be specified manually. Determining membership functions can be a time-consuming, trial-and-error process. Moreover the elicitation of rules from human experts can be an expensive, error-prone procedure. FL as an additional useful tool has nevertheless been successfully applied to some of the most interesting e-service areas described over the rest of the chapter.

## 3 Web Information Presentations, Search, and Mining

The rapid growth of web based e-service applications has provided a better information presentation and faster web search and mining to web users. Fuzzy set techniques, with their abilities to provide a very natural representation of human conceptualization, are making many more useful contributions to improve the quality to e-service applications. Clearly, the presentation for information and searching on products/services will benefit from the facility of linguistic descriptions and partial matching. Fuzzy set theory has a great potential for supporting such kinds of linguistic description and partial matching. Also, fuzzy sets based on intelligent agents can be used in e-services as a means adding in the current web information presentation.

Applying fuzzy set theory can improve the quality of web information search and retrieval in several ways. One can add a fuzzy quantifier to each term or concept. In addition, one can interpret the AND as fuzzy-MIN and OR as fuzzy-MAX functions. Alternatively, one can add agents in the user interface and assign certain tasks to them or use machine learning to learn user behavior or preferences to improve performance (Nikravesh 2006). A Conceptual Fuzzy Search (CFS) model developed by (Nikravesh 2006) can be used for intelligent information and knowledge retrieval through conceptual matching of both text and images. The selected query does not need to match the decision criteria exactly, which gives the system a more human-like behavior. The CFS can also be used for constructing fuzzy ontology or terms related to the context of search or query to resolve the ambiguity.

Fuzzy set techniques have also been applied into web mining and web user profile clustering. An important aspect of customer profiles in online customer relationship management is the clustering of similar profiles to create customer "segments" (Mobasher et al. 2000). Clustered user profiles are a good option when there exist insufficient data to build individual profiles. Most customer attributes such as "quality-conscious" and "calorie-conscious" are inherently fuzzy. Also, customer segments are not crisp. Thus, fuzzy set theory can play a major role in customer profile representations (Jain and Krishnapuram 2001). Bautista et al. (2000) used a genetic algorithm to build an adaptive consumer profile based on documents retrieved by users. A fuzzy classification and a genetic term-selection process provide a better utilization of valuable knowledge to learn the current and future interests of users.

Krishnapuram et al. (2001) introduced the notion of uncertainty in web usage mining, discovering clusters of user session profiles using robust fuzzy algorithms. In their approach, a user or a page can be assigned to more than one cluster. After pre-processing the log data, they created a dissimilarity matrix that is used by the fuzzy algorithms presented in order to cluster typical user sessions. To achieve this, they introduced a similarity measure that takes into account both the individual URLs in a web session, as well as the structure of the site.

# 4 Personalizations and Trust in E-Services

In an online environment, customers can be individually identified, observed, analyzed, and addressed easily. While the initial research efforts have been directed more towards contents, e-service is expected to dominate personalization efforts over the next years. Personalization includes customization of all interactions between the customer and the Internet intermediary (Chen et al. 2004). Web-based personalized service, also called e-service personalization (ESP), is concerned with building a closer relationship and understanding the needs of individual users to deliver right information to right users at the right time. ESP is experiencing widespread adoption in the application areas such as customer relationship management, e-commerce interaction and intimacy, and employee development (Lu et al. 2006a; Adomavicius and Tuzhilin 2005). For example, given a customer, how to pick the right advertisement to target? How to determine which product should be recommended to a customer? How to determine the content of a Web page that a customer views? Many researchers have recently endeavoured to provide personalization mechanisms for e-service. Personalized e-service has recently received considerable attention because online information and services provided need different users (Adomavicius and Tuzhilin 2005; Eirinaki and Vazirgiannis 2003).

As one of the most popular applications of personalization techniques, recommender systems have gained much attention in the past 10 years (Adomavicius and Tuzhilin 2005). Recommender systems aim at filtering out the uninterested items (or predicting the interested ones) automatically on behalf of the users according to their personal preferences. A recommendation system (Amoroso and Reinig 2004) considers user preferences, interests, or browsing behaviors when analyzing user behaviors for personalized services. It therefore can either predict whether a particular user will like a particular item, or to identify a set of items that will be of interest to a certain user (Karypis 2001).

Various approaches for recommender systems have been developed (Breese et al. 1998; Burke 2000, 2002; Zeng et al. 2004). The main types of these approaches adopted in recommender systems are the content-based (CB) approach, the collaborative filtering (CF) approach, the knowledge-based (KB) approach, and the hybrid approach. The CB approach mainly relies on the content and relevant profiles to generate personalized recommendations. Using the approach, a recommender system recommends some web objects, to a user, which are similar to what the user has been interested in the past (Mooney and Roy 1999). The CF approach offers recommendations based on the similarity of a group of users (Mobasher et al. 2000). The CF approach has been known to be the most popular recommendation approach. It has been used in various e-service applications such as recommending web pages, movies, articles and products. The CF approach can be divided into two types: user-based CF and item-based CF (Karypis 2001). The user-based CF approach is implemented in two main steps: (1) a set of k-nearest neighbours of a target user is selected. This is performed by computing correlations or similarities between user profiles and a target user; (2) producing a prediction value for the target user on unrated (or unvisited) items, and generating recommendations to the target user. The item-based CF approach first considers the relationships among items. Rather than finding user neighbours, the system attempts to find k similar items that are rated (or visited) by different users in some similar ways. Then, for a target item, related predictions can be generated. For example, by taking a weighted average of a target user's item ratings (or weights) on these neighbour items. The third type is KB approach; a new approach is new comparing with others. A knowledge-based recommender system attempts to suggest items based on inferences about a user's preferences. Such systems use knowledge in relevant to users and items to generate recommendations. In some sense, all recommendation techniques could be described as doing some kinds of inference. A knowledge-based recommender system avoids gather information about a particular user because its judgments are independent of individual taste (Burke 2000). Some of these systems employ the techniques of case-based reasoning for knowledge-based recommendation, such as Wasabi Personal Shopper (Burke 1999), a restaurant recommender system. All of the known recommendation techniques have strengths and weaknesses. A common thread in recommender system research is the need to combine recommendation techniques to achieve peak performance. For example, Fab (Balabanovic and Shoham 1997) implements a hybrid CB-CF system for recommending web pages. In Fab, user profiles based on the pages a user liked are maintained by using CB techniques. The profiles are directly compared to determine similarity between users in order to make CF predictions. Entrée (Burke 2002) initially is a knowledge-based recommender system that uses CBR techniques to select and rank restaurant. It was implemented to serve as a guide to attendees of s serial of conferences in 1996. Electronic Funding Information (ELFI) (Mooney and Roy 1999) provides suggestions on funding programs and agencies for researches. It is a hybrid recommender system by combining CB and KB approaches. ELFI is a web-based system that provides information about research funding. Guo and Lu (2006) proposed a hybrid recommendation approach by integrating item-based CF approach with semantic similarity analysis techniques.

There are still some problems with current recommendation approaches including the lack of scalability and sparsity, prediction accuracy, and lack the ability to provide recommendations or predictions for new users and new items (Guo and Lu 2006). Fuzzy approximate reasoning (Klir and Yuan 1995), fuzzy matching, and fuzzy similarity are being used in recommendation approaches to overcome these existing problems. Nasraoui and Petenes (2003) investigated the framework and presented an approach to provide a dynamic prediction in the web navigation space. Yager (2003) described a reclusive approach in which fuzzy set methods are used for the representation and subsequent construction of justifications and recommendation rules. Differing from CF, it is based solely on preferences of the single individuals for whom we provide the recommendation, without using preferences of other collaborators. It makes extensively use of an internal description of the items, and relies solely on the preferences of the target user. Carbo and Molina (2004) developed a CF-based algorithm in which ratings and recommendations can be linguistic labels represented by fuzzy sets. Perny and Zucker (1999, 2001) proposed a recommender system from a decision support perspective, noting that such applications position themselves between the archetypical problems of individual and group decision making. In that light, they pursued a hybrid approach that involves a number of fuzzy relations. Using appropriate fuzzy similarity measures, for each item $i$, and each user $u$, a neighbourhood of $k$ most similar elements is constructed and denoted $N_k(i)$, respectively $N_k(u)$; thanks to the use of neighborhoods, the entire search space does not need to be traversed in producing recommendations. Next, $Q\ (u,\ i)$ can be a self- or peer-evaluation of the confidence about $u$'s rating of $i$, to strengthen or diminish its impact in the generation of recommendations. Based on fuzzy similarity measures, a hybrid recommendation algorithm with fuzzy set theory was proposed. It is being used in a one-and-only item recommendation system in government e-services (Cornelis et al. 2005).

Under these general framework and approaches, some fuzzy techniques based recommender systems have been developed. For example, a personalized courseware recommendation system (PCRS) is proposed by Chen et al. (2004). This system is developed based on the fuzzy item response theory to provide web-based learning services. In the proposed fuzzy item response theory, the fuzzy theory is combined with the original item response theory (Baker and Frank 1992) to model uncertainly learning response. The PCRS can dynamically estimate learner ability based on the proposed fuzzy item response theory by collecting learner feedback information after studying the recommended courseware. Experiments show that the proposed fuzzy technology based personalized courseware recommendation system can recommend appropriate course materials to learners based on the individual ability, and help them to learn more efficiently and effectively.

The applications of fuzzy sets in e-service personalization also conduct some personalized e-service models. For example, Viswanathan and Childers (1999) considered online product categories as fuzzy sets. Products are said to have degrees of memberships in specific attributes. The memberships at the attribute level are then combined to obtain an overall degree of memberships of a product in a category. Fuzzy-set-based measures enable fine distinction among products and assist in the new product development, brand extension, and brand positioning. Another example is about a competitive structure. Fuzzy methods are useful in modeling market structure since they can handle the uncertainty associated with consumer choice and their next purchase (Nishio and Shiizuka 1995).

E-service has managed to place itself in the society. However, there are many hindrance factors that cause it to fail to reach its full potential, mainly on the dissatisfaction of customers, such as a low level of personal data security and mistrust of the technology (Manchala 2000). This has affected consumers' trust towards online business. Since the concept of trust is subjective, it creates a number of unique problems that obviate any clear mathematical result. Hence, fuzzy logic is currently being investigated as a possible best fit approach as it takes into account the uncertainties within e-commerce data and like human relationships, trust is often expressed by linguistic terms rather than numerical values. Nefti et al. (2005) identified two advantages of using fuzzy-logic to quantify trust in e-commerce applications. (1) Fuzzy inference is capable of quantifying imprecise data and quantifying uncertainty in measuring the trust index of the vendors. For example, in the trust model, the community comments variable in the fulfillment factor has a wide range of values as we may have a small or large number of customers providing positive or negative feedback to the vendor; the number of comments will affect the decision made by the associated evaluation module. (2) Fuzzy inference can deal with variable dependencies in the system by decoupling dependable variables. The membership functions can be used to generate membership degrees for each variable. Any defined fuzzy rule set will be applied to the output space (trust index) through fuzzy '*and*' and '*or*' operators. Such a fuzzy trust module can describe more effectively users' trust behavior in e-services.

# 5 E-Service Evaluations and Knowledge Discovery

Since the mid-1990s, businesses have spent quite a bit of time, money and effort developing web-based e-service applications. These applications are assisting businesses in building more effective customer relationships and gaining competitive advantage through providing interactive, personalized, faster e-services to fulfill customer demands (Chidambaram 2001). Businesses in the earlier stages of employing web-based applications had little data, knowledge and experience for assessing and evaluating the potential of e-services for organizational impacts and benefits. Organisational efforts were largely geared toward customer service provision with little to no thought identifying and measuring the costs involved in moving services online against the benefits received by adopting e-services. After several years experience of e-service provision, businesses now urgently need to plan their further development in e-services (Lu 2003). Importantly, businesses have obtained related e-service running data and knowledge, which makes it possible to identify in what items of investment for an e-service application effectively contribute to what benefit aspects of business objectives.

Recent reports concerning the success, quality, usability and benefit of e-services have led researchers to express increasing interest in evaluating and measuring the development of e-service applications (Wade and Nevo, 2005). Much research has been conducted to evaluate e-services from various views and using various methods. In general, the research in e-service evaluation can be classified under four major categories.

The first one is the evaluation for the features, functions or usability of e-service systems. It is often combined with the evaluation of the use of related websites.

Typical approaches used in this category of research are testing, inspection and inquiry (Hahn and Kauffman 2002). These approaches are often used together in analyzing a web search or a desk survey. For example, Ng et al. (1998) reported a desk survey of business websites and discussed the features and benefits of web-based applications. Smith (2001) proposed a set of evaluation criteria to New Zealand government websites. Lu et al. (2001) showed their assessment results for e-commerce development in the businesses of New Zealand.

The quality of websites needs to be measured using criteria focused on the effective website design (e.g., clear ordering of information, consistent navigation structure). However, from the information consumer's perspective the quality of a website may not be assessed independently of the quality of the information content that provides. Based on the information quality framework for the design of information systems defined Lee et al. (2002), Enrique et al. (2003) presented a computing-with-words based fuzzy method to measure the informative quality of Web sites used to publish information stored in XML documents.

The second category is the customer satisfactory evaluation. Various evaluation criteria and factors have been identified and related evaluation systems have been designed for obtaining customers' feedback and measuring the degree of their satisfaction to current e-services provided (Lu and Lu 2004). Questionnaire-based survey and multi-criteria evaluation systems are mainly used to conduct this kind of research. For example, Lin (2003) examined some customer satisfaction for e-commerce and proposes three main scales that play a significant role in influencing customer satisfaction: customer need, customer value and customer cost. In the meantime, a related topic, customer loyalty, has also been paid attentions. For example, Srinivasan et al (2002) explored the antecedents and consequences of customer loyalty in e-commerce.

During the evaluation of e-services, in particular of user satisfactory evaluation, fuzzy set techniques have been extended to discovery of fuzzy association rules (Kuok et al. 1998) and their extension to fuzzy sequential pattems (Hong et al. 1999). Fuzzy set theory provides a host of parameterized operators that can be used to model various aggregation strategies in web-based knowledge discovery (Jain and Krishnapuram 2001). Mela and Lehmann (1995) established a parametric link between fuzzy set theoretic techniques and commonly used preference formation rules in psychology and marketing. Setnes and Kaymak (2001) described an application of a fuzzy clustering algorithm to extract fuzzy rules from consumer response data collected by a sampling procedure. The rules are used to rank customers and the top n customers are considered targets. Fuzzy Adaptive Resonance (ART) can also be used for clustering customers in groups for targeting (Jain and Krishnapuram 2001).

The third category is e-service investment analysis that has been conducted for evaluating and justifying investment in an e-service application. For example, Giaglis et al. (1999) presented a case study of e-commerce investment evaluation. Furthermore, Drinjak et al. (2001) investigated the perceived business benefits of investing in e-service applications. While Amir et al. (2000) created a cost-benefit framework for online system management and evaluation. In particular, Lu and Zhang (2003) proposed a cost benefit factor analysis model in e-services and conducted analysis for e-service development of businesses in Australia based on a questionnaire survey.

Fuzzy set approaches have been used to summarise and analyse the survey results in the e-service evaluation in the form of linguistic knowledge that can be understood by merchants easily. Fuzzy set techniques are known to be effective for analysis even with sparse data especially when application-specific knowledge is available in terms of fuzzy rules. Hsu et al. (2002) proposed a fuzzy clustering approach for segment structure analysis of customer response to surveys.

Significant results have also been reported in the fourth category, the establishment of evaluation models, frameworks and systems. For example, Lee et al. (1999) created a model for evaluating the business value of business-to-business e-service through five propositions. Zhang and Dran (2000) developed a two-factor model for the website design and evaluation. More generally, Hahn and Kauffman (2002) presented a value-driven framework for evaluating e-commerce websites. A web-based fuzzy multi-criteria group decision support system has been developed by Lu et al. (2005) and used for the website evaluation. A group of users can use the online decision support system to input linguistic terms such as '*good*', '*very good*' for a set of selected websites respectively. Finally evaluation results will show the level of the user satisfactory for each attribute of each website.

## 6 Intelligent E-Service Systems

As e-services become common a large number of Internet-based systems have been developed to assist users in different aspects of e-services. Many software agents and other kinds of intelligent systems have been developed for a web-based framework to mainly perform tasks of intermediation and communication between users and the web (Yager 2000).

One possibility to facilitate the communication processes consists in the application of the fuzzy linguistic approach (Zadeh 1975), which provides a flexible representation model of information by means of linguistic labels. The application of fuzzy linguistic techniques enables e-service providers to handle information with several degrees of truth and solving the problem of quantifying qualitative concepts. Some examples of the use of fuzzy linguistic techniques in the design of intelligent e-service systems, in particular multi-agent systems, can be found in Delgado et al. (2001), Delgado et al. (2002) and Herrera-Viedma et al. (2006). These papers presented some new models of fuzzy linguistic intelligent systems that involve the use of fuzzy set techniques and other intelligent approaches to improve the information access on the Web. For example, a fuzzy linguistic multi-agent system can gather information on the Web with a hierarchical architecture of seven action levels (Herrera-Viedma et al. 2004) that improves information retrieval by means of the application of user profiles to improve the filtering activity.

E-negotiation is a typical example of intelligent e-service systems. Kowalczyk and Bui (2000) presented some aspects of a customable fuzzy e-negotiation agents (FeNAs) system for autonomous multi-issue negotiation in the presence of limited common knowledge and imprecise/soft constraints and preferences. The FeNAs use the principles of utility theory and fuzzy constraint-based reasoning in order to find a consensus that maximizes the agent's utility at the highest possible level of fuzzy constraint satisfaction subject to its acceptability by other agents. This information can be imprecise where constraints, preferences and priorities are defined as fuzzy constraints describing the level of satisfaction of an agent (and its user) with different potential solutions. The approach used by the FeNAs is based on modeling and solving negotiation as a fuzzy constraint satisfaction problem (FCSP) (Dubois et al. 1994; Kowalczyk 1999; Zadeh 1973). Finally, a consensus that maximizes the agent's utility at the highest possible level of fuzzy constraint satisfaction subject to its acceptability by other agents. Through applying these fuzzy set approaches, a variety of e-negotiation problems with incomplete common knowledge and imprecise/soft constraints can be handled.

Another example of fuzzy set techniques based e-service systems is an e-learning system. With the rapid growth of computer and Internet technologies, e-learning has currently become a major trend in the computer assisted teaching and learning field. Many researchers made efforts in developing e-learning systems to assist on-line learning. To promote learning efficiency and effectiveness, some systems applied fuzzy set approaches to fully consider learner's behaviors, interests, or habits and also to assist learners in selecting subject, topics, materials with an appropriate difficult level to learners through learners gives a fuzzy response of understanding percentage for the learned courseware. Results show that applying the proposed fuzzy set approaches to e-learning can achieve personalized learning and help learners to learn more effectively and efficiently.

The rapidly developing e-service activity of automated procurement will benefit from the kinds of intelligent decision systems that can be constructed using fuzzy technology (Yager 2000). Ngai and Wat (2005) developed a fuzzy decision support system for risk analysis in e-commerce development. Lu et al. (2005) developed a web-based fuzzy group decision support system (WFGDSS) based on the fuzzy group decision making method. This system first identified three factors from web users that may influence the assessment of utility of alternatives and the deriving of the group satisfactory solution. The first one is an individual's role (weight) in the ranking and selection of the satisfactory solutions. The second factor is an individual's preference for alternatives. The third factor is criteria for assessing these alternatives. The above-mentioned three factors also derive a crucial requirement for linguistic information processing techniques in an online group decision-making practice. Any individual role in an online decision process, a preference for alternatives, and a judgment for assessment-criteria are often expressed by linguistic terms. For example, an individual role can be described by using linguistic terms *strongly important person*, *general decision person* or *weakly important person*. Since these linguistic terms reflect the uncertainty, inaccuracy and fuzziness of decision makers, fuzzy set theory (Zadeh 1965) is directly applied to deal with them. The web-based GDSS uses a web environment as a development and delivery platform. This system allows decision makers distributed in different locations to participate in a group decision making activity through the web. It manages the group decision making process through criteria generation, alternative evaluation, opinion interaction and decision aggregation with the use of linguistic terms. This web based GDSS has a convenient and graphical user interface with visualization possibilities, and therefore is automatically available to large number of decision makers.

## 7 Conclusions

E-service intelligence is still in its infancy. The preliminary research seemed promising, but more research and developments should be followed soon. FL plays an important role for dealing with e-services as already briefly outlined by many successful applications in this chapter. We strongly believe the use of FL and related SC techniques in copy with traditional tools will significantly enhance the current development of e-services in general and the future intelligent e-services in particular.

## References


1. Adomavicius G, Tuzhilin A. (2005) Personalization technologies: a process-oriented perspective. Communications of the ACM 48: 83–90
2. Amir Y, Awerbuch B, Borgstrom RS (2000) A cost–benefit framework for online management of a metacomputing system. Decision Support Systems 28: 155–164
3. Amoroso DL, Reinig BA (2004) Personalization management systems: minitrack introduction, Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04). Big Island, Hawaii
4. Balabanovic M, Shoham Y (1997) Fab: content-based, collaborative recommendation. Communications of the ACM. 40: 66–72
5. Baker FB (1992) Item response theory: Parameter estimation techniques, Marcel Dekker, New York
6. Bautista MM, Vila MA, Larsen H (2000) Building adaptive user profiles by a genetic fuzzy classifier with feature selection. Proceedings of the Ninth IEEE International Conference on Fuzzy Systems. pp. 308–312
7. Breese JS, Heckerman D, Kadie C (1998) Empirical analysis of predictive algorithms for collaborative filtering. Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. Madison, Wisconsin, USA
8. Burke R (1999) The Wasabi personal shopper: a case-based recommender systems. Proceedings of the 11th Annual Conference on Innovative Applications of Artificial Intelligence. Menlo Park, CA, pp. 844–849
9. Burke R (2000) Knowledge-based recommender systems. In: Kent A Ed Encyclopedia of Library and Information Systems. Marcel Dekker, New York
10. Burke R (2002) Hybrid recommender systems: survey and experiments. User Modelling and User-Adapted Interaction 12: 331–370
11. Carbo J, Molina JM (2004) Agent-based collaborative filtering based on fuzzy recommendations. International Journal of Web Engineering and Technology 1: 414–426
12. Chen CM, Duh LJ, Liu CY (2004) A personalized courseware recommendation system based on fuzzy item response theory. Proceedings of 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04). Taipei, Taiwan, pp. 305–308
13. Chidambaram L (2001) The editor's column: why e-Service., e-Service Journal 1: 1–3

14. Cornelis C, Guo X, Lu J, Zhang GQ (2005) A fuzzy relational approach to event recommendation. Proceedings of the Second Indian International Conference on Artificial Intelligence (IICAI-05). Pune, India, pp. 2231–2242
15. Delgado M, Herrera F, Herrera-Viedma E, Martín-Bautista MJ, Martínez L, Vila MA (2002) A communication model based on the 2-tuple fuzzy linguistic representation for a distributed intelligent agent system on Internet. Soft Computing 6: 320–328
16. Delgado M, Herrera F, Herrera-Viedma E, Martín-Bautista MJ, Vila MA (2001) Combining linguistic information in a distributed intelligent agent model for information gathering on the Internet. In: Wang PP Ed Computing with Word. John Wiley & Son, pp. 251–276
17. Drinjak J, Altmann G, Joyce P (2001) Justifying investments in electronic commerce. Proceedings of the Twelfth Australia conference on Information Systems. Coffs Habour, Australia, pp. 187–198
18. Dubois D, Fargier H, Prade H (1994) Propagation and satisfaction of flexible constraints. In: Yager RR, Zadeh LA (Eds) Fuzzy Sets, Neural Networks arid Sott Computing. Van Nostrand Reinhold, New York, pp. 166–187
19. Eirinaki M, Vazirgiannis M (2003) Web mining for web personalization. ACM Transactions on Internet Technology 3: 1–27
20. Enrique H, Eduardo P, María DO, Juan CH, Yusef HM (2003) Evaluating the informative quality of web sites by fuzzy computing with words. In: Zhong N Ed Advances in Web Intelligence: First International AtlanticWeb Intelligence Conference AWIC 2003. Springer Berlin/Heidelberg, New York, pp. 62–72
21. Giaglis GM, Paul RJ, Doukidis GI (1999) Dynamic modelling to assess the business value of electronic commerce. International Journal of Electronic Commerce 3: 35–51
22. Guo X, Lu J (2006) Intelligent e-government services with recommendation techniques. International Journal of Intelligent Systems, Special issue on E-service intelligence. Accepted
23. Hahn J, Kauffman RJ (2002) Evaluating selling web site performance from a business value perspective. Proceedings of International conference on e-Business. Beijing, China, pp. 435–443
24. Herrera-Viedma E, Herrera F, Martínez L, Herrera JC, López AG (2004) Incorporating filtering techniques in a fuzzy multi-agent model for gathering of information on the web. Fuzzy Sets and Systems 148: 61–83
25. Herrera-Viedma E, Peis E, Morales-Del-Castillo JM, Anaya K (2006) Improvement of web-based service information systems using fuzzy linguistic techniques and semantic web technologies. In: Lu J, Ruan D, Zhang GQ (Eds) E-Service Intelligence: Methodologies, Technologies and Applications. Springer, Heideberg
26. Hong T, Kuo C, Chi S (1999) Mining fuzzy sequential patterns from quantitative data. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. pp. 962–966
27. Hsu T, Chu KM, Chan HC (2002) The fuzzy clustering on market segment. Proceedings of the Ninth IEEE International Conference on Fuzzy Systems. pp. 62 1–626
28. Jain V, Krishnapuram R (2001) Applications of fuzzy sets in personalization for e-commerce. Proceedings of IFSA-NAFIPS 2001 Conference. pp. 263–268
29. Karypis G (2001) Evaluation of item-based top-N recommendation algorithms. Proceedings of the ACM 10th International Conference on Information and Knowledge Management. Atlanta, Georgia, pp. 247–254
30. Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic: theory and applications. Upper Saddle River, Prentice Hall, NJ
31. Kowalczyk R (1999) On linguistic fuzzy constraint satisfaction problems. In: Zadeh LA, Kacprzyk J (Eds) Computing with Words in Intelligent Information Systems. Kluwer, pp. 166–187
32. Kowalczyk R, Bui V (2000) FeNAs: a fuzzy e-Negotiation agents system. Proceedings of the Conference on Computational Intelligence for Financial Engineering. NY, pp. 26–29
33. Krishnapuram R, Joshi A, Nasraoui O, Yi, L (2001) Low complexity fuzzy relational clustering algorithms for web mining. IEEE Transactions on Fuzzy Systems 9: 595–607

34. Kuok C, Fu A, Wong MH (1998) Mining fizzy association rules in databases. SIGMOD Record 27: 41–46
35. Lee C, Seddon P, Corbitt B (1999) Evaluating business value of internet-based business-to-business electronic commerce. Proceedings of 10th Australasian Conference on Information Systems. pp. 508–519
36. Lee YW, Strong DM, Kahn BK, Wang RY (2002) AIMQ: a methodology for information quality assessment. Inf. & Manag. 40: 133–146
37. Lin C (2003) A critical appraisal of customer satisfaction and e-commerce. Managerial Auditing Journal 18: 202–212
38. Lu J (2003) A model for evaluating e-commerce based on cost/benefit and customer satisfaction. Journal of Information Systems Frontiers 5: 265–277
39. Lu J, Lu Z (2004) Development, distribution and evaluation for online tourism services in China. Electronic Commerce Research Journal 4: 221–239
40. Lu J, Ruan D, Zhang GQ (2006a) E-service intelligence: methodologies, technologies and applications, Springer, Heideberg, to appear
41. Lu J, Ruan D, Zhang GQ, Zimmermann HJ (2006) Editorial preface. International Journal of Intelligent Systems. Special Issue on E-Service Intelligence, to appear
42. Lu J, Tang S, Mccullough G (2001) An assessment for internet-based electronic commerce development in businesses of New Zealand electronic markets. International Journal of Electronic Commerce and Business Media 11: 107–115
43. Lu J, Zhang GQ (2003) Cost benefit factor analysis in e-services. International Journal of Service Industry Management 14: 570–595
44. Lu J, Zhang GQ, Wu F (2005) Web-based multi-criteria group decision support system with linguistic term processing function. The IEEE Intelligent Informatics Bulletin 5: 35–43
45. Manchala DW (2000) E-commerce trust metrics and models. IEEE Internet Computing 4: 36–44
46. Mela C, Lehmann DR (1995) Using fuzzy set theoretic techniques to identify preference rules from interactions in the linear model: an empirical study. Fuzzy Sets and Systems 71: 165–181
47. Mobasher B, Dai H, Luo T, Sung Y, Nakagawa M, Wiltshire J (2000) Discovery of aggregate usage profiles for web personalization. Proceedings of the Web Mining for E-Commerce Workshop, WebKDD. Boston, USA
48. Mooney R, Roy L (1999) Content-based book recommending using learning for text categorization. Proceedings of the ACM SIGIR'99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley, CA, USA
49. Nasraoui O, Petenes C (2003) An intelligent web recommendation engine based on fuzzy approximate reasoning. Proceedings of the IEEE International Conference on Fuzzy Systems. pp. 1116–1121
50. Nefti S, Meziane F, Kasiran K (2005) A fuzzy trust model for e-commerce. Proceedings of Seventh IEEE International Conference on E-Commerce Technology (CEC'05). pp. 401–404
51. Ng H, Pan YJ, Wilson TD (1998) Business use of the world wide web: a report on further investigations. International Journal of Management 18: 291–314
52. Ngai EWT, Wat FKT (2005) Fuzzy decision support system for risk analysis in e-commerce development. Decision Support Systems 40: 235–255
53. Nikravesh M (2006) Concept-based semantic web search and Q&A. In: Lu J, Ruan D, Zhang GQ (Eds) E-Service Intelligence: Methodologies, Technologies and Applications. Springer, Heideberg, to appear
54. Nishio C, Shiizuka H (1995) Competitive market structures constructed from brand-switching data by fuzzy structural modeling. Proceedings of IEEE International Conference on Fuzzy Systems. pp. 819–824
55. Perny P, Zucker JD (1999) Collaborative filtering methods based on fuzzy preference relations. Proceedings of the EUROFUSE-SIC. pp. 279–285
56. Perny P, Zucker JD (2001) Preference-based search and machine learning for collaborative filtering: the 'Film-Conseil' movie recommender system. Revue I3, 1: 1–40

57. Ruan D (1997) Intelligent hybrid systems. Kluwer Academic Publishers, Boston
58. Schwab I, Pohl W, Koychev I (2000) Learning to recommend from positive evidence. Proceedings of the 5th International Conference on Intelligent User Interfaces. New Orleans, LA, USA
59. Setnes M, Kaymak U (2001) Fuzzy modeling of client preference from large data sets: an application to target selection in direct marketing. IEEE Transactions on Fuzzy Systems 9: 153–163
60. Smith AG (2001) Applying evaluation criteria to New Zealand government websites. International Journal of Information Management 21: 137–149
61. Srinivasan S, Anderson R, Ponnavolu K (2002) Customer loyalty in e-commerce: an exploration of its antecedents and consequences. Journal of Retailing 78: 41–50
62. Viswanathan M, Childers TL (1999) Understanding how product attributes enhance product categorization: development and validation of fuzzy set-based measures of gradednesss in product categories. Journal of Marketing Research 36: 75–94
63. Wade MR, Nevo S (2005) Development and validation of a perceptual instrument to measure e-commerce performance. International Journal of Electronic Commerce 10: 123
64. Yager RR (2000) Targeted e-commerce marketing using fuzzy intelligent agents. IEEE Intelligent Systems 15: 42–45
65. Yager RR (2003) Fuzzy logic methods in recommender systems. Fuzzy Sets and Systems 136: 133–149
66. Zadeh LA (1965) Fuzzy sets. Inform & Control 8: 338–353
67. Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Man Cybernetics 3: 28–44
68. Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning, Part I Information Sciences 8: 199–249, Part II. Information Sciences 8: 301–357, Part III. Information Sciences 9: 43–80
69. Zadeh LA (1996) Fuzzy logic = computing with words. IEEE Transactions on Fuzzy Systems 4: 103–111
70. Zeng C, Xing CX, Zhou LZ, Zheng XH (2004) Similarity measure and instance selection for collaborative filtering. International Journal of Electronic Commerce 8: 115–129
71. Zhang P, Dran VG (2000) Satisfiers and dissatisfiers: a two-factor model for website design and evaluation. Journal of American Association for Information Science (JASIS) 51: 1253–1268

# A Fuzzy Linguistic Recommender System to Advice Research Resources in University Digital Libraries

**Enrique Herrera-Viedma, Carlos Porcel, Antonio Gabriel López-Herrera and Sergio Alonso**

**Abstract** As it is known the Web is changing the information access processes. The Web is one of the most important information media. Furthermore, the Web is influencing in the development of other information media, as for example, newspapers, journals, books, libraries, etc. In this chapter we analyze its impact in the development of the University Digital Libraries (UDL). As in the Web, the growing of information is the main problem of the academic digital libraries, and similar tools could be applied in university digital libraries to facilitate the information access to the students and teachers. *Filtering systems* or *recommender systems* are tools whose objective is to evaluate and filter the great amount of information available on the Web to assist the users in their information access processes. Therefore, we present a model of fuzzy linguistic recommender system to help students and researchers to find research resources which could improve the services that render the UDL to their users.

## 1 Introduction

In last years the new concept of digital library is growing. **Digital libraries** are information collections that have associated services delivered to user communities using a variety of technologies. The information collections can be scientific, business or personal data, and can be represented as digital text, image, audio, video, or other media. This information can be digitalized paper or born digital material and the services offered on such information can be varied, and can be offered to individuals or user communities. Internet access has resulted in digital libraries that are increasingly used by diverse communities for diverse purposes, and in which sharing and collaboration have become important social elements. As digital libraries become commonplace, as their contents and services become more varied, people expect more sophisticated services from their digital libraries [4, 10, 11, 31].

The digital libraries are composed by human resources (staff) take over handle and enable users the access to the documents more interesting for them, taking into account their needs or interest areas. The library staff searches, evaluates, selects, catalogues, classifies, preserves and schedules the digital documents access [10, 11]. Some of the main digital libraries functions are the following:

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008 

- To evaluate and select digital materials to add in its repository.
- To preserve the security and conservation of the materials.
- To describe and index the new digital materials (catalogue and classify).
- To deliver users the material stored in the library.
- Other managerial tasks.

Digital libraries have been applied in a lot of contexts. We are going to center in an academic environment. **University digital libraries** provide information resources and services to students, faculty and staff in an environment that supports learning, teaching and research [5, 27].

The exponential increase of Web sites and documents is contributing to that Internet users not being able to find the information they seek in a simple and timely manner. Users are in need of tools to help them cope with the large amount of information available on the Web [25, 28]. Therefore, techniques for searching and mining the Web are becoming increasingly vital. Furthermore, the Web is influencing in the development of many organizations, as for example, banks, companies, universities, libraries, etc. In particular, we are interested in the development of academic digital libraries. As in the Web, the exponential growing of information is the main problem of these libraries because the library staff find troubles to perform the task of information delivery to the users. We could use those tools applied successfully in the Web context to solve the new problems appeared in UDL to facilitate the tasks of library staff and therefore, the information access to the students and teachers.

A traditional search function is normally an integral part of any digital library, but however users' frustrations are increased as their needs become more complex and as the volume of information managed by digital libraries increases. Digital libraries must move from being passive, with little adaptation to their users, to being more proactive in offering and tailoring information for individuals and communities, and in supporting community efforts to capture, structure and share knowledge [4, 11, 31]. So, the digital libraries should anticipate the users' needs and recommending about resources that could be interesting for them.

In this paper we study two techniques that applied together can contribute to achieve major advances in the activities of university digital libraries in order to improve their performance:

- **Information Filtering Tools:** An important tool to improve the information access on many environments concerns the way in which it is possible to filter the great amount of information available. Information filtering is a name used to describe a variety of processes involving the delivery of information to people who need it. Operating in textual domains, **filtering systems** or **recommender systems** evaluate and filter the great amount of information available in a specific scope to assist users in their information access processes [13, 32].
- **Fuzzy Linguistic Modeling (FLM):** The great variety of representations and evaluations of the information existing in Internet is the main obstacle to the information handling from what is very important the design of appropriate communication protocol. The problem becomes more noticeable when users take part in the process. This reveals the need of more flexible techniques to the informa-

tion representation and evaluation. To solve this problem we propose the use of **FLM** [16, 17, 34] to represent and handle flexible information by means of linguistic labels.

The paper is structured as follows. Section 2 revises the main aspects and models of information filtering techniques. Section 3 analyzes different approaches of FLM, the 2-tuple FLM [17, 19] and the multi-granular FLM [15, 18]. In Sect. 4 we present a model of fuzzy linguistic recommender systems to advice research resources in UDL. Finally, some concluding remarks are pointed out.

# 2 Preliminaries

## *2.1 Information Filtering*

Information gathering on the Internet is a complex activity. Finding the appropriate information, required for the users, on the World Wide Web is not a simple task. This problem is more acute with the ever increasing use of the Internet. For example, users who subscribe to Internet lists waste a great deal of time reading, viewing or deleting irrelevant e-mail messages. To improve the information access on the Web the users need tools to filter the great amount of information available across the Web. **Information Filtering** (IF) is a name used to describe a variety of processes involving the delivery of information to people who need it. It is a research area that offer tools for discriminating between relevant and irrelevant information by providing personalized assistance for continuous retrieval of information.

IF systems are characterized by [13]:

- applicable for unstructured or semi-structured data (e.g. web documents, e-mail messages),
- based on user profiles,
- handle large amounts of data,
- deal primarily with textual data and
- their objective is to remove irrelevant data from incoming streams of data items.

We can find some of the above features in Information Retrieval (IR) systems, but IF differs from traditional IR in that the users have long-term information needs that are described by means of user profiles, rather than ad-hoc needs that are expressed as queries posed to some IR system [2]. Traditionally IR develops storage, indexing and retrieval technology for textual documents. An user describes his information need in the form of a query to the IR system and the system attempts to find items that match the query within a document store. The information need is usually very dynamic and temporary, i.e., an user issues a query describing an immediate need. Furthermore, IR systems tend to maintain a relatively static store of information. Unlike IR systems, IF systems generally operate on continuous information streams, and always maintain a profile of the user interests needs throughout many uses of the system. As a result, IF systems tend to filter information based on more long-term interests.

Traditionally, these IF systems or recommender systems have fallen into two main categories [13, 29, 32]. **Content-based filtering systems** filter and recommend the information by matching user query terms with the index terms used in the representation of documents, ignoring data from other users. These recommender systems tend to fail when little is known about user information needs, e.g. when the query language is poor. **Collaborative filtering systems** use explicit or implicit preferences from many users to filter and recommend documents to a given user, ignoring the representation of documents. These recommender systems tend to fail when little is known about an user, or when he/she has uncommon interests [29]. In these kind of systems, the users' information preferences can be used to define user profiles that are applied as filters to streams of documents; the recommendations to an user are based on another user's recommendations with similar profiles. Many researchers think that the construction of accurate profiles is a key task and the system's success will depend to a large extent on the ability of the learned profiles to represent the user's preferences [30]. Several researchers are exploring hybrid content-based and collaborative recommender systems to smooth out the disadvantages of each one of them [3, 6, 12, 29].

On the other hand, we should point out that the **matching process** is a main process in the activity of filtering systems. The two major approaches followed in the design and implementation of IF systems to do the matching are the statistical approach and the knowledge based approach [13]. In our system, we have applied the statistical approach. This kind of filtering systems represents the documents and the user profiles as weighted vectors of index terms. To filter the information the system implements a statistical algorithm that computes the similarity of a vector of terms that represents the data item being filtered to an user's profile. The most common algorithm used is the Correlation or the Cosine measure between the user's profile and the document's vector.

The filtering activity is followed by a relevance feedback phase. Relevance feedback is a cyclic process whereby the user feeds back into the system decisions on the relevance of retrieved documents and the system then uses these evaluations to automatically update the user profiles.

Another important aspect that we must have in mind when we design a IF system is the method to gather user information. In order to discriminate between relevant and irrelevant information for an user and to provide him/her personalized information, we must have some information about this user, i.e. we must know the user preferences. Information about user preferences can be obtained in two different ways [13], **implicit** and **explicit mode**, although these ways not be mutually exclusive. The implicit approach is implemented by inference from some kind of observation. The observation is applied to user behavior or to detecting an user's environment (such as bookmarks or visited URL). The user preferences are updated by detecting changes while observing the user. On the other hand, the **explicit** approach, interacts with the users by acquiring feedback on information that is filtered, that is, the user expresses some specifications of what they desire. This last approach is very used. In [9] the personalization in digital libraries is studied. They conclude that the technology is still premature, but the next step of digital libraries services should be oriented towards the automation of the process of constructing of user profiles.

## 2.2 Fuzzy Linguistic Modeling

There are situations in which the information cannot be assessed precisely in a quantitative form but may be in a qualitative one. For example, when attempting to qualify phenomena related to human perception, we are often led to use words in natural language instead of numerical values. In other cases, precise quantitative information cannot be stated because either it is unavailable or the cost for its computation is too high and an "approximate value" can be applicable. The use of Fuzzy Sets Theory has given very good results for modeling qualitative information [34]. **FLM** is a tool based on the concept of **linguistic variable** [34] to deal with qualitative assessments. It has proven to be useful in many problems, e.g., in decision making [16], quality evaluation [24], models of information retrieval [20, 21], clinical decision making [8], political analysis [1], etc.

Next we analyze two FLM that we use in our system, i.e., the 2-tuple FLM [17, 19] and the multi-granular FLM [15, 18, 23].

### The 2-Tuple Fuzzy Linguistic Modeling

The **2-tuple FLM** [17, 19] is a kind of fuzzy linguistic modeling that mainly allows to reduce the loss of information typical of other fuzzy linguistic approaches (classical and ordinal [16, 14, 34]). Its main advantage is that the linguistic computational model based on linguistic 2-tuples can carry out processes of computing with words easier and without loss of information. To define it we have to establish the 2-tuple representation model and the 2-tuple computational model to represent and aggregate the linguistic information, respectively.

Let $S = \{s_0, ..., s_g\}$ be a linguistic term set with odd cardinality ($g + 1$ is the cardinality of $S$), where the mid term represents an assessment of approximately 0.5 and with the rest of the terms being placed symmetrically around it. We assume that the semantics of labels is given by means of triangular membership functions represented by a 3-tuple $(a, b, c)$ and consider all terms distributed on a scale on which a total order is defined $s_i \leq s_j \iff i \leq j$. In this fuzzy linguistic context, if a symbolic method [14, 16] aggregating linguistic information obtains a value $\beta \in [0, g]$, and $\beta \notin \{0, ..., g\}$, then an approximation function is used to express the result in $S$. To do this, we represent $\beta$ as a 2-tuple $(s_i, \alpha_i)$, where:

- $s_i$ represents the linguistic label, and
- $\alpha_i$ is a numerical value expressing the value of the translation from the original result $\beta$ to the closest index label, $i$, in the linguistic term set ($s_i \in S$).

This model defines a set of transformation functions between numeric values and 2-tuples: $\Delta(\beta) = (s_i, \alpha)$ y $\Delta^{-1}(s_i, \alpha) = \beta \in [0, g]$ [17].

The 2-tuple linguistic computational model is defined by presenting the comparison of 2-tuples, a negation operator and aggregation operators of 2-tuples:

1. Negation operator of 2-tuples: $Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha)))$.
2. Comparison of 2-tuples $(s_k, \alpha_1)$ and $(s_l, \alpha_2)$:
   - If $k < l$ then $(s_k, \alpha_1)$ is smaller than $(s_l, \alpha_2)$.
   - If $k = l$ then

   a) if $\alpha_1 = \alpha_2$ then $(s_k, \alpha_1)$ and $(s_l, \alpha_2)$ represent the same information,
   b) if $\alpha_1 < \alpha_2$ then $(s_k, \alpha_1)$ is smaller than $(s_l, \alpha_2)$,
   c) if $\alpha_1 > \alpha_2$ then $(s_k, \alpha_1)$ is bigger than $(s_l, \alpha_2)$.
3. Aggregation operators of 2-tuples. The aggregation of information consists of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a 2-tuple. In the literature we can find many aggregation operators which allow us to combine the information according to different criteria. Using functions $\Delta$ and $\Delta^{-1}$ that transform without loss of information numerical values into linguistic 2-tuples and viceversa, any of the existing aggregation operator can be easily extended for dealing with linguistic 2-tuples. Some examples are the arithmetic mean, the weighted average operator or the linguistic weighted average operator.

### The Multi-Granular Fuzzy Linguistic Modeling

In any fuzzy linguistic approach, an important parameter to determinate is the "granularity of uncertainty", i.e., the cardinality of the linguistic term set $S$ used to express the linguistic information. According to the uncertainty degree that an expert qualifying a phenomenon has on it, the linguistic term set chosen to provide his knowledge will have more or less terms. When different experts have different uncertainty degrees on the phenomenon, then several linguistic term sets with a different granularity of uncertainty are necessary (i.e. multi-granular linguistic information) [15, 18, 23]. The use of different label sets to assess information is also necessary when an expert has to assess different concepts, as for example it happens in information retrieval problems, to evaluate the importance of the query terms and the relevance of the retrieved documents [22]. In such situations, we need tools for the management of multi-granular linguistic information, i.e., we need to define a **multi-granular FLM**. In [15] we define a proposal of multi-granular FLM based on the ordinal FLM [16], and in [18] we define other one based on the 2-tuple FLM. In this paper, we follow that defined in [18] which uses the concept of **Linguistic Hierarchies** [7] to manage the multi-granular linguistic information.

**A linguistic hierarchy** is a set of levels, where each level is a linguistic term set with different granularity from the remaining of levels of the hierarchy [7]. Each level belonging to a linguistic hierarchy is denoted as **l(t,n(t))**, **t** being a number that indicates the level of the hierarchy and **n(t)** the granularity of the linguistic term

**Table 1** Linguistic Hierarchies

| | Level 1 | Level 2 | Level 3 |
|---|---|---|---|
| l(t,n(t)) | l(1,3) | l(2,5) | l(3,9) |
| l(t,n(t)) | l(1,7) | l(2,13) | |

set of the level **t**. Usually, linguistic hierarchies deal with linguistic terms whose membership functions are triangular-shaped, symmetrical and uniformly distributed in [0,1]. In addition, the linguistic term sets have an odd value of granularity representing the central label the value of **indifference** ("approximately 0.5"). The levels belonging to a linguistic hierarchy are ordered according to their granularity, i.e., for two consecutive levels **t** and **t+1**, $n(t+1) > \mathbf{n(t)}$. Therefore, each level $t+1$ provides a linguistic refinement of the previous level $t$.

Generically, we can say that the linguistic term set of level **t+1**, $S^{n(t+1)}$, is obtained from its predecessor level **t**, $S^{n(t)}$ as: $l(t, n(t)) \to l(t+1, 2 \cdot n(t) - 1)$. In Table 1 is presented the granularity needed in each linguistic term set of the level **t** depending on the value **n(t)** defined in the first level (3 and 7 respectively).

In [18] was demonstrated that the linguistic hierarchies are useful to represent the multi-granular linguistic information and allow to combine multi-granular linguistic information without loss of information. To do this, a family of transformation functions between labels from different levels was defined:

**Definition 1.** *Let* $LH = \bigcup_t l(t, n(t))$ *be a linguistic hierarchy whose linguistic term sets are denoted as* $S^{n(t)} = \{s_0^{n(t)}, ..., s_{n(t)-1}^{n(t)}\}$*. The transformation function between a 2-tuple that belongs to level t and another 2-tuple in level* $t' \neq t$ *is defined as:*

$$TF_{t'}^{t} : l(t, n(t)) \longrightarrow l(t', n(t'))$$

$$TF_{t'}^{t}(s_i^{n(t)}, \alpha^{n(t)}) = \Delta\left(\frac{\Delta^{-1}(s_i^{n(t)}, \alpha^{n(t)}) \cdot (n(t') - 1)}{n(t) - 1}\right)$$

As it was pointed out in [18] this family of transformation functions is bijective. This result guarantees the transformations between levels of a linguistic hierarchy are carried out without loss of information.

To define the multi-granular linguistic computational model we select a level to uniform the information (normally the most granularity level is selected) and then we can use the operators defined in 2-tuples model.

# 3 A Recommender System Based on Multi-granular Fuzzy Linguistic Modeling to Advice Research Resources in University Digital Libraries

In this section we present a Recommender System (RS) designed using the **content-based filtering approach** and assuming **a multi-granular FLM**. This RS is applied to advice users on the better research resources that could satisfy their information needs in an university digital library.

The users of an university digital library are usually the students and teachers that access to its information resources. Both manage and spread a lot of information about research information such as electronic books, electronic papers, electronic journals, official dailies and so on. Nowadays this amount of information is growing up and the users of these libraries are in need of automate tools to filter and to spread the information in a simple and timely manner.

We present a RS that follows the content-based approach. Moreover to improve the filtering process we incorporate in the system the possibility to manage multi-granular linguistic information, that is, it uses different label sets to represent the different concepts to be assessed in its recommending activity. Then, the system filters the incoming information stream and delivers it to the suitable researchers or students in accordance with their research areas. The system sends the users a mail with a summarized information about the resources, the calculated relevance degrees of the resources for the users and recommendations about others researchers or students with which they could collaborate.

In that follows, we present the system architecture, the required data structures and how the system works.

## *3.1 System Architecture*

The system architecture is shown in Fig. 1. As we can see in the figure, the system has three main components:

- **Resources management**. This module is the responsible of management the information sources from where the library staff receive all the information about research resources, and obtain an internal representation of these electronic resources. To management the resources, we represented them in accordance with their features (title, author(s), abstract, text, date, type and so on) and their scope, and to obtain this scope representation we use the **UNESCO terminology** for the science and technology [33]. This terminology is composed by three levels and each one is a refinement of the previous level. The first level includes general topics and they are codified by two digits. Each topic includes some disciplines codified by four digits in a second level. The third level is composed by subdisciplines that represent the activities developed in each discipline; these subdisciplines are codified by six digits. We are going to operate with the first and second levels, because we think the third level supply a discrimination level too much high and this could difficult the interaction with the users. Moreover, for each resource we store another kind of information that the system uses in the filtering process.

- **User profiles management**. The users can be researchers or students. In both cases, the system operates with an internal representation of the user's preferences or needs, that is, the system represents the users' preferences through user profiles. To define an user profile we are going to use the basic information about the user and his/her interest topics, defined too by the UNESCO terminology [33], i.e. each user has a list of UNESCO codes according to his/her information needs or interests. The research groups have assigned a set of UNESCO codes that define their research activity. So, initially the systems assign to each research or student the UNESCO codes of the research group which the user belongs. If the user doesn't belong to a group, the library staff assigns him/her the UNESCO codes by hand, in accordance with his/her interest areas. Afterwards the users can update their profiles by a feedback phase in which the users express some explicit specifications of their preferences.
- **Filtering process**. In this phase the system filters the incoming information to deliver it to the fitting users and this process is based in a Matching Process. As our system is a content-based filtering system, it filters the information by matching the terms used in the representation of user profiles against the index terms used in the representation of resources. Later we will study this process in detail taking into account the used data structures.

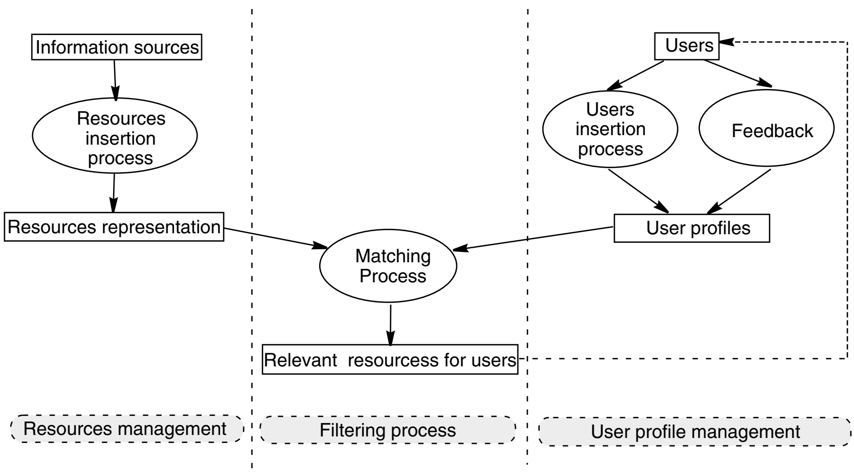



**Fig. 1** Structure of the system

### 3.2 Data Structures

In this subsection we are going to discuss the data structures we need to represent all the information about the users and research resources. We must have in mind that the system stores this information because it doesn't work with explicit user queries.

To characterize a resource, we use the title, author(s), abstract, journal (if is part of a journal, the system stores the journal name), book (if is a book chapter, the system stores the book title), official daily (if is part of an official daily, the system stores the daily title), date, source, text, link (when the system send the users information about a resource, it doesn't send all the information but summarized information and the link to access the resource), kind of resource (if is a book, a paper, a journal, an official daily and so on), target (this field indicates the kind of users that is directed the resource, that is researchers, students or anybody) and scope. To represent the resource scope we use the **vector model** where for each resource the system stores a vector $VR$, i.e. an ordered list of terms. To build this vector we follow the UNESCO terminology [33], specifically we use the second level. This level has 248 disciplines, so the vector must have 248 positions, one position for each discipline. In each position the vector stores the importance degree for the resource scope of the UNESCO code represented in that position.

To characterize an user we must distinguish if is a research or a student, although the system stores the same basic information: user's identity (usually his/her mail), password (necessary to access the system), dni (identity national document), name and surname, department and center (if the user is a students this information is not necessary), address, phone number, mobile phone and fax, web, email (elemental information to send the resources and recommendations), research group (is a string composed by 6 digits, 3 characters indicating the research area and 3 numbers identifying the group; if the user is a students this information is not necessary), collaboration preferences (if the user want collaborate with other researchers of other groups, with students, with anybody or with nobody), preferences about resources (the user choose the kind of desired resources, i.e. if he/she want only books, or papers, etc.) and interest topics. To represent the interest topics we use the vector model too where for each user the system stores a vector $VU$. To build this vector we follow the UNESCO terminology [33], specifically we use the second level. This level has 248 disciplines, so the vector must have 248 positions, one position for each discipline. In each position the vector stores the importance degree for the user research of the UNESCO code represented in that position. With all this information the system sets up the user profiles.

On the other hand, to represent the linguistic information we use different label sets, i.e. the communication among the users and the system is carried out by using multi-granular linguistic information, in order to allow a higher flexibility in the processes of communication of the system. Therefore the system uses different label sets ($S_1$, $S_2$, ...) to represent the different concepts to be assessed in its filtering activity. These label sets $S_i$ are chosen from those label sets that composes a $LH$, i.e., $S_i \in LH$. We should point out that the number of different label sets that we can use is limited by the number of levels of $LH$, and therefore, in many cases the label sets $S_i$ and $S_j$ can be associated to a same label set of $LH$ but with different interpretations depending on the concept to be modeled. In our system, we distinguish between three concepts that can be assessed:

- **Importance degree** ($S_1$) of an UNESCO code with respect to a resource scope or user preferences.
- **Relevance degree** ($S_2$) of a resource for a researcher or for a student.
- **Compatibility degree** ($S_3$) between a researcher and a student, between researchers of different groups and between different students.

In our system we use a linguistic hierarchy of three levels. Specifically we use the level 2 (5 labels) to assign importance degree ($S_1 = S^5$) and the level 3 (9 labels) to assign relevance degrees ($S_2 = S^9$) and compatibility degrees ($S_3 = S^9$). Using this linguistic hierarchy the linguistic terms in each level are:

- $S^3 = \{a_0 = Null = N,\ a_1 = Medium = M,\ a_2 = Total = T\}$.
- $S^5 = \{b_0 = Null = N,\ b_1 = Low = L,\ b_2 = Medium = M,\ b_3 = High = H,\ b_4 = Total = T\}$
- $S^9 = \{c_0 = Null = N,\ c_1 = Very\_Low = VL,\ c_2 = Low = L,\ c_3 = More\_Less\_Low = MLL,\ c_4 = Medium = M,\ c_5 = More\_Less\_High = MLH,\ c_6 = High = H,\ c_7 = Very\_High = VH,\ c_8 = Total = T\}$

Therefore, for a resource $i$ we have a vector representing its scope:

$$VR_i = (VR_{i1}, VR_{i2}, ..., VR_{i248}),$$

where each component $VR_{ij} \in S_1$, with $j = 1..248$, stores a linguistic label indicating the importance degree of the UNESCO code $j$ with regard to the resource $i$. These linguistic labels are assigned by the library staff when they add a new electronic resource.

To represent the interest topics in the user profiles we follow the same method, using a vector $VU$ for each user of the system. Then, for the user $x$, we have a vector:

$$VU_x = (VU_{x1}, VU_{x2}, ..., VU_{x248}),$$

where each component $VU_{xy} \in S_1$, with $y = 1..248$, stores a linguistic label indicating the importance degree of the UNESCO code $y$ with regard to the preferences of the user $x$. These linguistic labels are assigned by the library staff too, but the users can edit it when they want.

### *3.3 Operation of Recommender System*

Any university digital library must provide the next two kind of services [10]:

- **User registration**. The users accesses to the system to solicit the services offered by the university digital library. The system present a form where the users introduce their personal information, their collaboration preferences and their preferences about the kind of resources they want to receive. Finally the users define their interest topics setting up the UNESCO codes and the importance degrees. If the user belongs to a research group, the system shows him/her the UNESCO codes of the group, and the user can edit (add, delete, or assign new degrees) these codes to adjust them to his/her interest areas. The system registers the user and assigns him/her an identifier (usually it uses the mail address) and a password. To conclude the registration process, the system sends the user an email to confirm the inserted information.
- **Information and documents access services**. Once the users have their identifies and passwords, they can use the digital library services. Therefore they can performs their information access processes taking into account their profiles.

Next we describe the users insertion process, the resources insertion process, the filtering process and the feedback phase.

### Users Insertion Process

In order to gather information about users we use a hybrid approach between the explicit and implicit approach. When we insert a new user we use implicit information to generate the profile and afterwards the users can update their profiles following the explicit approach.

So, to add a new user into the system, it shows a form that the user must fill in introducing his/her personal information, collaboration preferences, preferences about the kind of resources he/she want to receive and so on. Then the system defines the user interest topics using the UNESCO codes of the research group which the user belongs. Each group or company has assigned one or more UNESCO codes, so when the system is inserting a new user, it assigns him/her the UNESCO codes of level 2 of the group which the user belong, with importance degree **Total** ($b_4 \in S_1$). The other positions have a value **Null** ($b_0 \in S_1$). The system presents this information to the users who can edit it if they want. The users who don't belong to a research group, must define their profiles manually, that is, they select the UNESCO codes and their importance degrees ($b_i \in S_1$) to establish their interest topics. Later the users can update their profiles always they want, accessing to the system and editing the UNESCO codes or the linguistic labels (in $S_1$) which they have assigned.

With this information the system defines and updates the user profiles which will use to filter information when a new resource arrives to the system.

*Example 1.* In this example we see the process of insertion of a new user. The user inserts all the information about him/her together with the user's identity $\mathcal{ID}$ and a password. Next, the system defines his/her interest topics. Let us suppose the user belong to a group which works in **Science of Nutriment**, because of this it has the UNESCO code **3206**; remember the group could have more UNESCO codes. Then, to define the vector of interest topics the system assigns the user this code (**3206**) with degree **Total** ($b_4 \in S_1$). With this information the user profile is represented by a vector of interest topics with the following values:

$$VU_{\mathcal{ID}}[x] = b_4, \quad \text{if } x = 100$$
$$VU_{\mathcal{ID}}[x] = b_0, \quad \text{otherwise.}$$

**Remark**. The UNESCO code 3206 is in the position 100 of the list so it is stored in $VU_{\mathcal{ID}}[100]$.

### Resources Insertion Process

This sub-process is carried out by the library staff that receive or find information about a resource and they want to spread this information. The experts introduce the interesting resources into the system and it automatically sends the information to the suitable users along with a relevance degree and collaborations possibilities.

As we said in the previous section, the system stores the general information about the resource and its scope. The scope is represented by a vector of UNESCO codes whereby to insert the resource the experts decide the UNESCO codes to assign it. Moreover, to manage the linguistic information, the experts also decide a linguistic label in $S_1$ to weight the importance degree of each UNESCO code of level 2 with regard to the resource.

Hence, when the library staff are going to insert a new resource, they access to the system, insert all the information about it, i.e. title, author(s), abstract, date, source, book name, journal name, daily name, link, text, kind of resource, target and finally they assess the importance degree of each UNESCO code of level 2 with regard to the resource. To do this, the system shows a list of UNESCO codes of level 2 and the library staff decide the codes to assign to the resource scope, selecting a code of the list and assign it a linguistic label to assess its importance degree. Then they accept and can either add another UNESCO code or finally the resource insertion.

*Example 2.* Now let us suppose the digital library receives a paper $i$ about an Science of Nutriment Conference. Then, he/she inserts the paper into the system, introducing all the available information and selecting from a list the UNESCO codes which match with the resource scope. In this example, the library staff could select the codes **3206 - Science of Nutriment** with importance degree **Total** ($b_4 \in S_1$) and **3309 - Food Technology** with degree **Very High** ($b_3 \in S_1$). Once the expert inserts this information, we have a vector $VR_i$ defining the resource $i$ with the following values:

$$VR_i[j] = b_4, \quad \text{if } j = 100$$
$$VR_i[j] = b_3, \quad \text{if } j = 118$$
$$VR_i[j] = b_0, \quad \text{otherwise.}$$

**Remark**. The UNESCO codes 3206 and 3309 are in the positions 100 and 118 of the list so they are stored respectively in $VR_i[100]$ and $VR_i[118]$.

### Filtering Process

As we have said, we are going to use the vector model [26] to represent the resources scope and the user interest topics. This vector model uses sophisticated similarity calculations to do the matching process, such as Euclidean Distance or Cosine Measure. Exactly we are going to use the Cosine Measure we described next.

The **Cosine Measure** is a similarity measure that is developed from the cosine of the angle between the vectors representing the scope resource ($VR$) and the user interest topics ($VU$), or between the vectors representing two users interest topics or between the vectors representing two scope resources. Its definition is [26]:

$$\sigma(VR, VU) = \frac{\sum_{k=1}^{n}(r_k \times u_k)}{\sqrt{\sum_{k=1}^{n}(r_k)^2} \times \sqrt{\sum_{k=1}^{n}(u_k)^2}}$$

where $n$ is the number of terms used to define the vectors (i.e. the number of UNESCO codes of level 2), $r_k$ is the value of term $k$ in the resource vector and $u_k$ is its value in the user vector. In mathematical terms this is the inner product of the resources and users vectors, normalized by their lengths. Using this cosine transforms the angular measure into a measure ranging from 1 for the highest similarity to 0 for the lowest. In the case of two users or two resources, this cosine measure is applied of the same way.

Angular measures representing a view of the resources and users items space from a fixed point, the origin. In addition, an angular measure does not consider the distance of each item from the origin, but only the direction. Hence two items that lie along the same vector from the origin will be judged identically, despite the fact that they may be far apart in the document space. This means that a one-paragraph announcement and an extensive, detailed paper about a topic might be judged to be equally relevant to a query. For example, suppose there are three notices, each described by the same two terms, with resource vectors:

$$VR_1 =< 1, 3 >,$$
$$VR_2 =< 100, 300 >, \ and$$
$$VR_3 =< 3, 1 > .$$

By the cosine measure, $\sigma(VR_1, VR_2) = 1.0$ and $\sigma(VR_1, VR_3) = 0.6$. The cosine measure views $R_2$ as more similar to $R_1$ than is $R_3$. It can be argued that in $R_1$ and $R_2$ the two terms have the same relative importance; that is, that the ratio of their values is the same.

Following this approach when a new resource has been inserted into the system, we compute the cosine measure $\sigma(VR_i, VU_j)$ between the new scope resource vector ($VR_i$) against all the user vectors ($VU_j$, $j = 1..m$ where $m$ is the number of users of the system) to find the fit users to deliver this information. If

$\sigma(VR_i, VU_j) \geq \alpha$, the system select the user $j$. Previously we have defined a threshold value ($\alpha$) to filter out the information. In this iteration, the system takes into account too the user preferences (kind of resource) to consider or not the user. The collaboration preferences are used to classify the selected users in two sets, the selected users that don't want to collaborate $\mathcal{U}_S$ and the selected users arranged to collaborate $\mathcal{U}_C$.

After this, the system has two sets of selected users $\mathcal{U}_S$ and $\mathcal{U}_C$ and for each user it has a value $\sigma(VR_i, VU_j) \geq \alpha$. The system apply to each $\sigma(VR_i, VU_j)$ the transformation function defined in definition 1 to obtain the relevance degree of the resource $i$ for the user $j$, expressed in the set $S_2$. Then, the system sends to the users of $\mathcal{U}_S$ the resource information and its calculated relevance degree by a linguistic label more effective than a number.

For the users in $\mathcal{U}_C$ the system performs other step; it calculates the collaboration possibilities between the selected users. To do it, between each two users $x, y \in \mathcal{U}_C$:

- to analyze if the users are researchers or students and take into account the users preferences about it. For example a researcher could want to collaborate only with others researches of different research group.
- to calculate the cosine measure between the users, $\sigma(VU_x, VU_y)$,
- to obtain the compatibility degree between $x$ and $y$, expressing $\sigma(VU_x, VU_y)$ as a linguistic label in $S_3$ (using the transformation function defined in definition 1) to send it to the user.

Finally the system sends to the users of $\mathcal{U}_C$ the resource information, its calculated relevance degree and the collaboration possibilities along with a compatibility degree. All the process is shown in the Fig. 2.

### Feedback Phase

This phase is related to the activity developed by the **filtering system** once users have taken some of the resources delivered by the system. As we said, user profiles represents the users' long-term information needs or interests and a desire property for user profiles is that they should be adaptable since users' needs could change continuously. Because of this, the system allows the users update their profiles to improve the filtering process with the needs of each one. In our system this feedback process is developed in the following steps:

- The users access the system entering their $\mathcal{ID}$ and password.
- The users can do the following operations:
  - to edit their collaboration preferences,
  - to edit their preferences about kind of desired resources,
  - to edit their interest topics:
    - to add new UNESCO codes with its importance degrees, i.e. linguistic labels $b_i \in S_1$.

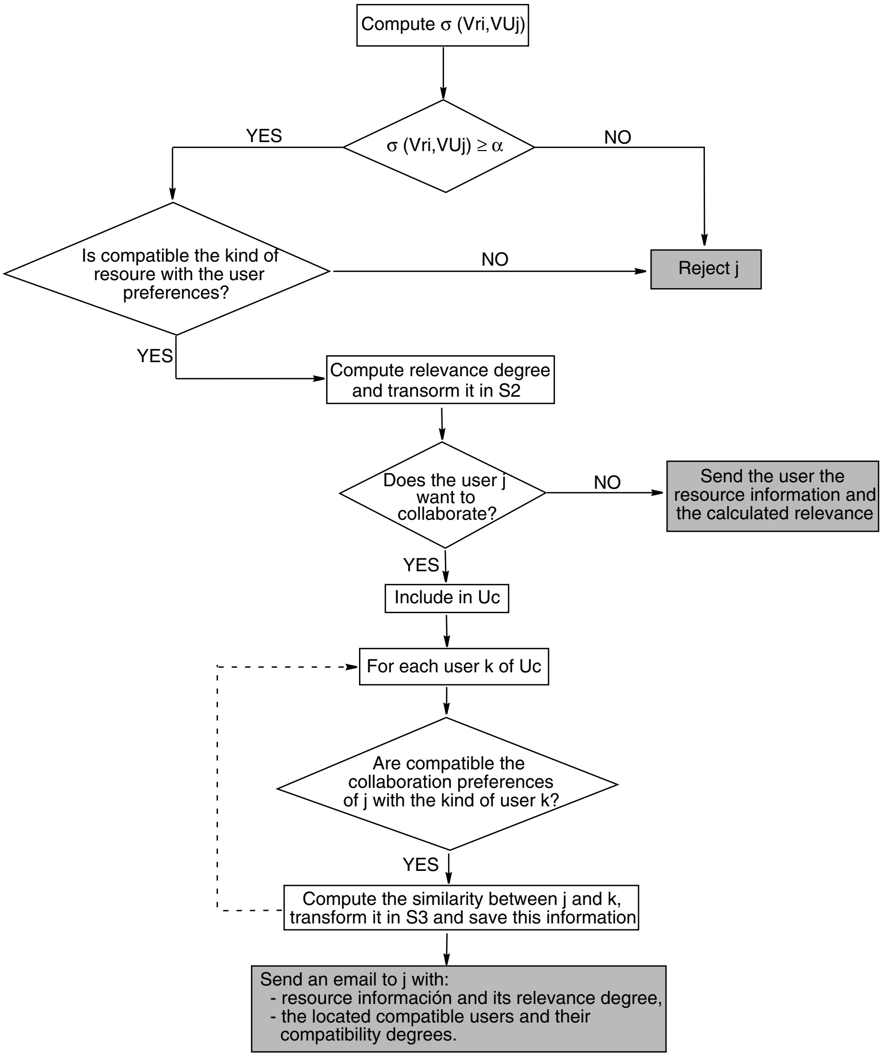



**Fig. 2** Matching process for an user $j$

- to delete an existing UNESCO code.
- to modify the importance degree (linguistic label $b_i \in S_1$) assigned to an existing UNESCO code.

*Example 3.* Assuming information given in 1, let us suppose the user $\mathcal{ID}$ wants to update his/her profile because $\mathcal{ID}$ thinks he/she should belong to the category **3309 - Food Technology**. In this case the user wants to add a new UNESCO code and assigns it an importance degree of **High** ($b_3 \in S_1$).

After this, the user $\mathcal{ID}$ has a new profile represented by a new vector with the following values:

$$VU_{\mathcal{ID}}[y] = b_4, \quad \text{if } y = 100$$
$$VU_{\mathcal{ID}}[y] = b_3, \quad \text{if } y = 118$$
$$VU_{\mathcal{ID}}[y] = b_0, \quad \text{otherwise.}$$

# 4 Concluding Remarks

The exponential increase of Web sites and electronic documents is contributing to that Internet users not being able to find the information they seek in a simple and timely manner. The impact of the new digital technologies in others organizations is causing the apparition of problems similar to the Web ones, as for example it happens in UDL. Hence, users of UDL need tools to assist them in their processes of information gathering because of the large amount of information available on these systems. We have presented two techniques that could contribute to solve this problem, the information filtering tools and multi-granular FLM. Then, we have defined a model of fuzzy linguistic recommender system to spread research resources in UDL using both techniques. The proposed system is oriented both researchers and student and advice them research resources that could be interesting for them. In particular, it is a personalized system based on both content-based filtering tools and the multi-granular FLM. The system filters the incoming information stream to spread the information to the fitting users and recommends them about collaboration possibilities. The multi-granular FLM has been applied in order to improve the users-system interaction and the interpretability of the system activities. Moreover, the system brings a extra value, that is, on the one hand it sends the users a linguistic relevance degree to justify the information mailing and on the other hand it recommends the user the collaboration possibilities with other users. However we think the system could improve, incorporating some features, such as incorporate a module to define the resources scope automatically, or apply new techniques that have been used in development of the recommender systems.

# References


1. Arfi B., Fuzzy decision making in politics. A linguistic fuzzy-set approach (LFSA). Political Analysis, 13 (1), 23–56, 2005.
2. Baeza-Yates R. and Ribeiro-Neto B., Modern Information Retrieval. Addison-Wesley, 1999.
3. Basu C., Hirsh H. and Cohen W., Recommendation as classification: Using social and content-based information in recommendation. In Proc. of the Fifteenth National Conference on Artificial Intelligence, 714–720, 1998.
4. Callan J., et. al., Personalisation and Recommender Systems in Digital Libraries. Joint NSF-EU DELOS Working Group Report. May 2003.
5. Chao H., Assessing the quality of academic libraries on the Web: The development and testing of criteria. Library & Information Science Research, 24, 169–194, 2002.
6. Claypool M., Gokhale A. and Miranda T., Combining content-based and collaborative filters in an online newpaper. In Proc. of the ACM SIGIR-99 Workshop on Recommender Systems-Implementation and Evaluation, 1999.

7. Cordón O., Herrera F. and Zwir I., Linguistic modelling by hierarchical systems of linguistic rules. IEEE Transactions on Fuzzy Systems, 10 (1), 2–20, 2001.
8. Degani R. and Bortolan G., The Problem of Linguistic Approximation in Clinical Decision Making. Int. J. of Approximate Reasoning, 2, 143–162, 1988.
9. Frias-Martinez E., Magoulas G., Chen S. and Macredie R., Automated user modeling for personalized digital libraries. International Journal of Information Management 26, 234–248, 2006.
10. Garcia E. and Garcia L.A., La Biblioteca Digital. Arco Libros S.L., 2001.
11. Gonçalves M. A., Fox E. A., Watson L. T. and Kipp N. A., Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. ACM Trans. Inf. Syst. 22, 2, 270–312, Apr. 2004.
12. Good N., Schafer J.B., Konstan J.A., Borchers A., Sarwar B.M., Herlocker J.L. and J. Riedl., Combining collaborative filtering with personal agents for better recommendations. In Proc. of the Sixteenth National Conference on Artificial Intelligence, 439–446, 1999.
13. Hanani U., Shapira B. and Shoval P., Information Filtering: Overview of Issues, Research and Systems. User Modeling and User-Adapted Interaction, 11, 203–259, 2001.
14. Herrera F. and Herrera-Viedma E., Aggregation operators for linguistic weighted information. IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems, 27, 646–656, 1997.
15. Herrera F., Herrera-Viedma E. and Martínez L., A Fusion Approach for Managing Multi-Granularity Linguistic Term Sets in Decision Making. Fuzzy Sets and Systems, 114, 43–58, 2000.
16. Herrera F., Herrera-Viedma E. and Verdegay J.L., Direct approach processes in group decision making using linguistic OWA operators. Fuzzy Sets and Systems, 79, 175–190, 1996.
17. Herrera F. and Martínez L., A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Transactions on Fuzzy Systems, 8 (6), 746–752, 2000.
18. Herrera F. and Martínez L., A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics, 31(2), 227–234, 2001.
19. Herrera F. and Martínez L., The 2-tuple linguistic computational model. Advantages of its linguistic description, accuracy and consistency. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 9, 33–48, 2001.
20. Herrera-Viedma E., Modeling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach. J. of the American Society for Information Science and Technology, 52(6), 460–475, 2001.
21. Herrera-Viedma E., An information retrieval system with ordinal linguistic weighted queries based on two weighting elements. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, 9, 77–88, 2001.
22. Herrera-Viedma E., Cordón O., Luque M., López A.G. and Muñoz A.M., A Model of Fuzzy Linguistic IRS Based on Multi-Granular Linguistic Information. International Journal of Approximate Reasoning, 34 (3), 221–239, 2003.
23. Herrera-Viedma E., Martínez L., Mata F. and Chiclana F., A Consensus Support System Model for Group Decision-making Problems with Multi-granular Linguistic Preference Relations. IEEE Trans. on Fuzzy Systems, 2005. To appear.
24. Herrera-Viedma E. and Peis E., Evaluating the informative quality of documents in SGML-format using fuzzy linguistic techniques based on computing with words. Information Processing & Management, 39(2), 195–213, 2003.
25. Kobayashi M. and Takeda K., Information retrieval on the web. ACM Computing Surveys, 32(2), 148–173, 2000.
26. Korfhage R.R., Information Storage and Retrieval. New York: Wiley Computer Publishing, 1997.
27. Lau Noriega J., Bibliotecas universitarias: Su importancia en el proceso de acreditacin. Gaceta Universitaria, 149, p. 1, 2002.
28. Lawrence S. and Giles C., Searching the web: General and scientific information access. IEEE Comm. Magazine, 37 (1), 116–122, 1998.

29. Popescul A., Ungar L.H., Pennock D.M. and Lawrence S., Probabilistic models for unified collaborative and content-based recommendation in sparce-data environments. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI), San Francisco, 437–444, 2001.
30. Quiroga L.M. and Mostafa J., An experiment in building profiles in information filtering: the role of context of user relevance feedback. Information Processing and Management, 38, 671–694, 2002.
31. Renda M.E. and Straccia U., A personalized collaborative Digital Library environment: a model and an application. Information Processing and Management, 41, 5-21, 2005.
32. Reisnick P. and Varian H.R., Recommender Systems. Special issue of Comm. of the ACM, 40 (3), 56–59, 1997.
33. Clasificacin UNESCO. Ministerio de Educación y Ciencia. http://www.mec.es/ciencia/jsp/plantilla.jsp?area=proyectos/invest&id=53
34. Zadeh L.A., The concept of a linguistic variable and its applications to approximate reasoning. Part I. Information Sciences, 8, 199-249, 1975. Part II, Information Sciences, 8, 301-357, 1975. Part III, Information Sciences, 9, 43–80, 1975.

# Fuzzy Measures in Image Processing

**Tamalika Chaira**

**Abstract** This chapter provides an overview of fuzzy measures and fuzzy integrals, measures of fuzziness, and their application in image processing in the areas of region based segmentation, thresholding, and color retreieval. This chapter also introduces a fuzzy color image retrieval method using a new type of membership function called beta membership function with Generalized Tversky's index as a measure of fuzziness. Lastly, the proposed method has been compared with some existing fuzzy and non-fuzzy methods.

## 1 Introduction

The concept of a fuzzy measure is one of the most important area in mathematics and so is the integral with respect to the fuzzy measure. The classical measure and the integral theory is based on the additivity of the set function. Additive property is sometimes important in some application but sometimes becomes ineffective in many reasonings in real world environment e.g., fuzzy logic, decision making, artificial intelligence etc. But with the introduction of fuzzy set theory by Prof. Lotfi A. Zadeh in 1965, which handles real life problems i.e. vagueness and ambiguity, the additive property of classical measures becomes a subject of controversy. During the seventies, M. Sugeno studied a common type of non- additive and monotonic set functions, called fuzzy measures. According to Sugeno, fuzzy measures are obtained by replacing the additivity requirement of classical measures with weaker requirements of monotonicity and continuity. Fuzzy measure permits to represent some background knowledge about the information sources that are being aggregated. Fuzzy integral is a general term for integral with respect to fuzzy measure. There are many kinds of fuzzy integrals developed by many researchers in defferent streams. Some of these are - the Choquet integral, Sipos integral, Sugeno integral, t-conorm integral etc.

The question as how fuzzy is a fuzzy set has been one of the issues associated with the development of fuzzy set theory. If we understand an image (or its segments) as fuzzy sets, then we have to answer the question how fuzzy the image is. The measure of uncertainty is the measure of fuzziness or fuzzy entropy. Kaufmann, De Luca Termini, Yager suggested some measures of fuzziness.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

This chapter reviews the theory of fuzzy measures and fuzzy integrals, measures of fuzziness, and fuzzy distance measures. Their application in image processing e.g. in image segmentation, image retrieval has also been studied extensively.

This chapter also introduces a method for color image retrieval, where a new type of membership function has been proposed called beta membership function to find the membership degree of the gray levels of an image histogram and a Generalized Tversky's index has been used as a fuzzy similarity measure. The method has been compared with the earlier existing fuzzy and non-fuzzy methods.

## 2 Fuzzy Measures and Fuzzy Integrals

Sugeno [1] in 1974 introduced the theory of fuzzy measures and fuzzy integrals. Fuzzy measure provides a structure for modeling the knowledge about the variables that are uncertain and unknown. It deals with the imprecise information where imprecision is defined as the uncertainty that an element $x$ belong to a set $X = \{x_1, x_2, \cdots, x_n\}$, but we do not know the degree of belongingness, $\mu$, to the set $X$. Therefore $'x'$ is assigned a membership value to represent the degree of belongingness. One useful feature of fuzzy measure is its ability to interact with fuzzy integrals to provide a tool in dealing with uncertainty.

Let $X = \{x_1, x_2, \cdots, x_n\}$ be a set. A fuzzy measure $g$ on a finite space $X$ is a mapping from subsets of $X$ into a unit interval such that $g\colon 2^X \to [0, 1]$, if it satisfies the following properties [2]:

1. Boundary conditions: $g(\phi) = 0$, $g(X) = 1$
2. Monotonicity: If $A \subset B, \forall A, B \in X$, then $g(A) \leq g(B)$
3. If $A_i$ is an increasing subsequence of subsets of $X$ i.e. $A_1 \subset A_2 \subset A_3 \subset \cdots, A_n$, then

$$\lim_{i \to \infty} g(A_i) = g(\bigcap_{i=1}^{\infty} A_i)$$

   A fuzzy measure, $g_\lambda$, is a Sugeno measure or $\lambda-$ fuzzy measure if it satisfies the following condition for some $\lambda > -1$.

4. $\forall A, B \subset X$ with $A \cap B = \phi$

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$$

For a finite set $A$, the fuzzy measure of set $A$ may be calculated as:

$$g(A) = \frac{1}{\lambda}[\prod_{x_i \in A} (1 + \lambda g_i) - 1], \quad \lambda \neq 0 \tag{1}$$

where $g_i = g(\{x_i\})$ is called fuzzy density.

Zero is always a solution and so the polynomial reduces from degree $n$ to $n$–1.

With $g(X) = 1$, the value of $\lambda$ may be calculated as:

$$\lambda + 1 = \prod_{i=1}^{n}(1 + \lambda g_i)$$

If $A_i = \{x_i, \cdots, x_n\}$, then values of $g(\{x_i, \cdots, x_n\})$ for $1 \le i \le n$ may be computed recursively as:

$g(\{x_i\}) = g_i$

$g(\{x_i, \cdots, x_n\}) = g(\{x_n\}) + g(\{x_i, \cdots, x_{n-1}\}) + \lambda g(\{x_n\})g(\{x_i, \cdots, x_{n-1}\})$

Some properties of fuzzy measure are:

If $f, g \in [0, 1]^X$ then

$$f \wedge g(x) = \min\{f(x), g(x)\}$$
$$f \vee g(x) = \max\{f(x), g(x)\}$$

Fuzzy integral is an integral of a real function with respect to a fuzzy measure. The performance of fuzzy integral is totally dependant on fuzzy measure. There are many types of fuzzy integrals but the most common fuzzy integrals are Sugeno integral [1] and Choquet integral [3].

A fuzzy integral of a function $f$ with respect to fuzzy measure $g$ may be defined as:

$$\int_A f(x) \circ g(\cdot) = \sup_{\alpha \in [0,1]} [\min(\alpha, g(f_\alpha))]$$

where $f_\alpha$ is a $\alpha$ level set of $f$.

Fuzzy integral is a Sugeno integral which is the generalization as a weighted minimum or median and thus presents a combination of fuzzy connectives - minimum and maximum. It is an alternative form of aggregation operator for fuzzy measures based on 'max' instead of sum. It combines data in numerical or ordinal scales with respect to fuzzy measure.

If $X = \{x_1, x_2, \cdots, x_n\}$ is a finite set, then Sugeno integral may be represented as:

$$(S)\int_A f(x) \circ g(\cdot) = \max_{i=1}^{n}[\min(f(x_i), g(A_i)] \tag{2}$$

where $f(x_1) \ge f(x_2) \ge \cdots \ge f(x_n)$ and $A_i = \{x_i, x_{i+1} \cdots x_n\}$

Choquet integral uses a combination of algebraic product and then addition becomes a generalization of operators such as arithmetic mean.

$$(C)\int_A f(x)\circ g(\cdot) = \sum_{i=1}^{n} g(A_i)[f(x_i) - f(x_{i-1})] \tag{3}$$

in which $f(x_1) \leq f(x_2)\cdots f(x_n)$ and $A_i = \{x_i, x_{i+1}\cdots x_n\}$ , $f(x_0) = 0$.

The enclosed subindices are the result of sort operation, e.g. if $f_1 \geq f_3 \geq f_2$, then

$$f(x_1) = f_1,\ f(x_2) = f_3,\ f(x_3) = f_2$$

The coefficients of fuzzy measure take the values as :

$$g(A_1) = g(\{x_1\}),\quad g(A_2) = g(\{x_1, x_3\}),\quad g(A_3) = g(\{x_1, x_2, x_3\}).$$

Based on the properties of fuzzy measure, fuzzy integral is an aggregation operator on multi-attribute fuzzy information. Examples for calculating fuzzy integrals are given below.

**Example**. Calculation of Sugeno measure:

Consider a set

$$X = \{a, b, c\}$$

Let the fuzzy density values are given as:

$$g^i = \begin{cases} 0.2 & \text{if } x_i\text{=a} \\ 0.3 & \text{if } x_i\text{=b} \\ 0.1 & \text{if } x_i\text{=c} \end{cases}$$

The value of λ may be calculated from the following equation

$$1 + \lambda = (1 + 0.2\lambda)(1 + 0.3\lambda)(1 + 0.1\lambda)$$

The solutions for λ are

$$\lambda = \{-21.44, 3.1\}.$$

Then the Sugeno measure may be written as follows:

| $E$ | *Sugeno measure* |
|---|---|
| $\{a\}$ | $g(\{a\}) = 0.2$ |
| $\{b\}$ | $g(\{b\}) = 0.3$ |
| $\{c\}$ | $g(\{c\}) = 0.1$ |

From the condition 4 as mentioned above

$$
\begin{array}{c|l|}
\{a, b\} & g(\{a, b\}) = g(\{a\}) + g(\{b\}) + \lambda g(\{a\})g(\{b\}) \\
 & \qquad\qquad = 0.686 \\
\{b, c\} & g(\{b, c\}) = g(\{b\}) + g(\{c\}) + \lambda g(\{b\})g(\{c\}) \\
 & \qquad\qquad = 0.4930 \\
\{a, c\} & g(\{a, c\}) = g(\{a\}) + g(\{c\}) + \lambda g(\{a\})g(\{c\}) \\
 & \qquad\qquad = 0.3620 \\
\{a, b, c\} & g(\{a, b, c\}) = 1
\end{array}
$$

**Example**. Calculation of Sugeno integral.

Let the choice of $f$ be given as follows:

$$
f(x_i) = \begin{cases} 0.7 & \text{if } x_i\text{=a} \\ 0.5 & \text{if } x_i\text{=b} \\ 0.1 & \text{if } x_i\text{=c} \end{cases}
$$

$$
\begin{aligned}
\int_A f(x) \circ g(\cdot) &= \max_{i=1}^{n}[\min(f(a), g_\lambda(a, b, c)), \min(f(b), g_\lambda(b, c)), \min(f(c), g_\lambda(c))] \\
&= \max[\min(0.7, 1), \min(0.50.4930), \min(0.1, 0.1)] \\
&= \max(0.7, 0.493, 0.1) \\
&= 0.7
\end{aligned}
$$

**Example**. Calculation of Choquet integral

The choice of $f$ in ascending order

$$
f(x_i) = \begin{cases} 0.2 & \text{if } x_i\text{=a} \\ 0.4 & \text{if } x_i\text{=b} \\ 0.7 & \text{if } x_i\text{=c} \end{cases}
$$

$$
\begin{aligned}
\int_A f(x) \circ g(\cdot) &= \sum_{i=1}^{n} g(A_i)[f(x_i) - f(x_{i-1}]] \\
&= g(\{a, b, c\})f(a) + g(\{b, c\})[f(b) - f(a)] + g(\{c\})[f(c) - f(a)] \\
&= 10.2 + 0.4930[0.4 - 0.2] + 0.1[0.7 - 0.4] \\
&= 0.2 + 0.098 + 0.03 \\
&= 0.328
\end{aligned}
$$

Some application of fuzzy measures and fuzzy integrals in image processing are as follows:

In **color image retrieval**, given a query image and an image database, we have to find out the image that is most similar to the query image with respect to color, texture or shape features. If we use color as a feature, then we have to find the similarity between the color features of the query image and the image in the database. The similarity measure may be Euclidean distance, Hamming distance or any other distances.

Frigui [2] proposed a dissimilarity measure using Choquet integral for retrieving color images. Let $X = [f_1, f_2, \cdots, f_n]$ be a set of $'n'$ features where $f_i$ represent color, texture or shape feature and $x^k, x^l$ are two n- dimensional feature vectors that represent the two images to be compared. Then Choquet integral, as a dissimilarity measure between two images, may be written as:

$$C_g(x^k, x^l) = \sum_{i=1}^{n} [(|x_i^k - x_i^l| - |x_{i-1}^k - x_{i-1}^l|)g(\{f_i, \cdots, f_n\})]$$

where $_{(i)}$ denotes the feature indices and

$$|x_1^k - x_1^l| \leq |x_2^k - x_2| \leq^| x_n^k - x_n^l|.$$

They represented $g^i = g(\{f_i\})$ as the relevence of feature $i$ and it depends on user's perception of similarity. All the features are initialized with the same values using

$$g^i = \frac{1}{n}, i = 1, 2, \cdots, n.$$

$\lambda$ is calculated from (1).

Then fuzzy measures are computed recursively as:

$$g\{f_i\} = \frac{1}{n};$$

$$g\{f_i, \cdots, f_n\} = g\{f_n\} + g(\{f_i, \cdots, f_{n-1}\}) + \lambda g\{f_n\}.g\{f_i, \cdots, f_{n-1}\}$$

With these fuzzy measures, they calculated the Choquet integral and based on the dissimilarity, the system retrieves the $K$ most similar images. The user then classifies the retrieved images as relevant or irrelevant images. If the user is not satisfied with the relevant images, the feature relevance weights are again updated [2].

For each image, $C_g$ is computed to retrieve the most similar image.

Li et al. [4] argued that using Choquet integral [2, 5], a complex similarity adjustment is reflected. The fuzzy measure is defined over a power set of a given set of features and Choquet integral is used to aggregate the similarity of these features. So there is a complex interaction between the features. Thus they suggested the efficacy of Sugeno's integral based similarity measure for content based image retrieval by formulating a subjective feedback information. Sugeno Integral is having a non-linear property i.e. the max - min relationship and so this complex interation does not arise. Using a set of features, they calculated the dissimilarity between the images using Sugeno integral.

**In region based image segmentation**, regions are extracted on the basis of color, texture or shape. There are many clustering algorithms in literature. Clustering could be the process of organizing objects into groups whose members are similar in some way. It involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. Clustering may also be thought of as a form of data compression, where a large number of samples are converted into a small number of representative clusters. Depending on the data and the application, different types of similarity measures may be used to identify classes, where the similarity measure controls how the clusters are formed. Similarity measures may be Euclidean, Minkowiski or any other distances. The most widely used clustering algorithm is fuzzy clustering algorithms that was originally introduced by Professor Jim Bezdek in 1981 [6].

Pham and Yan [7] segmented color image data using Choquet fuzzy integral as the distance measure. As other fuzzy clustering methods, they suggested that their segmentation technique does not require an initial estimate of cluster center. They found the cluster centers using mountain clustering technique. To express the similarity between image data with respect to $RGB$ components, they used $\pi$ membership function and used Choquet integral for calculating the degree of color similarity between each pixel $p_i$ with respect to all cluster centres $c_k$:

$$\sigma = \sum_{j=1}^{3} g(A)[f(x_j) - f(x_{j-1}]$$

$\sigma$ is the degree of similarity. The greater the value, the more is the similarity.

$$x_1 = R_{pi}, x_2 = G_{pi} x_3 = B_{pi}$$

and

$$A = [x_1 = R_{ck}, x_2 = G_{ck}, x_3 = B_{ck}]$$

Then segmentation is carried by assigning each pixel to the cluster centre having a maximum degree of similarity.

## 3 Measures of Fuzziness

The measure of uncertainty in a fuzzy set is the measure of fuzziness or fuzzy entropy. It tells how fuzzy the fuzzy set is. Zadeh [8] first introduced the idea of fuzzy entropy in 1969. De Luca and Termini [9] introduced axiom contruction of entropy of a fuzzy set. The distance measure is also used to define entropy. It is a measure that finds the difference between two fuzzy sets. The use of distance measure to define fuzzy entropy is due to Kaufmann [10] whereas Yager [11] defined entropy as the distance from a fuzzy set and its complement. Entropy measures are used to know the uncertainty in the process. Some entropy measures have influence from Shannon probabilistic entropy which is commonly used as a measure of information.

Let $X$ be a universal set and $F(X)$ is the class of all fuzzy sets of $X$. The properties of measures of fuzziness or entropy between two fuzzy sets $A$ and $B$ such that $A, B \in X$ with membership functions $\mu_A(x_i)$ and $\mu_B(x_i)$ are given as follows:

1. $I(A) = 0$ or minimum iff $\mu_A(x_i) = 0$ or $1, \forall x \in A$
2. $I(A) = 0$ or maximum iff $\mu_A(x_i) = 0.5, \forall x \in A$
3. $I(A) = I(\bar{A})$
4. $I(A) \geq I(A^\star)$ where $A^\star$ is a sharpened version of $A$
   such that $\mu^\star_A(x_i) \leq \mu_A(x_i)$ if $\mu_A(x_i) \leq 0.5$
   $\mu^\star_A(x_i) \geq \mu_A(x_i)$ if $\mu_A(x_i) \geq 0.5$

Many authors suggested many definitions for fuzzy entropy.

**a) Entropy**
as defined by De Luca and Termini [9] is the Logarithmic entropy which is given by:

$$H(A) = \frac{1}{n \log 2} \sum_{i=1}^{n} S(\mu_A(x_i)) \tag{4}$$

with Shannon's function

$$S(\mu_A(x_i)) = -\mu_A(x_i) ln(\mu_A(x_i)) - (1 - \mu_A(x_i)) ln(1 - \mu_A(x_i))$$

This entropy is quiet different from the classical entropy because no probabilistic concept has been used.

**b) Index of fuzziness**
Kaufmann [10] defined the entropy or measure of fuzziness as the degree of ambiguity or fuzziness present in a set by measuring the distance between the membership values of the set $A$ and its nearest ordinary set. The index of fuzziness is defined as:

$$I(A) = \frac{2}{n^k} d(A, \widetilde{A})$$

where $d(A, \widetilde{A})$ denotes the distance between set $A$ and its nearest ordinary set $\widetilde{A}$. An ordinary set $\widetilde{A}$ nearest to the fuzzy set $A$ is defined as

$$\mu_{\widetilde{A}}(x_i) = \begin{cases} 0 \text{ if } \mu_A(x_i) \leq 0.5 \\ 1 \text{ if } \mu_A(x_i) > 0.5 \end{cases}$$

There are two types of indices of fuzziness: - linear index of fuzziness and quadratic index of fuzziness. These are defined as follows:

(i) If $k = 1, d$ becomes Hamming distance and the linear index of fuzziness becomes:

$$L.I = \frac{2}{k} d(A, \widetilde{A})$$

and

$$d(A, \widetilde{A}) = \sum_{i=1}^{n} (\mu_A(x_i) - \mu_{\widetilde{A}}(x_i))$$
$$= \sum_{i=1}^{n} \mu_{A \cap \widetilde{A}}$$

$\mu_{A \cap \widetilde{A}}$ denotes the intersection (common) of the membership value of the element $a_i$ at $i^{th}$ point in the fuzzy set $A$ and its complement. So the linear index of fuzziness (L.I) may be written as:

$$L.I. = \frac{2}{n} \sum_{i=0}^{n-1} \min(\mu_A(a_i), 1 - \mu_{\bar{A}}(a_i)) \tag{5}$$

ii) If $k = 0.5, d$ becomes Euclidean distance and the quadratic index of fuzziness (Q.I.) is defined as:

$$Q.I. = \frac{2}{\sqrt{n}} [\sum_{i=0}^{n-1} \{\mu_A(a_i) - \mu_{\widetilde{A}}(a_i)\}^2]^{\frac{1}{2}}$$
$$Q.I. = \frac{2}{\sqrt{n}} [\sum_{i=0}^{n-1} \min(\mu_A(a_i), 1 - \mu_{\bar{A}}(a_i))^2]^{\frac{1}{2}} \tag{6}$$

**c) Yager's measure**

Yager [11] defined fuzzy entropy or measure of fuzziness as the relationship between fuzzy set and its complement. The distance between the fuzzy set $A$ and its complement $\bar{A}$ is denoted as:

$$D(X, \bar{X}) = \sum_{i=1}^{n} [\mu_A(x_i) - \mu_{\bar{A}}(x_i)]^{\frac{1}{p}}$$
$$\mu_{\bar{A}}(x_i) = 1 - \mu_A(x_i), \quad i = 1, 2, \cdots, n$$

The measure of fuzziness is defined as:

$$S(A, B) = 1 - \frac{D(X, \bar{X})}{X^{\frac{1}{p}}} \tag{7}$$

Another entropy as defined by Pal and Pal [12] is the exponential entropy and is defined as:

$$H(A) = \frac{1}{n\sqrt{(e)} - 1} \sum_{i=1}^{n} (S(\mu_A(x_i)) - 1) \tag{8}$$

with Shannon's function

$$S(\mu_A(x_i)) = \mu_A(x_i) e^{(1-\mu_A(x_i))} + (1 - \mu_A(x_i)) e^{\mu_A(x_i)}$$

**d) Index of non-fuzziness**
The index of non-fuzziness implies the amount of non-fuzziness or crispness in $\mu_A$ by computing the distance from its complement version.

$$D(X, \bar{X}) = \sum_i |\mu_A(a_i) - \mu_{\bar{A}}(a_i)|, \quad i = 1, 2, \cdots, n \tag{9}$$

**e) Fuzzy geometrical measure**
In pattern recognition we often want to measure the geometric properties of regions in an image, but these regions are not crisply defined rather, it is sometimes regarded as fuzzy subsets of an image. Fuzzy geometrical measures such as fuzzy compactness may be used to measure the geometrical fuzziness of different regions of an image. Many of these geometric properties include area, perimeter, compactness etc. The standard approach to image analysis and recognition begins with the image segmentation where the image is segmented into various regions. The optimization of these measures (e.g. minimization of fuzzy compactness regarding to the cross-over point of membership function) may be applied to make fuzzy and/or crisp pixel classifications.

Rosenfeld [13] extended the concept of digital picture geometry to the fuzzy subsets and generalized some of the standard geometric properties among the regions of fuzzy subsets. Here $\mu(a_{ij})$ is replaced by $\mu$ for simplicity.

The area of $\mu$ is defined as:

$$a(\mu) = \int \mu$$

where the integral is taken over the region in which $\mu > 0$. If $\mu$ is a piecewise constant as in a digital image, then $a(\mu)$ is the weighted sum of the areas of the regions on which $\mu$ has constant values, weighted by these values. For a piecewise constant, perimeter of $\mu$ is defined as:

$$P(\mu) = \sum_{m,n} \sum_{k} |\mu_m - \mu_n||A_{mnk}|$$
$$m, n = 1, 2, \cdots, r, m < n, k = 1, 2, \cdots, r_{mn}$$

This is the weighted sum of the arcs $A_{mnk}$ along which $m^{th}$ and $n^{th}$ regions having constant $\mu$ values, $\mu_m$ and $\mu_n$ respectively, meet weighted by the absolute difference of these values. The compactness of $\mu$ is defined as:

$$\text{Comp}(\mu) = \frac{a(\mu)}{P^2(\mu)} \tag{10}$$

## 4 Fuzzy Similarity Measure

In day to day life, we often encounter situation where we have to distinguish between similar groups or elements or find the similarity between two images or patterns. In that case, similarity measure is the basic concept in a fuzzy set that is used to find the similarity. Suppose there are $n$ elements, $\{x_1, x_2, \cdots, x_n\}$ and $N$ groups, $\{G_1, G_2, \cdots, G_N\}$. We may want to know the similarity or the relation between the two groups or between the two elements. So two types of questions may arise:

1. At what degree the groups are similar.
2. At what degree the elements are in the same group.

For answer to these questions, we have to use similarity measure between fuzzy sets and fuzzy elements.

Let $X$ be a universal set and $F(X)$ is the class of all fuzzy sets of $X$. The properties of similarity measure and distance measure between two fuzzy sets $A$ and $B$ such that $A, B \in F(X)$ with membership functions $\mu_A(x_i)$ and $\mu_B(x_i)$ are given as follows:

1. $S(A, B) = S(B, A), \forall A, B \in F(X)$.
2. For three fuzzy sets $A, B, C, \forall A, B, C \in F(X)$ if $A \subset B \subset C$ then $S(A, B) \geq S(A, C)$ and $S(B, C) \geq S(A, C)$.
3. Similarity degree is bounded i.e. $0 \leq S(A, B) \leq 1$ , if $A, B$ are normalized.

### Fuzzy Measure based on Tversky's model

Tversky [14] in 1977 proposed a feature contrast model. Instead of considering stimuli as points in a metric space, Tversky characterized them as set of binary points features. In other words stimulus $'a'$ is characterized by $'A'$ set of features that the stimulus possesses. Equivalently, a feature set is a set of logic predicates, which are true for the stimulus in question. Let $a, b$ be two stimuli and $A$ and $B$ be the sets of features respectively, and $s(a, b)$ be a similarity between $a$ and $b$. Then Tversky's theory is based on the following assumptions:

1. **Matching**: $s(a, b) = F(A \cap B, A - B, B - A)$
2. **Monotonicity**: $s(a, b) > s(a, c)$
   whenever $A \cap C \subseteq A \cap B, A - B \subseteq A - C, B - A \subseteq C - A$

A function that satisfies matching and monotonicity is called matching function.

Tolias [15] proposed a similarity measure called Generalized Tversky's index (GTI). In the work [16], the generalized Tversky's index has been extended to find the similarity in gray/color images. It has been defined as:

$$GTI(A, B; \alpha, \beta) = \frac{f(A_n \cap B_n)}{f(A_n \cap B_n) + \alpha.f(A_n - B_n) + \beta.f(B_n - A_n)}$$

$f$ is non-negative and increasing function. Here $A_n$, $B_n$ are two sets of predicates (which are the histograms of two images $A$ and $B$) on the measurements $\alpha, \beta \geq 0$. GTI compares the saliency of the common features to the saliency of distinctive features. $A_n \cap B_n$ denotes the common features i.e. features that are common to both images $A$ and $B$. $A_n - B_n$ is the distinctive features that are present in image $A$ and not in image $B$. The above equation has been written as:

$$GTI(A, B; \alpha, \beta = \frac{\sum_{i=0}^{L} \min(\mu_A(n), \mu_B(n))}{\sum_{i=0}^{L} \min(\mu_A(n), \mu_B(n)) + \alpha. \min(\mu_A(n), 1 - \mu_B(n)) + \beta. \min(1 - \mu_A(n), \mu_B(n))}$$

When the membership values of the gray levels of the histograms are similar, it may be inferred that the common features of the two images are more and the distinctive feature is less. This in turn implies that the two images are similar. The values of $\alpha, \beta$ determine the relative importance of the distinctive features in the similarity assessment. When $\alpha \neq \beta$, a directional similarity measure is obtained that focuses on the distinctive features of image $A$, when $\alpha > \beta$. Likewise when $\beta > \alpha$, the focus is on the distinctive features of image $B$. GTI provides a set theoretic index for similarity assessment based on human perception. In the present work, the values of $\alpha$ and $\beta$ have been chosen as 0.5.

Some applications of fuzzy distance measures/measures of fuzziness in image processing are given below.

In image thresholding, given an input image, a binarized output image is obtained for a certain threshold gray level. The output image contains two regions- object and background. Many authors suggested different methods to threshold an image using measures of fuzziness/fuzzy similarity measures. Huang and Wang [17] minimized Shannon entropy and Yager's measure for thresholding an image. They used inverse membership function that evaluates the relationship of the $(i, j)^{th}$ pixel, $a_{ij}$, of image $A$ and has been defined as:

$$\mu_A(a_{ij}) = \begin{cases} \frac{1}{1+c.|a_{ij}-\mu_0|)} & \text{if } a_{ij} \leq t, \text{ for background} \\ \frac{1}{1+c.|a_{ij}-\mu_1|)} & \text{if } a_{ij} > t, \text{ for object} \end{cases}$$

where the average gray level of the background and object region are respectively given by the relation:

$$\mu_0 = \frac{\sum_{f=0}^{t} f\,\text{count}(f)}{\sum_{f=0}^{t} f\,\text{count}(f)}$$

and

$$\mu_1 = \frac{\sum_{f=t+1}^{L-1} f\,\text{count}(f)}{\sum_{f=t+1}^{L-1} f\,\text{count}(f)}$$

where $'\text{count}(f)'$ denotes the number of occurrences of the gray level of the image. The constant ‘c’ is taken as the inverse of the difference of the maximum value and the minimum value of the image, $c = \frac{1}{\text{max}-\text{min}}$, and has been used to make the membership value of all the points lying between 0.5 and 1 i.e. $0.5 \leq \mu_A(a_{ij}) \leq 1$. This method has also been extended to multilevel thresholding.
The threshold is given as:

$$\text{T} = [\frac{1}{MNln2} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} S(\mu_A(a_{ij}))]$$

with Shannon’s function

$$S(\mu_A(a_{ij})) = -\mu_A(a_{ij}) ln(\mu_A(a_{ij})) - (1 - \mu_A(a_{ij})) ln(1 - \mu_A(a_{ij}))$$

$M$, $N$ is the size of an image.

This $(T)$ is calculated for all the gray levels. The optimum threshold is the minimum of the threshold values calculated for all the gray levels.
Chaira and Ray [18, 19, 20] minimized indices of fuzzziness and maximizes the fuzzy compactness for image thresholding using a different type of membership function, Gamma membership function, for finding the membership values of the pixels of an image. They argued that using Gamma membership function the thresholded results are better.

For any threshold $'t'$, the membership functions for the background and object are:

$$\mu_A(a_{ij}) = \begin{cases} \exp(-c|a_{ij} - \mu_0|) & \text{if } a_{ij} \leq t, \text{ for background} \\ \exp(-c|a_{ij} - \mu_1|) & \text{if } a_{ij} > t, \text{ for object} \end{cases}$$

The constant is same as described in Huang’s method and has been used to make the membership value of all the pixels to lie between [0, 1].
For each threshold, linear index of fuzziness is calculated as:

$$L.I. = \frac{2}{n}\sum_{i=0}^{L-1}\min(\mu_A(a_i), \mu_{\tilde{A}}(a_i)), \quad n = \{0, 2, \cdots, L-1\}$$

$L$ is the maximum gray level of the image. The actual threshold value is selected corresponding to the minimum value of the indices of fuzziness calculated for all the gray levels. This method has also been extended to multilevel thresholding.

## 5 The Author's Proposed Method for Color Image Retrieval

Given a query image and a database of images, we have to retrieve the image in the database that is most similar to the query image with color as a feature. For color image retrieval, the similarity between two sets of features, extracted from the database image and the query image is used as a match measure. The match measure, which is a Generalized Tversky's index, has been used to retrieve those images present in the database image, which are similar to the query image. This measure is a human perception model and has been presented to reflect the degree of resemblance between the image in the database and the query image. The features used for the color images are the $RGB$ colors present in the images which are calculated from the histograms of the images. The gray levels of the histograms are fuzzified using a newly proposed membership function, **Beta membership function**.

### Scheme for Finding the Membership Value

As the images are considered to be fuzzy, so the membership values of the pixels of the images are initially calculated. A new type of membership function is used here that has been derived from Beta membership function with some assumptions. It is nothing but another new type of membership function.

### Beta Membership function

Beta distributions have two free parameters, which are labeled according to one of two notational conventions. The usual definition calls these $\alpha$ and $\beta$. The plots shown in Fig. 1 are for various values of $(\alpha, \beta)$ with $\alpha = 1$ and $\beta$ ranging from 0.25 to 3.00.

The domain is [0, 1] and the probability function $P(x)$ is given by

$$\begin{aligned} P(x) &= \frac{(1-x)^{\beta-1}x^{\alpha-1}}{B(\alpha, \beta)} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{\beta-1}x^{\alpha-1} \end{aligned} \quad (11)$$

$x$ should be less than 1.

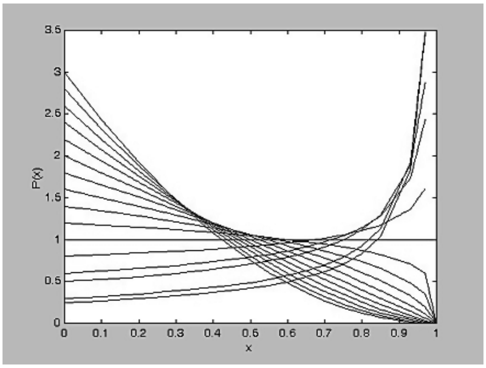



**Fig. 1** Beta distribution

Case:- When $\alpha = 1$, then

$$P(x) = \frac{\Gamma(1+\beta)}{\Gamma(1)\Gamma(\beta)}(1-x)^{\beta-1}$$
$$= c1\frac{\Gamma(1+\beta)}{\Gamma(\beta)}(1-x)^{\beta-1} \quad (12)$$

Constant $c1$ has been chosen as 0.4. Replacing $'x'$ by $|x - \nu|$ in (12), $\nu$ being the referencing parameter, the membership function may be written as:

$$\mu_A(x) = (1 - |x - \nu|)^{\beta-1}$$

For color image retrieval, the location parameter $\nu$ is the histogram of the query image for the channel whose membership function is to be found out. So $\nu$ is replaced by $h_B$ , the histogram of the query image. Image $A$ is one of the image in the database and image $B$ is the query image. The membership function of the database image for each color channel may be written as:

$$\mu_{A(R,G,B)}(n) = (1 - |h_{A(R,G,B)}(n) - h_{B(R,G,B)}(n)|)^{\beta-1} \quad (13)$$

Likewise, the membership values of the query image will be 1 as:

$$\mu_{B(R,G,B)}(n) = (1 - |h_{B(R,G,B)}(n) - h_{B(R,G,B)}(n)|)^{\beta-1} \quad (14)$$

where $n = \{0, 1, 2 \cdots L-1\}$ are the gray levels of the image.

The retrieval of a query image is based on the fuzzy similarity between two images. For computing the similarity between the images, the membership values for the color pixel intensities in both the database image and the query image are initially computed from the (13), (14). Then GTI is calculated for each color channel $R$, $G$, and $B$ between the images in the database and the query image. The total similarity for all the color channels in an image is $\sum_{R,G,B} S$, $S$ is the GTI measure. The image in the database whose GTI value is maximum, is the most similar to the query image.

### 5.1 Experimental Results and Discussion

The aim is to develop an efficient color retrieval. Retrieval is classified as accurate and topmost retrieval if for a given query image, the image which is perceptually most similar to the image in the database, is retrieved. For time complexity 16 bins have been used to represent each of the 1-D color histograms.

The first experiment has been performed on the database of flag images having 70 flags of different countries, second experiment has been performed on a database of VisTex(Vision Texture) color textured images containing 100 images, and third experiment has been performed on a database of color logo images containing 200 logos of different comapnies.

The value of $\beta$ in the experimentation has been selected as $\beta = 2.5$. Also the histograms of the images are normalized.

20 query images from each of the database have been picked and each is presented as a query image. For each of these queries, similar images in the database are manually listed and also the precision and recall rates [17, 21] are calculated for measuring the accuracy of the retrieval scheme. These rates have been calculated at positions 5, 10, 15, and 20. If $'n_1'$ is the number of images retrieved in top 5, 10, 15, and 20 positions that are closed to query and $n_2$ is the number of images in the database that are similar to the query then:

Recall rate = $\frac{n_1}{n_2}$ ,

Precision rate = $\frac{n_1}{n}$ , $n = 5, 10, 15, \text{or } 20$

In the present work, for calculating the precision and recall rates, the average of 20 precision/recall values have been taken. The query image is at the extreme top left of the figures in Figs. 2, 3, and 4. The color flag, color textured, and color logo images are ranked according similarity of colors present in the query image. As the query image is one of the images in the database, so the query image is the exact similar image. Other images are ranked according to the similarity in descending order. Table 1, 2, and 3 show the retrieval efficiency of the proposed method and the

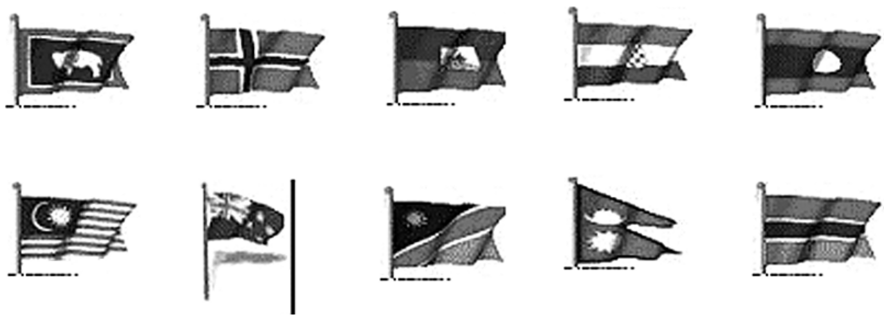


**Fig. 2** Sample result on color flag images

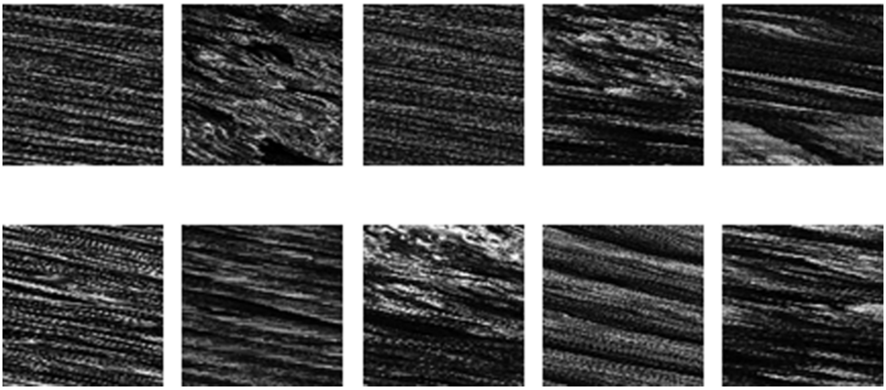


**Fig. 3** Sample result on color textured images

other three methods, i.e., Tolias et al. [14], Vertan and Boujemaa [22], and Swain and Ballard [23] for comparison. Also, Figs. 5, 6 and 7 show the plots of precision and recall of the proposed methods and other three methods.

From Table 1, 2, and 3 it is observed that using the proposed method, the precision rate in the case of color flag images is 0.84 and the recall rate is 0.35 when $n \leq 5$. For color textured images the precision rate is 0.80 and the recall rate is 0.34 when $n \leq 5$. For color logo images the precision rate is 0.75 and the recall rate is 0.29 when $n \leq 5$ . Results of Tolias et al., show the precision rate of 0.51 and the recall rate of 0.22 for color flag images. For color textured images, the precision rate is 0.66 and the recall rate is 0.28 when $n \leq 5$. For color logo images, the precision rate is 0.45 and the recall rate is 0.18 when $n \leq 5$ . In the case Vertan and Boujemaa, precision rate in the case of color flag images is 0.78 and the recall rate is 0.32 when $n \leq 5$. For color textured images the precision rate is 0.77 and the recall rate is 0.33 when $n \leq 5$. For color logo images, the precision rate is 0.70 and the recall rate is 0.27 when $n \leq 5$. With the non-fuzzy/crisp method, Swain and Ballard obtained the precision rate is 0.43 and the recall rate is 0.20 when $n \leq 5$ with color textured images. For color flag images, the precision rate is 0.81 and the recall rate is 0.33


**Fig. 4** Sample result on color logo images

**Table 1** Recall/precision of color textured images

| Method | Precision | | | | Recall | | | |
|---|---|---|---|---|---|---|---|---|
| Beta | 0.80 | 0.63 | 0.55 | 0.47 | 0.34 | 0.47 | 0.56 | 0.62 |
| Tolias et al. | 0.66 | 0.44 | 0.37 | 0.30 | 0.28 | 0.43 | 0.49 | 0.55 |
| Vertan | 0.77 | 0.62 | 0.49 | 0.40 | 0.33 | 0.46 | 0.50 | 0.53 |
| Swain, Ballard | 0.43 | 0.34 | 0.31 | 0.27 | 0.20 | 0.23 | 0.32 | 0.35 |

**Table 2** Recall/precision of color flag images

| Method | Precision | | | | Recall | | | |
|---|---|---|---|---|---|---|---|---|
| Beta | 0.84 | 0.68 | 0.57 | 0.48 | 0.35 | 0.56 | 0.69 | 0.77 |
| Tolias et al. | 0.51 | 0.38 | 0.32 | 0.27 | 0.22 | 0.27 | 0.33 | 0.42 |
| Vertan | 0.78 | 0.65 | 0.57 | 0.44 | 0.32 | 0.52 | 0.62 | 0.67 |
| Swain, Ballard | 0.81 | 0.65 | 0.52 | 0.45 | 0.33 | 0.53 | 0.65 | 0.73 |

**Table 3** Recall/precision of color logo images

| Method | Precision | | | | Recall | | | |
|---|---|---|---|---|---|---|---|---|
| Beta | 0.75 | 0.53 | 0.52 | 0.41 | 0.29 | 0.39 | 0.49 | 0.60 |
| Tolias et al. | 0.45 | 0.25 | 0.22 | 0.20 | 0.18 | 0.21 | 0.27 | 0.36 |
| Vertan | 0.70 | 0.50 | 0.44 | 0.39 | 0.27 | 0.36 | 0.47 | 0.57 |
| Swain, Ballard | 0.47 | 0.33 | 0.31 | 0.28 | 0.19 | 0.25 | 0.35 | 0.40 |

when $n \leq 5$. With color logo images,the precision rate is 0.47 and the recall rate is 0.19 when $n \leq 5$.

So with the proposed Beta membership function, the retrieved results are much higher than Tolias et al., Vertan and Boujemaa and a little better than non-fuzzy method for color flag images. In case of textured images, the results are also better than Vertan and Boujemaa but much better than Tolias et al and Swain and Ballard method. While in case of color logo images, the results are also better than Vertan and Boujemaa, Tolias et al., and Swain and Ballard method. As GTI is a human

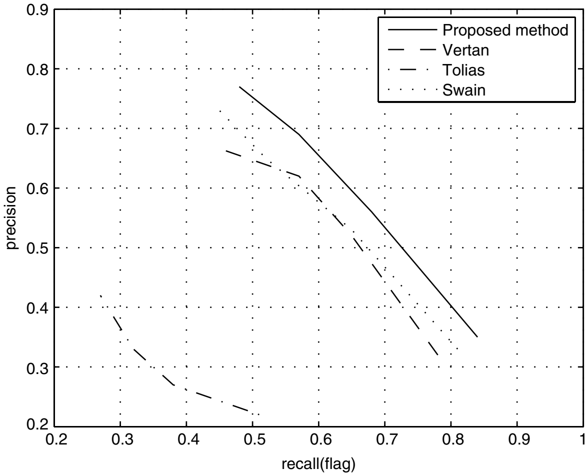



**Fig. 5** Precision/Recall of color flag images

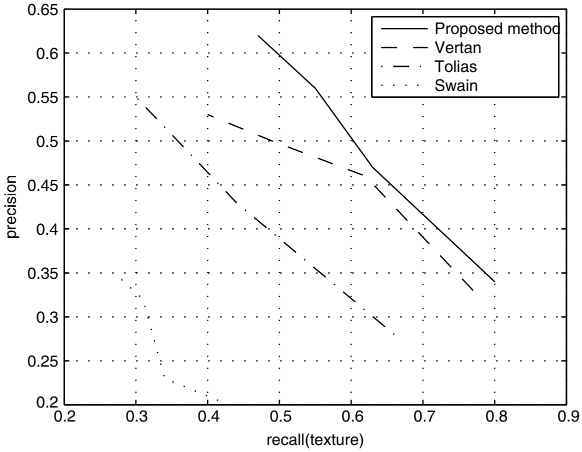



**Fig. 6** Precision/Recall of color texture images

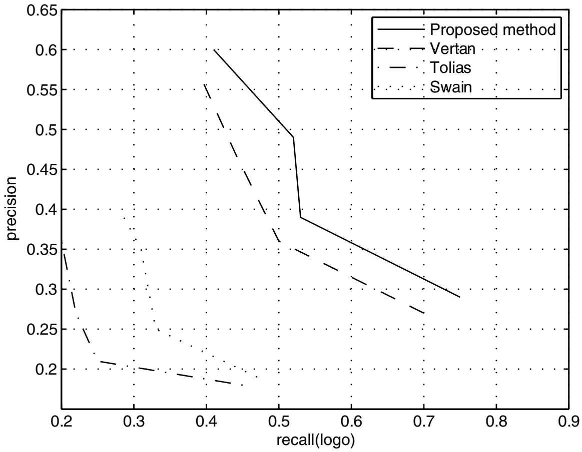



**Fig. 7** Precision/Recall of color logo images

perceptual model that takes into account both distinctive and common features, the results will be better. Sample results of the proposed method on color flag images and textured images are shown below.

The following graphs show the recall vs. precision of color flag, color texture, and color logo images.

## 6 Conclusion

In this chapter, fuzzy measures and fuzzy integrals, different kinds of measures of fuzziness, and fuzzy distance measures are reviewed. Also their application in region based segmentation, color image retrieval, and thresholding are discussed.

Apart from this, a new fuzzy based method has been proposed for color image retrieval using a new membership function called Beta membership function, when

Generalized Tversky's index has been used as a similarity measure. Generalized Tversky's index is a human perceptual measure that takes into account the common features as well as distinctive features.The results are found much better when compared with the existing fuzzy and non-fuzzy methods.

## References


1. Sugeno M (1977) Fuzzy measures and fuzzy integrals – A survey. In: Gupta, Saridis and Gaines (eds.) Fuzzy Automata and Decision Processes: 89–102.
2. Frigui H (2001) Interactive image retrieval using fuzzy sets. Pattern Recognition Letters 22: 1021–1031.
3. Murofishi T, Sugeno M (1991) A theory of fuzzy measures: representation, the Choquet integral, and null sets. J. of Mathematical Analysis and Applications 159: 532–549.
4. Li J, Chi Z, Chen G (2004) Image retrieval using Sugeno integral. Third IEEE Intl. Conferance on Image and Graphics (ICIG04), HongKong, China.
5. Y S Choi, D Kim, R Krishnapuram (1998) Relevance feedback for content based image retrieval using the Choquet integral. IEEE Conferance on Multimedia ans Expo, 1207–1210.
6. Pham T D, Yan H (1999) Color image segmentation using fuzzy integral and mountain clustering. Fuzzy Sets and Systems 107: 121–130.
7. Pham T D, Yan H (1999) Color image segmentation using fuzzy integral and mountain clustering. Fuzzy Sets and Systems 107: 121–130.
8. L. A. Zadeh, Fuzzy sets and systems, Proc. of Symposiam on systems theory, Polytechnic Institute of Brooklyn, NY, USA, 1965.
9. Luca De A, Termini S (1979) Entropy and energy measures of fuzzy sets: In Gupta M M, Ragade R K ,and Yager R R (eds) Advances in Fuzzy sets Theory and Applications. North Holland, Amsterdam.
10. Kaufmann A (1980) Introduction to the theory of fuzzy subsets: Fundamental theoretical elements-1. Academic Press, NewYork.
11. Yager R R (1979) On the measure of fuzziness and negation. I. Membership in unit interval. Intl. Journal Gen. Systems 5: 221–229.
12. Pal N R, Pal S K (1989) Entropy, a new definition and its application. IEEE Proc.36: 284–295.
13. Rosenfeld A (1984) The fuzzy geometry of image subsets, Pattern Recognition Letters 2: 311–317.
14. Tolias Y K, Panas S M et al. (2001) Generalized fuzzy indices for similarity matching. Fuzzy Sets and Systems 120: 255–270.
15. Tversky A (1977) Features of similarity. Psychological Review 84(4): 327–352.
16. Chaira T, Ray A K (2005) Fuzzy measures for color image retrieval. Fuzzy Sets and Systems Vol. 150(3): 545–560.
17. Huang L K, Wang M J J (1995) Image thresholding by minimizing the measure of fuzziness. Pattern Recognition 28: 41–51.
18. Chaira T, Ray A K (2004) Threshold selection using fuzzy set theory. Pattern Recognition Letters Vol. 25: 865–874.
19. Chaira T, Ray A K (2004) Fuzzy measures for threshold selection. The Journal of Fuzzy Mathematics Vol. 12. No.1 (in press).
20. Chaira T (2003) Image segmentation and color retrieval: a fuzzy and intuitionistic fuzzy set theoretic approach. PhD theses, Indian Institire of Technology, Kharagpur, India.
21. Bimbo A D (2001) Visual information Retrieval, Morgan Kaufmann Publishers Inc., San Francisco, USA.
22. Vertan C, Boujemaa N (2000) Using Fuzzy histograms and distances for color image retrieval. Challenge of image retrieval, Brighton: 1–6.
23. Swain, M J, Ballard D H (1991) Color indexing. International journal of computer vision 7(1): 11–32.

# Type II Fuzzy Image Segmentation

**Hamid R. Tizhoosh**

**Abstract** Type II fuzzy sets are high-level representation of vague data with, compared to ordinary fuzzy sets, greater capability for uncertainty management. Theoretical aspects of type II fuzzy systems have been extensively investigated, and the research is still ongoing. Many image processing tasks accompanied with different types of imperfection. In this chapter, the applications of type II fuzzy sets for image segmentation will be discussed. Global and spatial type II segmentation schemes will be systematically introduced and examples will be provided.

## 1 Introduction

In many image processing applications, we often have to segment gray-level images into meaningful regions. In some cases, the image contains a background and one or more objects. Image segmentation mainly serves as a tool for feature extraction and object recognition. Image thresholding, the simplest form of segmentation, is generally a two-class clustering procedure. Many segmentation techniques already exist [5, 19, 38, 39] and extensive research has introduced new and more robust thresholding techniques [12, 16, 26, 33]. Sankura and Sezginb list over 40 different thresholding techniques [25].

Fuzzy techniques are suitable for development of new image processing algorithms because as nonlinear knowledge-based methods, they are able to remove grayness ambiguities in a robust way [29]. The concept of *ultrafuzziness* aims at capturing/eliminating the uncertainties within fuzzy systems using regular (type I) fuzzy sets. Ultrafuzzy sets should not only remove the vagueness/imprecision in the data but also the uncertainty in assigning membership values to the data, which is generally the reflection of inherent stochasticity.

This chapter is organized as follows: Sect. 2 describes briefly the type II fuzzy sets. In Sect. 3 a measure of ultrafuzziness will be reviewed. Sect. 4 introduces image thresholding using type II fuzzy sets by means of the measure of ultrafuzziness. A simple type-II approach for segmenting images will be introduced in Sect. 5. Finally, the chapter will be summarized with some conclusions and directions for future work in Sect. 6.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

## 2 Type II Fuzzy Sets

Type I fuzzy sets [37], and fuzzy inference systems in particular, have become a powerful tool in dealing with vague and imprecise data. The main problem with type I fuzzy sets, however, is the uncertain assignment of a membership degree to an element/pixel regardless of which shape we use and what algorithm is applied. Membership functions are usually defined by an expert and are based on his (mainly) subjective knowledge. The fact that different fuzzy approaches to solve a given problem mainly differ in the way that they define the membership function is for the most part due to this dilemma. To remove this uncertainty, type II fuzzy sets should be utilized.

There are different sources of uncertainties in type I fuzzy sets [14, 32]; the meanings of the words that are used, and noisy measurements. "Type I fuzzy sets are not able to directly model such uncertainties because their membership functions are totally crisp. On the other hand, type-2 fuzzy sets are able to model such uncertainties because their membership functions are themselves fuzzy." [Mendel et al. [14]]

The term *footprint of uncertainty* (FOU) is used in literature to verbalize the irregular shape of a type II fuzzy set (gray area in Fig. 1)[13, 14]. The FOU implies that there is a distribution that sits on top of that gray area. When they all equal one, the resulting type II fuzzy sets are called *interval type II fuzzy sets*. Since we are not certain about membership assignment fuzzy sets of type II are fuzzy sets for which the membership function does not deliver a single value for every element, but rather an interval.

**Definition** A type II fuzzy set $\tilde{A}$ is defined by a type II membership function $\mu_{\tilde{A}}(x, u)$, where $x \in X$ and $u \in J_x \subseteq [0, 1]$, i.e. [13],

$$\tilde{A} = \left\{ \left( (x, u), \mu_{\tilde{A}}(x, u) \right) | \quad \forall x \in X, \quad \forall u \in J_x \subseteq [0, 1] \right\}, \tag{1}$$

in which $0 \leq \mu_{\tilde{A}}(x, u) \leq 1$. $\tilde{A}$ can also be expressed in usual notation of fuzzy sets as

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u)/(x, u) \qquad J_x \subseteq [0, 1], \tag{2}$$

where the double integral denotes the union over all $x$ and $u$.

In order to define a type II fuzzy set one can define a type I fuzzy set and assign upper and lower membership degrees to each element to (re)construct the footprint of uncertainty (Fig. 1). A more practical definition for a type II fuzzy set can be given as follows [32]:

$$\tilde{A} = \{(x, \mu_U(x), \mu_L(x))| \quad \forall x \in X, \mu_L(x) \leq \mu(x) \leq \mu_U(x), \mu \in [0, 1]\} \tag{3}$$

To define the lower and upper membership functions, $\mu_L$ and $\mu_U$, linguistic hedges can be employed to modify the skeleton function [32]. Hedges are generally available as pairs, which represent diagonally different modification of a basic term.

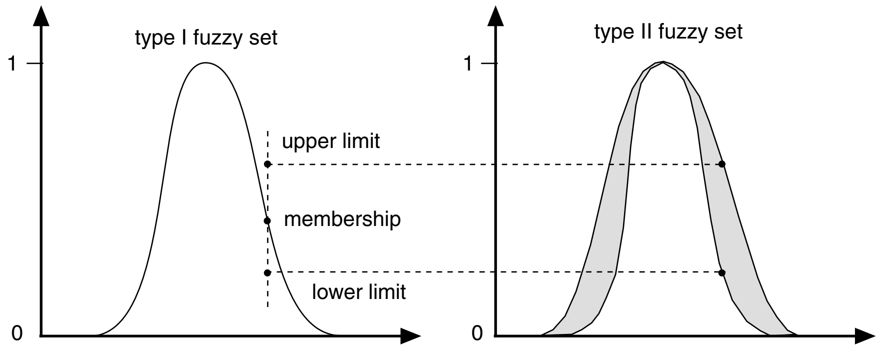



**Fig. 1** Constructing type II fuzzy sets. The interval between lower and upper membership values (**gray region**) represents the *footprint of uncertainty* [32]

It seems that it is more practical to use a linguistic hedge and its reciprocal value to represent the footprint of uncertainty:

$$\mu_U(x) = [\mu(x)]^{\frac{1}{\alpha}}, \tag{4}$$

$$\mu_L(x) = [\mu(x)]^{\alpha}, \tag{5}$$

where $\alpha \in (1, \infty)$. Well-established linguistic hedges like *dilation* and *concentration*

$$\mu_U(x) = [\mu(x)]^{0.5}, \tag{6}$$

$$\mu_L(x) = [\mu(x)]^{2}, \tag{7}$$

and *deaccentuation* and *accentuation*

$$\mu_U(x) = [\mu(x)]^{0.75}, \tag{8}$$

$$\mu_L(x) = [\mu(x)]^{1.25}, \tag{9}$$

can be used to represent the FOU.

## 3 Measuring Ultrafuzziness

If we interpret images, their segments, histograms or thresholds as type II fuzzy sets, then the question arises how ultrafuzzy is a fuzzy set, that is, to what degree are membership values uncertain. If the degrees of membership can be defined without any uncertainty (ordinary or type I fuzzy sets), then the ultrafuzziness diminishes (=0). However, if the individual membership values can only be given as an interval, then the amount of ultrafuzziness will increase. The maximal ultrafuzziness (= 1) is comparable to total ignorance in measure theory, where we absolutely do not have any knowledge about the nature of membership degrees of the problem at hand. Tizhoosh [32] has defined a measure of ultrafuzziness $\tilde{\gamma}$ for an $M \times N$ image subset $\tilde{A} \subseteq X$ with $L$ gray levels $g \in [0, L-1]$, histogram $h(g)$ and membership function $\mu_{\tilde{A}}(g)$, as follows:

$$\tilde{\gamma}(\tilde{A}) = \frac{1}{MN} \sum_{g=0}^{L-1} h(g) \times [\mu_U(g) - \mu_L(g)], \tag{10}$$

where upper and lower memberships are calculated according to (4) and (5). For the spatial case, the ultrafuzziness can be calculated as follows [32] (also see [1, 24]):

$$\tilde{\gamma}(\tilde{A}) = \frac{1}{MN} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} [\mu_U(g_{ij}) - \mu_L(g_{ij})]. \tag{11}$$

This basic definition relies on the assumption that the singletons sitting on the FOU are all equal in height (which is the reason why interval-based type II sets are used). Hence, it can only measure the variation in the length of the FOU.

The measure of ultrafuzziness $\tilde{\gamma}$ has the following properties [32][1]:

1. Minimum Ultrafuzziness: $\tilde{\gamma}(\tilde{A}) = 0$ if $\mu_{\tilde{A}}$ is a type I fuzzy set ($\forall g \in X \mu_U(g) = \mu_L(g)$).
2. Equal Ultrafuzziness: $\tilde{\gamma}(\tilde{A}) = \tilde{\gamma}(\bar{\tilde{A}})$ ($\bar{\tilde{A}}$: complement set)
3. Reduced Ultrafuzziness: $\tilde{\gamma}(\tilde{A}) \geq \tilde{\gamma}(\tilde{A}^{'})$ if $\tilde{A}^{'}$ is an intensified (*crisper*) version of $\tilde{A}$ ($\tilde{A}^{'}$ has a shorter/narrower FOU than $\tilde{A}$).
4. Maximum Ultrafuzziness: $\tilde{\gamma}(\tilde{A}) = 1$ if $\forall g \in X \quad \mu_U(g) - \mu_L(g) = 1$.

## 4 Type II Image Thresholding

A large number of fuzzy techniques for image thresholding have been introduced in literature. Fuzzy clustering, for instance, considers thresholding as a two-class clustering problem [7, 8, 10, 11]. The rule-based approach, as the most popular fuzzy scheme, uses fuzzy if-then rules to find the suitable threshold [30]. Fuzzy-geometrical approaches optimize geometrical measures such as compactness, and index of area coverage [17, 18, 20, 27, 28, 29]. The largest class of fuzzy threshoding

[1] also see the definition of intuitionistic entropy [1].

algorithms is based on information-theoretical measures and minimizes/maximizes measures of fuzziness and image information such as index of fuzziness or crispness, fuzzy entropy, and fuzzy divergence [2, 3, 6, 21, 28, 29, 31, 35, 36].

The most common measure of image fuzziness is the linear index of fuzziness [4, 9, 23, 29, 31, 36]. For an $M \times N$ image subset $A \subseteq X$ with $L$ gray levels $g \in [0, L-1]$, histogram $h(g)$ and membership function $\mu_A(g)$, the linear index of fuzziness $\gamma_l$ can be defined as follows:

$$\gamma_l(A) = \frac{2}{MN} \sum_{g=0}^{L-1} h(g) \times min[\mu_A(g), 1 - \mu_A(g)]. \quad (12)$$

For the spatial case, the fuzziness can be calculated as follows:

$$\gamma_l(A) = \frac{2}{MN} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} min[\mu_A(g_{ij}), 1 - \mu_A(g_{ij})]. \quad (13)$$

Other types of fuzzy negation may be considered as well.

To measure the global or local image fuzziness, a suitable membership function $\mu_A(g)$ should be defined. Different functions are already used in the literature, such as the standard S-function [22, 23] and the Huang/Wang-function [6]. Tizhoosh [28] defined the suitable threshold as an $LR$-type fuzzy number (Fig. 2). Using a fuzzy number seems to be more natural since we usually try to segment the image by means of a preferably single number (a unique threshold for the entire image). Only if this fails, which occurs in many applications, are advanced techniques for adaptive thersholding employed. By moving/shifting the fuzzy threshold across the dynamic range one can calculate the fuzziness and detect positions with minimum/maximum values (Fig. 3). The general algorithm for image thresholding using type I fuzzy sets is given in Tabel 1.

Tizhoosh [32] introduced type-II thresholding using the measure of ultrafuzziness. The general algorithm for image thresholding based on type II fuzzy sets is summarized in Table 2. Thresholding based on this scheme can be formulated as solving the following equation:

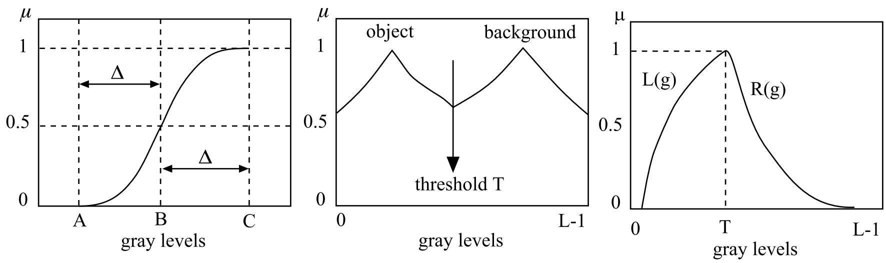



**Fig. 2** Different membership functions for image thresholding. From left to right: S-function used by Pal et al. [18], function used by Huang/Wang [6], and threshold as a fuzzy number used by Tizhoosh [28]. [source: [32]]

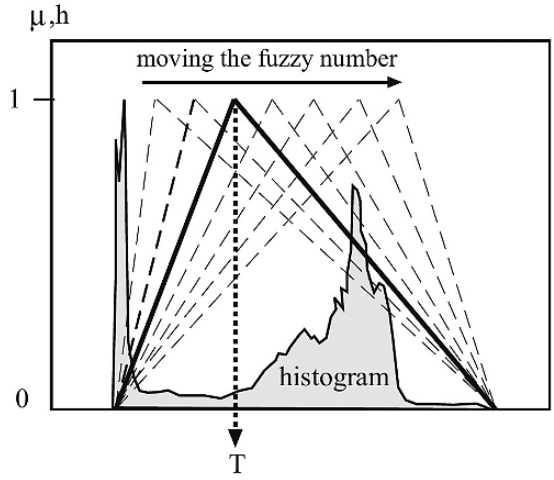



**Fig. 3** The membership function is moved/shifted over the gray-level range to calculate the fuzziness in each position. The maximum fuzziness indicates the optimal threshold. [source: [32]]

$$\frac{\partial}{\partial T}\tilde{\gamma}(\tilde{A}) = \frac{\partial}{\partial T}\frac{1}{MN}\sum_{g=0}^{L-1} h(g) \times [\mu_U(g, T) - \mu_L(g, T)] = 0. \tag{14}$$

Tizhoosh [32] has also compared type I and type II thresholding using different test images and their gold standards. The results are compared using a measure of performance $\eta$, which compares the individual gold images with the binary result delivered by type I and type II thresholding. Based on the misclassification error [25, 34], the performance measure has been defined as

**Table 1** General algorithm for thresholding based on type I fuzzy sets

| |
|---|
| Select the shape of membership function |
| Select the suitable measure of fuzziness (e.g. (12)) |
| Calculate the image histogram |
| Initialize the position of membership function |
| Shift the membership function along the gray-level range (Fig. 3) |
|     Calculate fuzziness in each position, for instance using (12) |
|     Locate the position $g_{opt}$ with minimum/maximum fuzziness |
| Threshold the image with $T = g_{opt}$ |

**Table 2** General algorithm for thresholding based on type II fuzzy sets [32]

| |
|---|
| Select the shape of skeleton membership function $\mu(g)$ |
| Initialize $\alpha$ |
| Calculate the image histogram |
| Initialize the position of membership function |
| Shift the membership function along the gray-level range |
|     Calculate the upper and lower membership values $\mu_U(g)$ and $\mu_L(g)$ |
|     Calculate ultrafuzziness in each position (10) |
|     Find out the position $g_{opt}$ with maximum ultrafuzziness |
| Threshold the image with $T = g_{opt}$ |

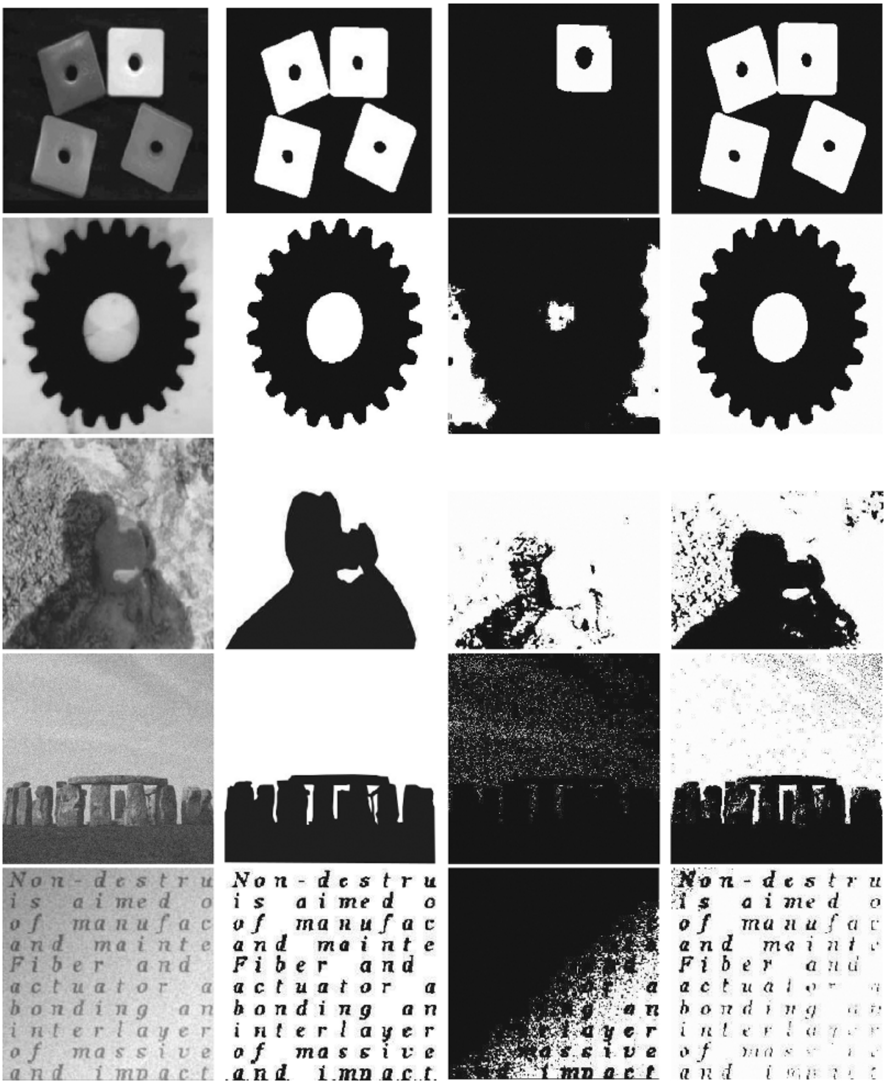


**Fig. 4** From Left to right: Original image, gold standard (created manually), type I segmentation, and type II segmentation

$$\eta = 100 \times \frac{|B_O \cap B_T| + |F_O \cap F_T|}{|B_O| + |F_O|}, \tag{15}$$

where $B_O$ and $F_O$ denote the background and foreground of the original (ground-truth) image, $B_T$ and $F_T$ denote the background and foreground area pixels in the resulting image, and $|\cdot|$ is the cardinality of the set. Sample results are presented in Fig 4. The performance measure $\eta$ for every algorithm is listed in Table 3.

As apparent from Table 3, type II thresholding has an average performance of

**Table 3** Performance of type I and type II methods based on comparison of their results with the gold standards (see Fig.4,(15))

| Image | $\eta_{\text{Type I}}$ | $\eta_{\text{Type II}}$ |
|---|---|---|
| blocks | 71.21 | 98.98 |
| gearwheel | 64.47 | 98.21 |
| shadow | 75.75 | 94.39 |
| stones | 39.96 | 96.99 |
| text | 36.37 | 93.44 |
| $m$ (average) | 57.55 | 94.60 |
| $\sigma$ (standard deviation) | 18.19 | 3.22 |

94.60% with a standard deviation of 3.22%. In contrast, the type I algorithm yields a 57.55% average performance and 18.19% standard deviation, and is clearly inferior to the type II algorithm.

## 5 Type II Segmenation

There exists a large number of techniques applied to image segmentation [5, 19, 38, 39]. The most natural way to employ the ultrafuzziness for segmenting images would be to design a type II fuzzy systems consisting of fuzzy rules applied to each pixel. However, we already know that even type I fuzzy inference has a relatively high cost if applied in an spatial fashion. The benefit of uncertainty management via type II fuzziness diminishes because the high computational expense that one has to pay if type II inference is conducted on spatial windows. Therefore, one should be looking for more feasible implementations of type II image segmentation especially when realtime processing is a given constraint.

In this section, a simple and straightforward type II segmentation procedure will be designed in order to demonstrate how one can exploit the advantage of type II fuzzy sets without complex coding and/or high computation time.

Assume that the camera man in Fig. 5 should be segmented. Since the object of interest (the camera man) is dark, the primary membership function $\mu$, as the skeleton, around which the footprint of uncertainty will be constructed, should be defined first. The primary membership function is usually defined based on given characteristics of the image at hand. For the test image, we can define a fuzzy set **dark pixels** using a $z$-shaped membership function on the interval $[g_{\min}, \frac{g_{\min}+g_{\max}}{2}]$:

$$\mu(g) = \begin{cases} 1 - 8\left(\dfrac{g - g_{\min}}{A}\right)^2 & g_{\min} \leq g \leq \frac{B}{4}, \\[2ex] \left(\dfrac{2AB - 8g}{A}\right)^2 & \frac{B}{4} < g < \frac{B}{2}, \\[2ex] 0 & g \geq \frac{B}{2}, \end{cases} \tag{16}$$

**Fig. 5** Test image: The dark object (**camera man**) should be represented and segmented with a type II fuzzy set

where $A = g_{\max} - g_{\min}$, and $B = g_{\max} + g_{\min}$.

Assuming that the primary (skeleton) membership function $\mu$ exists, the lower and upper membership functions can be calculated:

$$\mu_L(i, j) = \mu(g(i, j))^{\alpha(i,j)}, \tag{17}$$

$$\mu_U(i, j) = \mu(g(i, j))^{\frac{1}{\alpha(i,j)}}. \tag{18}$$

The ultrafuzzifier $\alpha$ for each spatial point $(i, j)$ can be determined with regard to some characteristics of the corresponding neighborhood:

$$\alpha(i, j) = F \times \min\left(1, \frac{\max_k g(i+k, j+k) - \min_k g(i+k, j+k)}{L-1}\right), \tag{19}$$

where $k \in \{-n, \ldots, -1, 0, 1, \ldots, n\}$ ($n = \lfloor \frac{W}{2} \rfloor$, $W =$ window size) and the constant $F \in (1, \infty)$ is an amplification factor controlling the length/width of the FOU. This means if the intensity of the center pixel considerably deviates from its immediate surrounding, then the uncertainty of segmentation increases.

In order to save computation time we may just operate on the upper and lower membership functions. A weighted average would be the ideal way to determine a single value for each pixel. The weight $w_L$ for lower membership function $\mu_L$ can be calculated from the ratio of the local and global minimum gray levels:

**Table 4** Type II fuzzy image segmentation

| |
|---|
| Determine the spatial window size (e.g. $3 \times 3, 5 \times 5, \ldots$) |
| Determine $g_{\min} = 255$, $g_{\max} = 0$ |
| For all spatial windows do |
| calculate $\alpha(i, j)$ (19) |
| calculate the weights $w_L$ and $w_U$ (20 and 21) |
| calculate the primary membership $\mu$ |
| calculate the the upper membership $\mu_U(i, j)$ (18) |
| calculate the the lower membership $\mu_L(i, j)$ (17) |
| assign new gray level $g'(i, j)$ (22) |

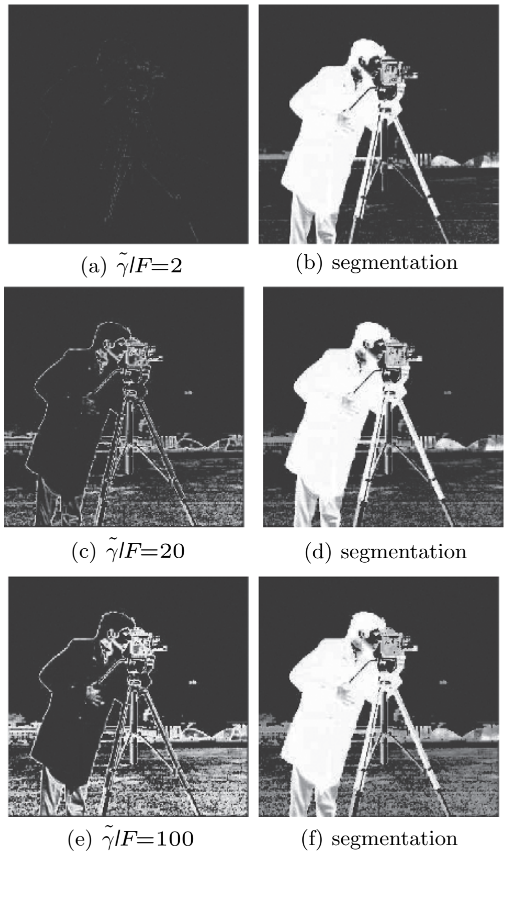



**Fig. 6** Results of type II segmentation. Left to right: the ultrafuzziness $\tilde{\gamma}$ for different $F$ values, and corresponding segmented image

$$w_L(i,j) = \frac{\min_k g(i+k, j+k)}{g_{\max}}. \tag{20}$$

If the spatial neighborhood is dark, then the lower membership will be amplified. Analogously, the weight $w_U$ for the upper membership function $\mu_U$ can be given as

$$w_U(i,j) = \frac{\max_k g(i+k, j+k)}{g_{\max}}. \tag{21}$$

Hence, the segmented pixels $g'$ can be given as

$$g'(i,j) = (L-1) \times \frac{w_U(i,j)\mu_U(g(i,j)) + w_L(i,j)\mu_L(g(i,j))}{w_U(i,j) + w_L(i,j)}. \tag{22}$$

The following algorithm in Table 4 can segment the image based on the ultrafuzziness of gray-level intensities. Figure 6 shows the results of segmentation. As it can be seen, for slight amplification of ultrafuzziness ($F = 2$) only the vicinity of very strong edges can cause uncertainty. As we widen the FOU ($F = 20$ and $F = 100$) more and more weak edges are included.

## 6 Concluding Remarks

Image segmentation is one of the most challenging tasks in image processing and computer vision. Extracting meaningful image regions, due to the inherent imperfections of digital images, requires capable algorithms. Fuzzy sets have been extensively used to develop a new generation of image processing algorithms. However, they suffer from a major drawback; the definition of membership functions is usually uncertain. A large number of researchers in the past have therefore focused on tuning and refinement of fuzzy systems.

Type II fuzzy sets seem to remove the membership uncertainty. This could lead to a new and robust class of segmentation algorithms capable of achieving a higher level of accuracy. How type II fuzzy sets should be embedded within existing fuzzy techniques, however, still needs to be investigated.

In this chapter two simple schemes were discussed to demonstrate how type II fuzzy sets can be implemented with respect to practical limitations. The preliminary results suggest that type II fuzzy sets are superior to their ordinary type I counterparts. Extending existing fuzzy techniques, and deeper and more comprehensive elaborations on advantages and challenges of type II fuzzy sets remains a subject for future work.

## References

1. P. Burillo, H. Bustince, Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets, Fuzzy Sets and Systems, Vol. 78, pp. 305–316, 1996.
2. T. Chaira, A.K. Ray, Segmentation using fuzzy divergence, Pattern Recognition Letters, Vol. 24(12), pp. 1837–1844, 2003.
3. Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition, World Scientific, Singapore, 1997.
4. A. De Luca, S. Termini, A definition of a nonprobabilistic entropy in the setting of fuzzy set theory, Information and Control, Vol. 20, pp. 301–312, 1972.
5. R.M. Haralick, Image segmentation survey, in Fundamentals in Computer Vision, O. D. Faugeras (ed.), Cambridge University Press, Cambridge, pp. 209–224, 1983.
6. L.K. Huang, M.J. Wang, Image thresholding by minimizing the measure of fuzziness. Pattern Recognition, Vol. 28, pp. 41–51, 1995.
7. C.V. Jawahar, P.K. Biswas, A.K. Ray, Investigations on Fuzzy Thresholding Based on Fuzzy Clustering, Patern Recognition, Vol. 30(10), pp. 1605–1613, 1997.
8. C.V. Jawahar, P.K. Biswas, A.K. Ray, Analysis of fuzzy thresholding schemes, Patern Recognition, Vol. 33(8), pp. 1339–1349, 2000.
9. A. Kaufmann, Introduction to the Theory of Fuzzy Subsets - Fundamental Theoretical Elements, Vol. 1. Academic Press, New York, 1975.
10. R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Transactions on Fuzzy Systems, Vol.1(2), pp. 98–110, 1993.
11. R. Kruse, F. Höppner, F., Klawonn, Fuzzy-Clusteranalyse (in German), Vieweg, Braunschweig, 1997.
12. K.C. Lin, Fast image thresholding by finding the zero(s) of the first derivative of between-class variance, Machine Vision and Applications, Vol. 13(5-6), pp. 254–262, 2003.
13. J.M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems, Prentice Hall, USA, 2001.
14. J.M. Mendel and R.I. Bob John, Type-2 Fuzzy Sets Made Simple, IEEE Transactions on Fuzzy Systems, Vol. 10(2), pp. 117–127, 2002.
15. H.B. Mitchell, Pattern recognition using type-II fuzzy sets, Information Sciences, Vol. 170, pp. 409–418, 2005.
16. P.L. Rosin, Unimodal thresholding, Pattern Recognition, Vol. 34(11), pp. 2083–2096, 2001.
17. S.K. Pal, A. Ghosh, A., Fuzzy geometry in image analysis, Fuzzy Sets and Systems, Vol. 48, pp. 23–40, 1992.
18. S.K. Pal, A. Rosenfeld, Image enhancement and thresholding by optimization of fuzzy compactness, Pattern Recognition Letters, Vol. 7, pp. 77–86, 1988.
19. N.R. Pal, S.K. Pal, A Review on Image Segmentation Techniques. Pattern Recognition, Vol. 26, pp. 1277–1294, 1993.
20. S.K. Pal, A. Ghosh, Index of area coverage of fuzzy image subsets and object extraction, Pattern Recognition Letters, Vol. 11, pp. 831–841, 1990.
21. N.R. Pal, D. Bhandari, D.D. Majumder, Fuzzy divergence, probability measure of fuzzy events and image thresholding, Patern Recognition Letters, Vol. 13, pp. 857–867, 1992.
22. S.K. Pal, C.A. Murthy, Fuzzy thresholding: mathematical framework, bound functions and weighted moving average technique, Pattern Recognition Letters, Vol. 11, pp. 197–206, 1990.
23. N.R. Pal, J.C. Bezdek, Measures of Fuzziness: A Review and several New Classes,. In: Yager, R.R., Zadeh, L.A. (eds.): Fuzzy Sets, Neural Networks, and Soft Computing, Van Nostrand Reinhold, New York, pp. 194–212, 1994.
24. R. Sambuc, Functions Φ-flous. Aplication a laide au diagnostic en pathologie thyrodene, Thse, Universite de Marseille, 1975.
25. B. Sankur, M. Sezgin, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging, Vol. 13(1), pp. 146–165, 2004.
26. L. Snidaro, G.L. Foresti, Real-time thresholding with Euler numbers, Pattern Recognition Letters, Vol. 24(9-10), pp. 1543–1554, 2003.

27. W. Tao, H. Burkhardt, An Effective Thresholding Method Using a Fuzzy Compactness Measure, Proc. of 12th ICIP, Vol. 1, pp. 47–51, 1994.
28. H.R. Tizhoosh, Über Binarisierung und Potentiale der Fuzzy-Ansätze. In: Kruse, R., Dassow, J., Informatik'98, Springer, pp. 97–106, 1998.
29. H.R. Tizhoosh, Fuzzy Image Processing (in German), Springer, Heidelberg, Germany, 1998.
30. H.R. Tizhoosh, G. Krell, T. Lilienblum, C.J. Moore, B. Michaelis, Enhancement and Associative Restoration of Electronic Portal Images in Radiotherapy. International Journal of Medical Informatics, Elsevier Science Ireland, Vol. 49(2), pp. 157–171, 1998.
31. H.R. Tizhoosh, H. Haußecker, H., Fuzzy Image Processing: An Overview. In: Jähne, B., Haußecker, H., Geißler, P. (Eds.), Handbook on Computer Vision and Applications, Academic Press, Boston, 1998.
32. H.R. Tizhoosh, Image Thresholding Using Type II Fuzzy Sets, Pattern Recognition, Vol. 38, pp. 2363–2372, 2005.
33. H. Yan, Unified formulation of a class of image thresholding techniques, Pattern Recognition, Vol. 29(12), pp. 2025–2032, 1996.
34. W. A. Yasnoff, J. K. Mui, and J. W. Bacus, Error measures for scene segmentation, Pattern Recogn, Vol. 9, pp. 217–231, 1977.
35. Q. Wang, Z. Chi, R. Zhao, Image Thresholding by Maximizing the Index of Nonfuzziness of the 2-D Grayscale Histogram, Computer Vision and Image Understanding, Vol. 85(2), pp. 100–116, 2002.
36. S.D. Zenzo, L. Cinque, S. Levialdi, S., Image Thresholding Using Fuzzy Entropies, SMC, Vol. 28(1), pp. 15–23, 1998.
37. L.A. Zadeh, Fuzzy sets, Information and Control, Vol. 8, pp. 338–353, 1965.
38. Y. Zhang, A survey on evaluation methods for image segmentation, Pattern Recognition, Vol. 29(8), pp. 1335–1346, 1996.
39. Y.-J. Zhang, Advances in Image And Video Segmentation, IRM Press, 2006.

# Image Threshold Computation by Modelizing Knowledge/Unknowledge by Means of Atanassov's Intuitionistic Fuzzy Sets

**Humberto Bustince, Miguel Pagola, Pedro Melo-Pinto, Edurne Barrenechea and Pedro Couto**

**Abstract** In this chapter, a new thresholding technique using Atanassov's intuitionistic fuzzy sets (A-IFSs) and restricted dissimilarity functions is introduced. We interpret the intuitionistic fuzzy index of Atanassov as the degree of unknowledge/ignorance of an expert for determining whether a pixel of an image belongs to the background or the object of the image. Under these conditions we construct an algorithm on the basis of A-IFSs for detecting the threshold of an image. Then we present a method for selecting from a set of thresholds of an image the best one. This method is based on the concept of fuzzy similarity. Lastly, we prove that in most cases our algorithm for selecting the best threshold takes the threshold calculated with the algorithm constructed on the basis of A-IFSs.

## 1 Introduction

The division of an image into regions is called segmentation. In reality, the segmentation of digital images is the process of dividing an image into disjointed parts, regions or classes so that each one has very specific attributes or properties. Each of these classes represents an object of the image.

One of the most commonly used methods for segmenting images is global thresholding or segmentation by levels of gray. In this method the objects of the image are supposedly characterized by different levels of gray. If the image is made up of one object, global thresholding is classifying the pixels into two regions, background and object. This classification is done establishing a threshold $t$ from which the pixels with the highest intensities belong to the background (or object) and those with lowest intensities belonging to the object (or background).

We will denote the coordinates of each pixel on the image $Q$ with $(x, y)$; the intensity or level of gray of the pixel $(x, y)$ with $q(x, y)$ so that $0 \leq q(x, y) \leq L - 1$ for each $(x, y) \in Q$, where $L - 1$ is the highest level of gray of our grayscale.

Normally an image is composed of many objects, therefore in practice it is necessary to choose more than one threshold in order to segment the image. In this chapter we will only consider images containing a single object.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* © Springer 2008

The choice of the best threshold $t$ is a difficult process mainly for two reasons:

1. The histogram does not determine the image in a single way, that is, there can be two equal histograms that correspond to two different images. Ambiguity in the predicate (property or condition) that must be met by the pixels of the object, due to the presence of overlap in the histogram associated with the image.
2. Presence of noise.

Many methods have been proposed for determining the threshold $t$ of an image [14, 15, 16, 18, 22, 23, 24, 25] nevertheless, considering that fuzzy set theory [30] has worked well in the treatment of models that present ambiguity and highly noisy data, this theory is an interesting alternative for determining the best threshold, in order to obtain a good segmentation of the image considered [4, 11, 17, 18, 19, 21]. Within the framework of this theory the most popular algorithms are those that use the concept of fuzzy entropy [4, 13, 17]. In this sense one of the most commonly used is the following:

a. Assign $L$ fuzzy sets $Q_t$ to each image $Q$. Each one is associated to a level of intensity $t$, ($t = 0, 1, \cdots, L-1$), of the grayscale $L$ used.
b. Calculate the entropy of each one of the $L$ fuzzy sets $Q_t$ associated with $Q$.
c. Take, the *best threshold* gray level $t$, associated with the fuzzy set corresponding to the lowest entropy. The justification for this choice is explained in [13, 17].

The main problem of this algorithm is step (a). The question considered is: which is the best function we should take in order to assign the membership of each element/pixel of the image to the associated fuzzy set? [27]. In order to solve this problem we will associate each pixel with three numerical values:

- A value for representing its membership to the background, which we will interpret as the knowledge of the expert of the membership of the pixel to the background. We obtain this value by means of the membership function associated with the set that represents the background. In this chapter this function will be constructed by the expert using restricted dissimilarity functions (see [3]).
- A value for representing its membership to the object, which we will interpret as the knowledge of the expert of the membership of the pixel to the object. This value is obtained by means of the membership function associated with the set that represents the object. This function will also be constructed using restricted dissimilarity functions.
- A value for representing the unknowledge/ignorance of the expert in determining the membership functions described in the first two items. This value will be represented by Atanassov's intuitionistic index.

Under these conditions, if Atanassov's intuitionistic index associated with a pixel has a value of zero, it means that the expert is positively sure about the belongingness of the pixel to the background or to the object. However, if the expert does not know that the pixel belongs to the background or to the object he must represent its membership to both with the value 0.5, and under these conditions we can say that the expert has used the greatest unknowledge/ignorance/intuition allowed in the construction of the functions of membership to the background and to the object respectively. For this reason we will use the A-IFSs (Atanassov's Intuitionistic Fuzzy Set [1, 2]).

## 2 A-IFSs

In 1983 K. Atanassov [1] introduced A-IFSs in the following way:

Let $X$ be an ordinary finite non-empty set. An *A-IFS* in $X$ is an expression $\tilde{A}$ given by

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x))|x \in X\}$$

where

$$\mu_{\tilde{A}} : X \longrightarrow [0, 1]$$
$$\nu_{\tilde{A}} : X \longrightarrow [0, 1]$$

satisfy the condition $0 \leq \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \leq 1$ for all $x$ in $X$.

The numbers $\mu_{\tilde{A}}(x)$ and $\nu_{\tilde{A}}(x)$ denote respectively the degree of membership and the degree of non-membership of the element $x$ in set $\tilde{A}$. We will represent as A-IFSs(X) the set of all the A-IFSs in $X$.

We know that fuzzy sets are represented exclusively by the membership function,

$$A = \{(x, \mu_A(x))|x \in X\}.$$

Hereinafter, fuzzy sets will have associated (besides the membership degree that defines them) a non-membership degree given by one minus the membership degree, that is:

$$A = \{(x, \mu_A(x), \nu_A(x))|x \in X\} = \{(x, \mu_A(x), 1 - \mu_A(x))|x \in X\}. \quad (1)$$

Since $\mu_A(x) + \nu_A(x) = \mu_A(x) + 1 - \mu_A(x) = 1$, in this sense fuzzy sets will be considered as a particular case of A-IFSs [1, 5, 8, 12]. We will represent as $FSs(X)$ the set of all the fuzzy sets in $X$. We also will represent as $\mathbb{I}$ the following fuzzy set $\{(x, \mu_{\mathbb{I}}(x) = 1)|x \in X\}$.

We will call [1]

$$\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x)$$

*Atanassov's intuitionistic fuzzy index* of the element $x$ in the A-IFS $\tilde{A}$. Naturally, if the set $A$ considered is fuzzy, then $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) = 1 - \mu_A(x) - 1 + \mu_A(x) = 0$.

For every $\tilde{A}, \tilde{B} \in$ A-IFSs(X) the following expressions are known [1, 2, 7, 8, 12]:

1. $\tilde{A} \leq \tilde{B}$ if and only if $\mu_{\tilde{A}}(x) \leq \mu_{\tilde{B}}(x)$ and $\nu_{\tilde{A}}(x) \geq \nu_{\tilde{B}}(x)$ for all $x \in X$
2. $\tilde{A} \preceq \tilde{B}$ if and only if $\mu_{\tilde{A}}(x) \leq \mu_{\tilde{B}}(x)$ and $\nu_{\tilde{A}}(x) \leq \nu_{\tilde{B}}(x)$ for all $x \in X$
3. $\tilde{A} = \tilde{B}$ if and only if $\tilde{A} \leq \tilde{B}$ and $\tilde{B} \leq \tilde{A}$
4. K. Atanassov [1, 2] defined the complementary of an IFS as follows: $\tilde{A}_c = \{(x, \nu_{\tilde{A}}(x), \mu_{\tilde{A}}(x)) | x \in X\}$.
5. $\vee(\tilde{A}, \tilde{B}) = \{(x, \vee(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)), \wedge(\nu_{\tilde{A}}(x), \nu_{\tilde{B}}(x))) | x \in X\}$.
6. $\wedge(\tilde{A}, \tilde{B}) = \{(x, \wedge(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)), \vee(\nu_{\tilde{A}}(x), \nu_{\tilde{B}}(x))) | x \in X\}$.

## 2.1 Entropy on A-IFSs

Much has been written on entropy on A-IFSs (see [5, 26]). We will focus on the definition and characterization of this concept presented in [5]. In this chapter entropy on A-IFSs is defined as a magnitude that measures the degree of A-IFS that a set has with respect to the fuzziness of the said set . The idea can be specified in the following conditions:

1. The entropy will be null when the set is a $FSs(X)$,
2. the entropy will be maximum if the set is totally intuitionistic; that is, $\mu(x) = \nu(x) = 0$ for all $x \in X$,
3. as in fuzzy sets, the entropy of an A-IFS will be equal to its respective complementary,
4. if the degree of membership and the degree of non-membership of each element increase, the sum will do as well, and therefore, this set becomes more fuzzy, and therefore the entropy should decrease.

Taking into account the above considerations, we have the following definition.

**Definition 1.** *A real function $IE : A - IFSs(X) \rightarrow R^+$ is called an entropy on A-IFSs(X) if it has the following properties:*
*$(IE1)$ $IE(\tilde{A}) = 0$ if and only if $A \in FSs(X)$,*
*$(IE2)$ $IE(\tilde{A}) = Cardinal(X) = n$ if and only if $\mu_{\tilde{A}} = \nu_{\tilde{A}} = 0$ for all $x \in X$,*
*$(IE3)$ $IE(\tilde{A}) = IE(\tilde{A}_c)$ for all $\tilde{A} \in$ A-IFSs(X),*
*$(IE4)$ if $\tilde{A} \preccurlyeq \tilde{B}$, then $IE(\tilde{A}) \geq IE(\tilde{B})$.*

In [5] it is proven that the most general form of IE is:

$$IE(\tilde{A}) = \frac{1}{n} \sum_{k=1}^{n} \pi_{\tilde{A}}(x_k) \qquad (1)$$

being $n = Cardinal(X)$. In the literature there are other definitions of entropy [26] which are not of use in the application that we present here. In this chapter we will use the expression (1).

## 3 Best Thresholding with A-IFSs

In general terms the algorithm we propose for calculating the best threshold of an image $Q$ is made up of the following steps:

A. Construct $L$ fuzzy sets $Q_{Bt}$ associated with the image $Q$. These sets represent the background of the image $Q$. Each one is associated with a level of intensity $t$, $(t = 0, 1, \cdots, L-1)$, of the grayscale $L$ used.
B. Construct $L$ fuzzy sets $Q_{Ot}$ associated with the image $Q$. These sets represent the object of the image $Q$ considered. Each one is associated with a level of intensity $t$, $(t = 0, 1, \cdots, L-1)$, of the grayscale $L$ used.
C. Represent the unknowledge/ignorance of the expert in the construction of the sets corresponding to items $(A)$ and $(B)$ by means of Atanassov's intuitionistic fuzzy index.
D. Construct $L$ intuitionistic fuzzy sets of Atanassov $\tilde{Q}_{Bt}$ associated with the background of the image.
E. Calculate an entropy $IE$ of each one of the $L$ intuitionistic fuzzy sets of Atanassov $\tilde{Q}_{Bt}$ by means of the expression (1).
F. Take the best threshold $t$ associated with the intuitionistic fuzzy set of Atanassov $\tilde{Q}_{Bt}$ corresponding to the lowest entropy $IE$.

## 4 A Model of the Algorithm with Restricted Dissimilarity Functions

In this section we present a possible development of the steps of the algorithm described in Sect. 3. We also justify each one of them.

### *4.1 Steps* $(A)$ *and* $(B)$

For each image $Q$ we will construct $L$ fuzzy sets $Q_{Bt}$ and another $L$ fuzzy sets $Q_{Ot}$, with $t = 0, 1, \cdots, L-1$. The fuzzy sets $Q_{Bt}$, $Q_{Ot}$ are associated with the background and with the object of the image respectively.

The membership function of each element to the set $Q_{Bt}$ $(Q_{Ot})$ must express the relationship between the intensity $q$ of the pixel in question and its membership to the background (object).

We will define these functions using the concept of *restricted dissimilarity functions*, so that the lower the dissimilarity between the gray level of any pixel $q$ and the mean of the intensities of the background (object) is, the greater will be the value of the membership function $\mu_{Q_{Bt}}(q)$ $(\mu_{Q_{Ot}}(q))$.

For each $t$, the mean of the intensities of gray of the pixels that belong to the background $m_{Bt}$ and the mean of the intensities of gray of the pixels that belong to the object $m_{Ot}$ are given by the following expressions:

$$m_B(t) = \frac{\sum_{q=0}^{t} qh(q)}{\sum_{q=0}^{t} h(q)} \qquad m_O(t) = \frac{\sum_{q=t+1}^{L-1} qh(q)}{\sum_{q=t+1}^{L-1} h(q)},$$

$h(q)$ being the number of pixels of the image with intensity $q$.

Evidently, the first thing we must do is to introduce the concept of restricted dissimilarity function and analyze its main properties. This concept is studied in [3].

**Definition 2.** *A function $d : [0,1]^2 \to [0,1]$ is called a restricted dissimilarity function, if it satisfies the following conditions:*

*1) $d(x, y) = d(y, x)$ for all $x, y \in [0, 1]$;*
*2) $d(x, y) = 1$ if and only if $x = 0$ and $y = 1$ or $x = 1$ and $y = 0$;*
*3) $d(x, y) = 0$ if and only if $x = y$;*
*4) For all $x, y, z \in [0, 1]$, if $x \leq y \leq z$, then $d(x, y) \leq d(x, z)$ and $d(y, z) \leq d(x, z)$.*

Next we present a method for the construction of restricted dissimilarity functions from automorphisms. Constructions of these functions from implication operators can also be found in [3].

We know that a continuous, strictly increasing function $\varphi : [a, b] \to [a, b]$ with boundary conditions $\varphi(a) = a$, $\varphi(b) = b$ is called an *automorphism* of the interval $[a, b] \subset \mathbb{R}$. In this chapter we always use the interval $[0, 1]$.

**Proposition 1.** *If $\varphi_1, \varphi_2$ are two automorphisms of the unit interval, then*

$$d(x, y) = \varphi_1^{-1}(\vee(\varphi_2(x) - \varphi_2(y), \varphi_2(y) - \varphi_2(x)))$$

*is a restricted dissimilarity function.*

**Proposition 2.** *Under the conditions of Proposition 1:*

$$\varphi_1(x) = \varphi_2(x) \text{ for all } x \in [0, 1] \text{ if and only if } d(1, x) = \varphi_1^{-1}(1 - \varphi_1(x)).$$

Once the concept of restricted dissimilarity function has been set and a mechanism for generating such functions has been given, we present a method for the construction of fuzzy sets using these functions.

Under these conditions we will construct fuzzy sets $Q_{Bt}$ and $Q_{Ot}$ by means of $F$ functions with special properties in the following way:

*Let $d$ be a restricted dissimilarity function and let $F : [0, 1] \to [0.5, 1]$ a non increasing function such that*

$$\begin{cases} F(x) = 1 \textit{ if and only if } x = 0 \\ F(x) = 0.5 \textit{ if and only if } x = 1 \end{cases}$$

*In these conditions, given an image $Q$ and an intensity threshold $t$ set, we construct the membership functions of each intensity to the sets $Q_{Bt}$ and $Q_{Ot}$ in the following way:*

$$\mu_{Q_{Bt}}(q) = F(d(\frac{q}{L-1}, \frac{m_B(t)}{L-1}))$$

$$\mu_{Q_{Ot}}(q) = F(d(\frac{q}{L-1}, \frac{m_O(t)}{L-1})).$$

Remark: possible F functions are the following: $F(x) = 1 - 0.5x$, $F(x) = \frac{1}{1+x}$, $F(x) = e^{-x^2.Ln2}$.

In the following proposition (see [3]) we present the main properties of these constructions.

**Proposition 3.** *Under the conditions of the construction above, the following items hold:*

1. $0.5 \le \mu_{Q_{Bt}}(q) \le 1$ *for all* $q$*;*
2. $0.5 \le \mu_{Q_{Ot}}(q) \le 1$ *for all* $q$*;*
3. $\mu_{Q_{Bt}}(q) = 1$ *if and only if* $q = m_B(t)$*;*
4. $\mu_{Q_{Ot}}(q) = 1$ *if and only if* $q = m_O(t)$*;*
5. *If* $q_1 \le q_2 \le m_B(t)$*, then* $\mu_{Q_{Bt}}(q_1) \le \mu_{Q_{Bt}}(q_2)$*,*
   *If* $m_B(t) \le q_1 \le q_2$*, then* $\mu_{Q_{Bt}}(q_1) \ge \mu_{Q_{Bt}}(q_2)$*,*
   *If* $q_1 \le q_2 \le m_O(t)$*, then* $\mu_{Q_{Ot}}(q_1) \le \mu_{Q_{Ot}}(q_2)$*,*
   *If* $m_O(t) \le q_1 \le q_2$*, then* $\mu_{Q_{Ot}}(q_1) \ge \mu_{Q_{Ot}}(q_2)$*.*

It is important to point out that, as a result of the first and the second items of Proposition 3, the membership functions constructed are always greater than or equal to 0.5 (see[17]).

*Example 1.* Let the function $F(x) = 1 - 0.5x$. If we take the automorphisms $\varphi_1(x) = \varphi_2(x) = x$ for all $x \in [0, 1]$ we have the restricted dissimilarity function $d(x, y) = |x - y|$. In these conditions the construction above provides the fuzzy sets $Q_{Bt}$ and $Q_{Ot}$ represented by the following membership functions:

$$\begin{cases} \mu_{Q_{Bt}}(q) = F(d(\frac{q}{L-1}, \frac{m_B(t)}{L-1})) = 1 - 0.5d(\frac{q}{L-1}, \frac{m_B(t)}{L-1}) = 1 - 0.5|\frac{q}{L-1} - \frac{m_B(t)}{L-1}| \\ \mu_{Q_{Ot}}(q) = F(d(\frac{q}{L-1}, \frac{m_O(t)}{L-1})) = 1 - 0.5d(\frac{q}{L-1}, \frac{m_O(t)}{L-1}) = 1 - 0.5|\frac{q}{L-1} - \frac{m_O(t)}{L-1}| \end{cases}$$

We take the following image (Fig. 1) on the grayscale from $0-255$.

For $t = 40$ we have $m_B(t) = 17.03$ and $m_O(t) = 79.26$, so that the fuzzy set $Q_{B40}$ is given by Fig. 2 and $Q_{O40}$ is given by Fig. 3.

Considering the construction method of fuzzy sets presented in this section, we propose Algorithm 1 for generating the $L$ fuzzy sets $Q_{Bt}$ associated with the background of the image $Q$ and the $L$ fuzzy sets $Q_{Ot}$ associated with the object of the same image $Q$.

(1) Choose the automorphisms $\varphi_1$ and $\varphi_2$ and construct the function $d$.
(2) Take the function $F$.
(3) FOR $t := 0$ TO $L-1$ DO
    (3 − 1) Calculate $m_B(t)$ and $m_O(t)$;
    (3 − 2) Construct
$$\mu_{Q_{Bt}}(q) = F(d(\tfrac{q}{L-1}, \tfrac{m_B(t)}{L-1}));$$
$$\mu_{Q_{Ot}}(q) = F(d(\tfrac{q}{L-1}, \tfrac{m_O(t)}{L-1}));$$
ENDFOR.

Alg. 1

## 4.2 Step $(C)$. Justification for the use of A-IFSs

Once the sets $Q_{Bt}$ and $Q_{Ot}$ have been constructed for each $t \in \{0, 1, \cdots, L-1\}$ by means of Algorithm 1 we must ask ourselves if the construction chosen is the best (see[27]).

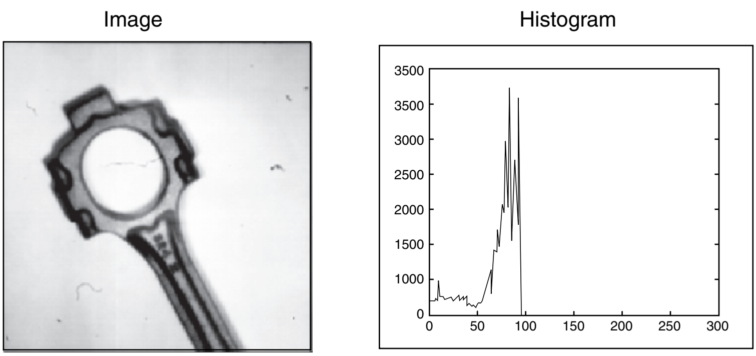



**Fig. 1** Image and histogram of Example 1

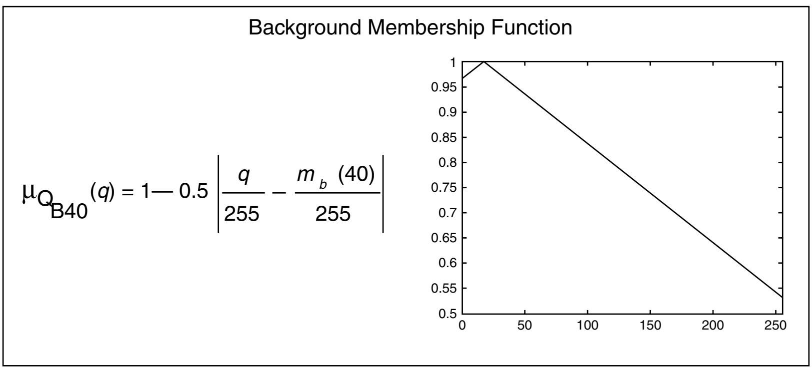



**Fig. 2** Background membership function for $t = 40$

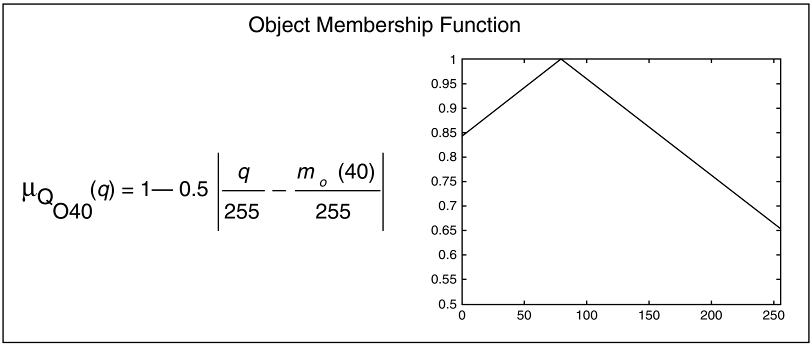



**Fig. 3** Object membership fuction for $t = 40$

We know that in order to separate the object from the background it is necessary to accurately determine the property that must be fulfilled by the pixels that belong to the object. This property establishes the form of the membership function associated with the set that represents the object. Normally this property is not positively known and therefore the choice of membership function is conditioned by the *unknowledge/ ignorance* of the expert who has been asked to construct a membership function.

As we remarked in the introduction, we will interpret Atanassov's intuitionistic index $\pi$ as the *unknowledge/ignorance* of the expert in assigning the membership value of a certain pixel to the background or the object of the image. Under this interpretation of $\pi$, we will consider that $\mu_{Q_{Bt}}$ ($\mu_{Q_{Ot}}$) indicates the expert's degree of knowledge of the pixel belonging to the background (object).

In any case the following conditions must be fulfilled:

1. The unknowledge that the expert uses in the choice of the membership of a pixel must be zero if he is certain that the pixel belongs to the object or the background.
2. The unknowledge/ignorance must decrease with respect to the certainty of the expert as to whether the pixel belongs to the background or to the object.
3. The unknowledge/ignorance must have the least possible influence on the choice of the membership degree. In this chapter we will work on the basis that, in the worst of cases, the unknowledge will have a maximum influence of 25 percent.

In this context, $\pi(q)$ is the quantification of the unknowledge/ignorance of the expert in the selection of the membership functions $\mu_{Q_{Bt}}(q)$ and $\mu_{Q_{Ot}}(q)$. The three considerations above justify the choice of the following expression for $\pi(q)$:

$$\pi(q) = (1 - \mu_{Q_{Bt}}(q))(1 - \mu_{Q_{Ot}}(q)).$$

Evidently, this expression is not the only possible one.

### 4.3 Step (D)

In this section we will associate an A-IFS (using the index $\pi$ described in the subsection above) with each one of the fuzzy sets $Q_{Bt}$ and $Q_{Ot}$, in the following way:

$$\begin{aligned}\tilde{Q}_{Bt} &= \{(q, \mu_{\tilde{Q}_{Bt}}(q), \nu_{\tilde{Q}_{Bt}}(q))|q = 0, 1, \cdots, L-1\}, \text{ given by}\\ &\mu_{\tilde{Q}_{Bt}}(q) = \mu_{Q_{Bt}}(q)\\ &\nu_{\tilde{Q}_{Bt}}(q) = 1 - \mu_{\tilde{Q}_{Bt}}(q) - \pi(q) = (1 - \mu_{Q_{Bt}}(q)) \cdot \mu_{Q_{Ot}}(q)\end{aligned}$$

and

$$\begin{aligned}\tilde{Q}_{Ot} &= \{(q, \mu_{\tilde{Q}_{Ot}}(q), \nu_{\tilde{Q}_{Ot}}(q))|q = 0, 1, \cdots, L-1\}, \text{ given by}\\ &\mu_{\tilde{Q}_{Ot}}(q) = \mu_{Q_{Ot}}(q)\\ &\nu_{\tilde{Q}_{Ot}}(q) = 1 - \mu_{\tilde{Q}_{Ot}}(q) - \pi(q) = (1 - \mu_{Q_{Ot}}(q)) \cdot \mu_{Q_{Bt}}(q)\end{aligned}$$

*Example 2.* In Fig. 4 we show Atanassov's intuitionistic fuzzy sets $\tilde{Q}_{B40}$ and $\tilde{Q}_{O40}$ associated with the image in Fig. 1.

### 4.4 Step (E)

**Proposition 4.** *For each $t \in \{0, 1, \cdots, L-1\}$,*

$$\begin{aligned}IE(\tilde{Q}_{Bt}) =& IE(\tilde{Q}_{Ot}) = IE(\vee(\tilde{Q}_{Bt}, \tilde{Q}_{Ot})) = IE(\wedge(\tilde{Q}_{Bt}, \tilde{Q}_{Ot}))\\ =& \frac{1}{N \times M} \sum_{q=0}^{L-1} h(q)(1 - \mu_{Q_{Bt}}(q))(1 - \mu_{Q_{Ot}}(q)) \qquad (2)\end{aligned}$$

*holds.*

**Proof.** Direct. ■

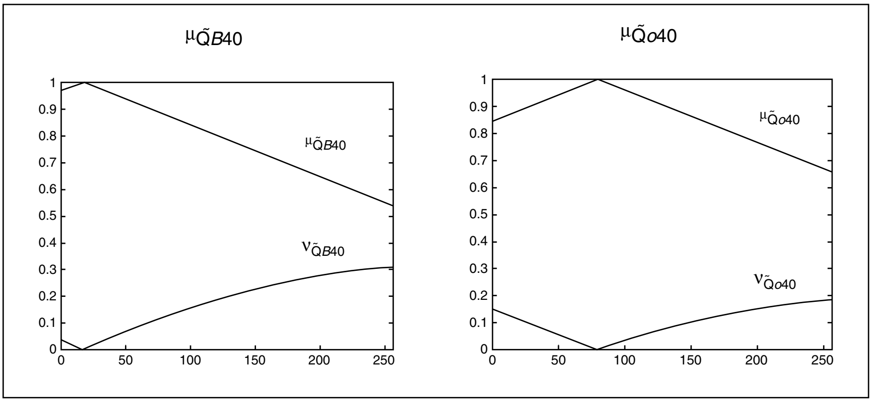



**Fig. 4** Atanassov's intuitionistic fuzzy sets $\tilde{Q}_{B40}$ and $\tilde{Q}_{O40}$

### 4.5 Step $(F)$. Justification for the Minimum Value of IE

**Proposition 5.** *Under our constructions, for each $t \in \{0, 1, \cdots, L-1\}$,*

$$0 \leq IE(\tilde{Q}_{Bt}) \leq 0.25$$

*holds.*

**Proof.** We know by Proposition 3 that the following inequalities hold for all $q$: $0.5 \leq \mu_{Q_{Bt}}(q) \leq 1$ and $0.5 \leq \mu_{Q_{Ot}}(q) \leq 1$. Therefore $1 - \mu_{Q_{Bt}}(q) \leq 0.5$ and $1 - \mu_{Q_{Ot}}(q) \leq 0.5$. Using these inequalities in the expression (2) we have that $IE(\tilde{Q}_{Bt}) \leq 0.5 * 0.5 = 0.25$. Evidently $IE(\tilde{Q}_{Bt}) \geq 0$. ■

From our constructions we deduce the following two items:

1. If $\mu_{Q_{Bt}}(q) \to 1$, then $d(\frac{q}{L-1}, \frac{m_B(t)}{L-1}) \to 0$, therefore $q \approx m_B(t)$. In this case the pixels with intensity $q$ are such that this intensity is very close to the average intensity of the pixels that represent the background. This fact enables us to assure that the pixel in question belongs to the background.
2. If $\mu_{Q_{Ot}}(q) \to 1$, then $d(\frac{q}{L-1}, \frac{m_O(t)}{L-1}) \to 0$, therefore $q \approx m_O(t)$. In this case the pixels with intensity $q$ are such that this intensity is very close to the average intensity of the pixels that represent the object. This fact enables us to assure that the pixel in question belongs to the object.

Therefore, the most representative set of the background $\tilde{Q}_{Bt}$ is that whose membership degrees are closest to one. Identical reasoning can be made for the most representative set of the object $\tilde{Q}_{Ot}$. In any case these sets are obtained by taking, from among all the sets constructed (one for each value of $t$), the set with the lowest intuitionistic fuzzy entropy $IE$. This is due to the fact that expression (2) is close to zero when for each $q$ the following holds: $\mu_{Q_{Bt}}(q) \to 1$ or $\mu_{Q_{Ot}}(q) \to 1$, which is the best possible situation as has been made clear in the two items above.

Please note that it can never happen that $\mu_{Q_{Bt}}(q) = 1 = \mu_{Q_{Ot}}(q)$, the pixel either belongs to the background or it belongs to the object, never to both at the same time.

### 4.6 Algorithm

The theoretical considerations studied in subsections 4.1–4.5 enable us to write the following general algorithm (Algorithm 2) for calculating the threshold of an image using A-IFSs.

Given an image $Q$ of $N \times M$ pixels on the grayscale of 0 to $L-1$.

(1) Take the automorphisms $\varphi_1$ and $\varphi_2$ and construct the function $d$.
(2) Take $F$.
(3) FOR t:= 0 TO $L-1$ DO:
    (3.1) Calculate $m_B(t)$ and $m_O(t)$;
    (3.2) FOR q:= 0 TO $L-1$ DO:
        (3.2.1) Construct
            $\mu_{Q_{Bt}}(q) = F(d(\frac{q}{L-1}, \frac{m_B(t)}{L-1}))$;
            $\mu_{Q_{Ot}}(q) = F(d(\frac{q}{L-1}, \frac{m_O(t)}{L-1}))$;
        (3.2.2) Construct $\pi(q) = (1-\mu_{Q_{Bt}}(q))(1-\mu_{Q_{Ot}}(q))$
        END FOR;
    (3.3) Construct $\tilde{Q}_{Bt}$;
    (3.4) Calculate $IE(\tilde{Q}_{Bt})$;
END FOR.
(4) Take the best threshold value $t$ associated with the set $Q_{Bt}$ corresponding to the lowest intuitionistic fuzzy entropy of Atanassov.

Alg. 3

## 5 Experimental Results

We have applied Algorithm 2 to each one of the 250 images taken from website: http://www.cs.cmu.edu/ cil/vision.html, in the following conditions: $\varphi_1(x) = \varphi_2(x) = x$ for all $x \in [0, 1]$, i. e., $d(x, y) = |x - y|$, and $F(x) = 1 - 0.5x$.

When constructing the binary image from each image using the thresholds obtained we have visually observed that these thresholds produce a good binarization. This fact can be seen in Fig. 5. In this figure we present the thresholds and their corresponding binarized image, from 21 images taken from the set above.

## 6 Comparison with Other Methods

In this section we set ourselves two objectives:

1. To provide a method for selecting from a set of thresholds of an image $Q$, the best one.
2. To prove experimentally on a large set of images taken randomly, that we obtain the best threshold with the algorithms that use A-IFSs, (Algorithm 2).

### *6.1 Best Threshold Selection*

In order to select the best threshold from a set of $n$ thresholds $(t_1, t_2, \cdots, t_n)$, we will use the concept of *fuzzy similarity*. The similarity between sets has been amply

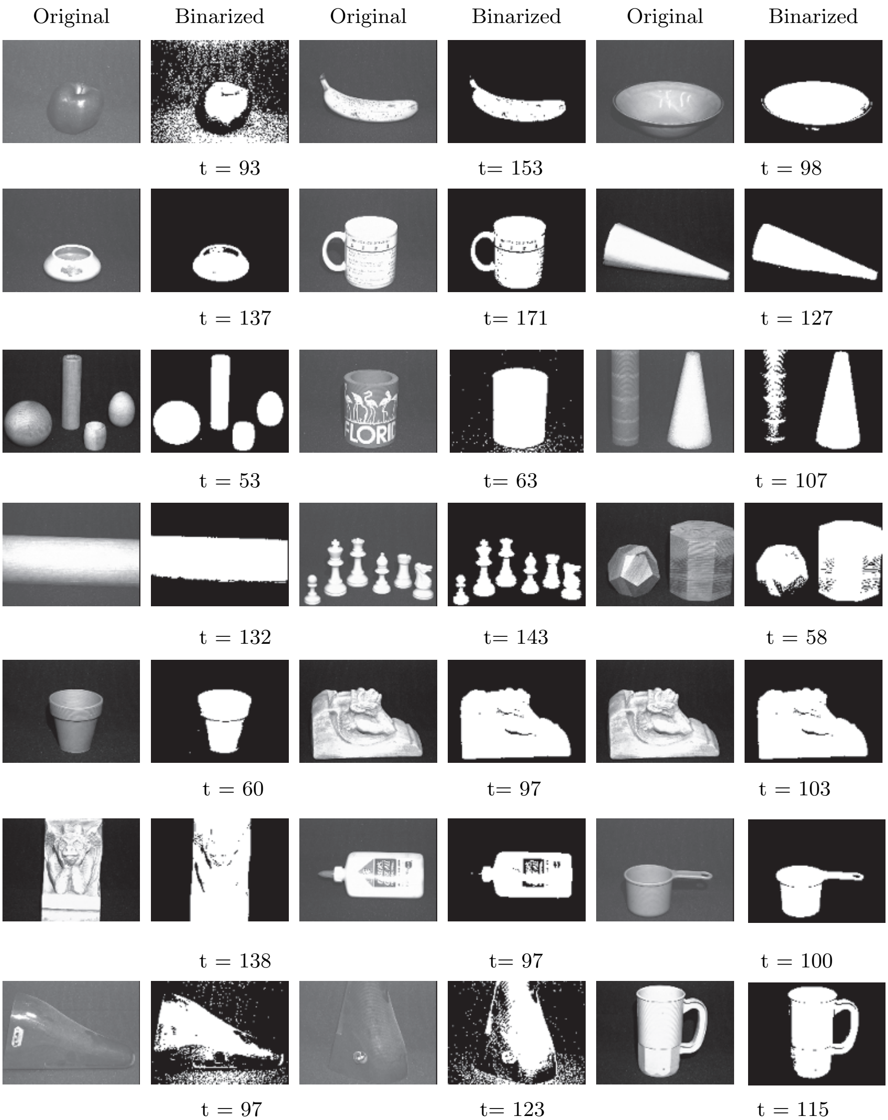



**Fig. 5** Binarized images obtained with Algorithm 2

studied (see [3, 6, 9, 10]), evidently, $SM(A, B) = 1$ if and only if $A = B$. In this chapter we will use the following expression called *Similarity measure based on contrast de-enhancement* (see [6, 10])):

$$SM(A, B) = \frac{1}{n}\sum_{i=1}^{n} 1 - |\mu_A(x_i) - \mu_B(x_i)| \quad \text{for all } A, B \in FSs(X).$$

Under these conditions we present the following algorithm (see [3]).

(1) Select the functions $d$ and $F$ from the construction method developed in Subsection 4.1.
(2) FOR $t := t_1$ TO $t_n$ DO:
    (2.1) Construct $Q_{Bt}$;
    (2.2) Calculate the similarity of $Q_{Bt}$ with the set $\mathbb{I}$;
    ENDFOR.
(3) Take the best threshold value $t$ associated with the set $Q_{Bt}$ corresponding to the greatest similarity with $\mathbb{I}$.

Alg. 3

The selection of the greatest similarity between the set $Q_{Bt}$ and the set $\mathbb{I}$ is justified bearing in mind that in these conditions $\mu_{Q_{Bt}}(q)$ is the one closest to 1 for most of the values of $q$. Considering our constructions, it results that we take as the best threshold the value of $t$ associated with the set with the intensities $q$ most grouped around the mean $m_{Bt}$, i.e. $|\frac{q}{L-1} - \frac{m_{Bt}(q)}{L-1}| \to 0$ for all $q$.

## 6.2 *Selection from Otsu's Threshold, a Fuzzy Threshold and the One Obtained with Algorithm 2*

In this section, using Algorithm 3, we will select the best threshold of the image $Q$ from a set of three thresholds: The one obtained with the classical Otsu method (see [20]), the one achieved with the fuzzy algorithm described in the introduction (steps $(a)-(c)$) and the one obtained with Algorithm 2.

For the fuzzy algorithm we will employ in step $(a)$ the construction developed in Subsection 4.1 and in step $(b)$ Yager's fuzzy entropy with $d = 1$ (see [28, 29]).

In order to calculate the threshold with the A-IFSs method we use Algorithm 2 in the conditions described in the section of experimental results (Sect. 5).

For step 1 of Algorithm 3 we will use $d(x, y) = |x - y|$ for all $x, y \in [0, 1]$ and $F(x) = 1 - 0.5x$.

Under these conditions on a set of 250 images taken from the above-mentioned website, we have found that the greatest similarity, and therefore the best threshold (from our point of view), is provided by Algorithm 2 in 51% of the images, the Fuzzy method in 25% and the Otsu method in 24%.

In Fig. 6 we present a sample composed of 6 images taken from the collection above with their corresponding binary images using the methods Otsu, Fuzzy and A-IFS.

The last column shows the numerical values corresponding, in each case, to the similarity between the fuzzy set constructed from the corresponding threshold (Otsu, Fuzzy and A-IFS) and set $\mathbb{I}$. We have indicated in bold type the highest numerical value. Please note that in all of the cases, except for the second to last row, the greatest similarity is that of the numerical value provided by Algorithm 2.

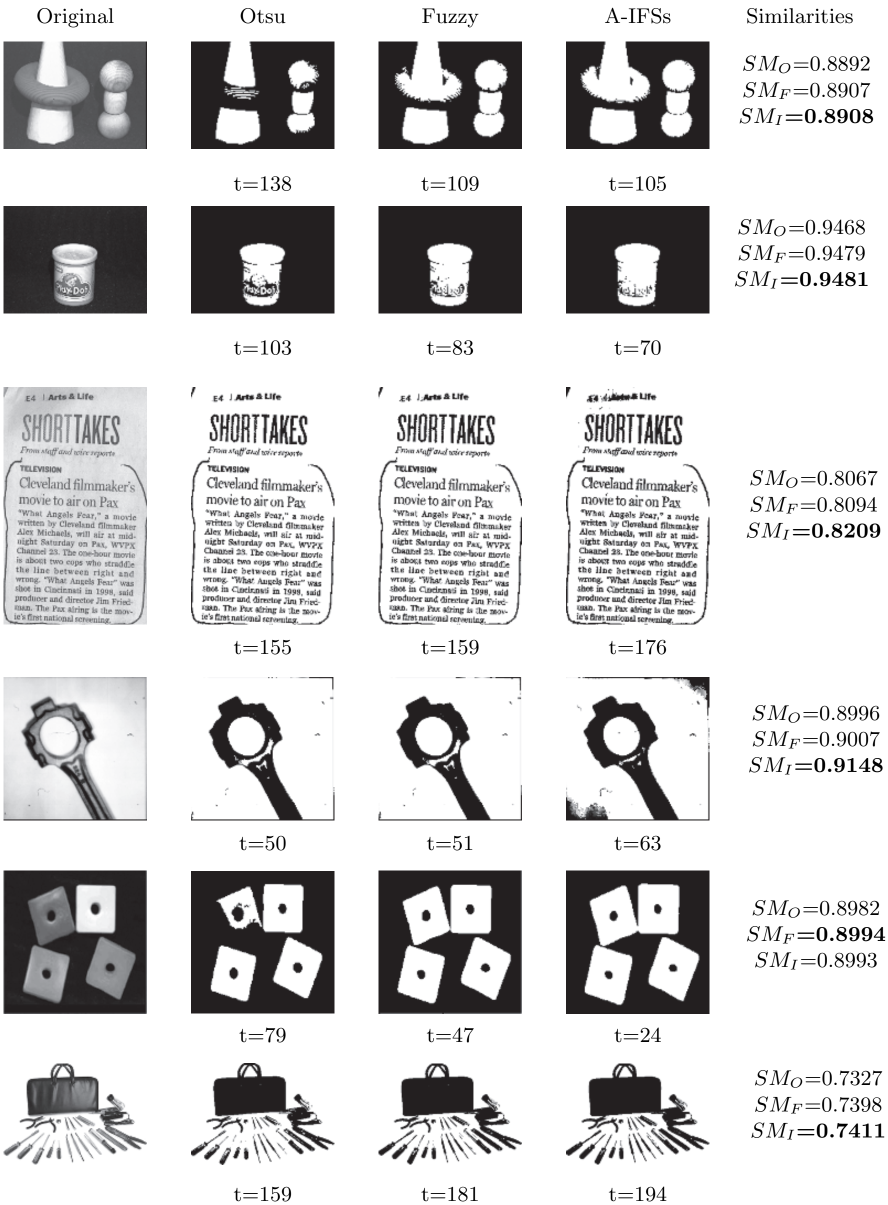



**Fig. 6** Binary images obtained with the Otsu, Fuzzy and A-IFSs algorithms

We would point out that the best threshold selection carried out by Algorithm 3 does not always coincide with the image that visually separates background from the object best. As we know, the selection of best threshold is very subjective and depends in each case on the application we are working on. For example, in medicine the result provided by Algorithm 3 is only indicative and it is the doctor who has to decide, in supervision mode, which is the best threshold. (We have conducted experiments that indicate the criteria used by many doctors provide results that are not very good when the images used are not provided with methods used in medicine).

## 7 Conclusions

We know that the difficulty of the fuzzy techniques for determining the thresholding of an image mainly lies in the determination of the optimal membership function. In this chapter we have presented a method that considers the difficulty encountered by experts for finding the best function. We have quantified the ambiguity in the selection of such functions in the intuitionistic indices of the A-IFS sets that represent the image. This way we have been able to use entropy IE in the same sense as fuzzy entropy is used in the fuzzy algorithms.

Considering the experimental results, we can say that, in general terms, the algorithm proposed in the chapter (Alg. 2) provides us with better results than the fuzzy algorithms. We think there are two reasons why this happens:

1. The use of restricted dissimilarity functions in the constructions of the membership functions. Another advantage of this concept is that, since they can be generated from automorphisms, the program developed enables us to construct different restricted dissimilarity functions for their later execution in an almost immediate way.
2. The representation of the ignorance/unknowledge of the expert by means of Atanassov's intuitionistic index.

With respect to the time efficiency of the A-IFS algorithm we have seen that it is practically the same as that of the fuzzy algorithm. The reason is that, once the fuzzy sets are constructed, in Algorithm 2 we construct the non-membership functions and afterward we calculate the IE by means of sums of differences. In the fuzzy case, once the same membership functions have been constructed, we calculate the fuzzy entropy, which also generally consists of sums of differences.

Finally, we must point out that the method presented for the best threshold selection gives good results under experimental conditions. This method enables us to assure that the thresholds calculated with Atanassov's intuitionistic fuzzy techniques are good. Nevertheless, we have also found in experiments that the threshold selected by Algorithm 3 does not often coincide with the one that would be visually selected by a human. This is due to the fact that the choice of best threshold is very difficult and the method presented in this chapter does not always take into account the particular interests of the field in which it is going to be applied. For this reason, in the near future we must improve our algorithms with heuristic techniques that will enable us to adapt the algorithms to the corresponding application. Nevertheless, we must indicate that in most cases the algorithms developed in these chapters improve the threshold calculated with respect to those calculated with other techniques.

It is important to point out that the results achieved with Algorithm 2 when $Q_{Bt}$ is replaced by $Q_{Ot}$ are the same.

**Acknowledgment** This work is supported by Research Project MTM2005-08982-C04-04.

## References

1. Atanassov K., *Intuitionistic fuzzy sets*, Fuzzy Sets and Systems, Vol.20, pp. 87–96, 1986
2. Atanassov K., *Intuitionistic fuzzy sets. Theory and applications*, Springer-Verlag, Heidelberg, New York, 1999
3. Barrenechea E., *Image Thresholding with Interval-valued Fuzzy Sets. Edge Detection. Contrast*, Ph. D. thesis, Universidad Pública de Navarra, 2005.
4. Bezdek J.C., Keller J., Krisnapuram R. and Pal N. R., *Fuzzy Models and algorithms for pattern recognition and image processing*, in The Handbooks of Fuzzy Sets Series, Series Editors: D. Dubois and H. Prade, Kluwer Academic Publishers, Boston/London/Dordrecht, 1999
5. Burillo P. and Bustince H., *Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets*, Fuzzy Sets and Systems Vol. 78, pp. 81–103, 1996
6. Bustince H., Barrenechea E. and Pagola M., *Restricted equivalence functions*, Fuzzy Sets and Systems Vol. 157, pp. 2333–2346, 2006
7. Bustince H. and Burillo P., *Perturbation of intuitionistic fuzzy relations*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems Vol. 9 (1), pp. 305–316, 2001
8. Bustince H., Kacprzyk J. and Mohedano V., *Intuitionistic fuzzy generators. Application to intuitionistic fuzzy complementation*, Fuzzy Sets and Systems, Vol. 114, pp. 485–504, 2000
9. Bustince H., Mohedano V., Barrenechea E. and Pagola M, *Proximity functions. Application to fuzzy thresholding*, in: Proceedings EUSFLAT 2005
10. Chaira T. and Ray A.K., *Fuzzy measures for color image retrieval*, Fuzzy Sets and Systems, Vol. 150, pp. 545–560, 2005
11. Chi Z., Yan H. and Pham T., *Optimal image thresholding*, Fuzzy algorithms: with application to image processing and pattern recognition, World Scientific, Singapore, pp. 45–84, 1998
12. Cornelis C., Deschrijver G. and Kerre E., *Intuitionistic Fuzzy Connectives Revisited*, in: Proceedings of the Ninth International Conference IPMU 2002, pp. 1839–1844, Annecy-France July 1–5, 2002
13. Forero M.G., *Fuzzy thresholding and histogram analysis*, in Fuzzy Filters for Image Processing, Edited by M. Nachtegael, D. Van der Weken, D. Van de Ville and E.E. Kerre, Springer, pp. 129–152, 2003
14. Forero M.G. and Rojas O., *New formulation in image thresholding using fuzzy logic*, in 11th Portuguese conference on pattern recognition RECPAD2000, pp. 117–124, 2000
15. Forero M.G., Sierra E.L., Alvarez J., Pech J., Cristobal G., Alcalá L. and Desco M., *Automatic sputum color image segmentation for tuberculosis diagnosis*, in Proceedings of SPIE: Algorithms and systems for optical information processing, Edited by V. Javidi Dahram and Psaltis Demetri, Vol. 4471, pp. 251–261, 2001
16. Glasbey C.A., *An analysis of histogram-based thresholding algorithms*, in CVGIP: Graphical models and image processing, 55 (6), pp. 532–537, 1993
17. Huang L.K. and Wang M.J., *Image thresholding by minimizing the measure of fuzziness*, Pattern recognition, Vol. 28(1), pp. 41–51, 1995
18. Jan J.S.R., Sun C.T. and Mizutani E., *Fuzzy Sets*, Neuro-fuzzy and soft computing, pp. 13–46, 1997
19. Lin C.T. and Lee G., *Fuzzy measures*, Neural fuzzy systems: A neuro-fuzzy synergism to intelligent systems, Prentice Hall, Upper Saddle River, pp. 63–88, 1996
20. Otsu N., *A threshold selection method from gray level histograms*, IEEE Transactions on Systems, Man and Cybernetics, Vol. 9, pp. 62–66, 1979

21. Pal N. R. and Pal S. K., *A review of image segmentation techniques*, Pattern recognition, Vol. 26, 1277–1294, 1993
22. Parker J.R., *Advanced method in grey-level segmentation*, Algorithms for image processing and computer vision, John Wiley and Sons, New York pp. 116–149, 1997
23. Pratt W.K., *Image segmentation*, Digital image processing, John Wiley and Sons, New York, pp. 597–627, 1991
24. Sankur B. and Sezgin M., *Survey over image thresholding techniques and quantitative perfomance evaluation*, J. Electron. Imaging, Vol. 13 (1), pp. 146–165, 2004
25. Sahoo P.K., Soltani S., Wong A.K.C. and Chen Y.C., *A survey of thresholding techniques*, Computer vision, graphics and image processing, Vol. 41, pp. 233–260, 1988
26. Szmidt E. and Kacprzyk J., *Entropy for intuitionistic fuzzy sets*, Fuzzy Sets and Systems, Vol. 118(3), 467–477, 2001
27. Tizhoosh H. R., *Image thresholding using type II fuzzy sets*, Pattern Recognition, Available on line, 2005. In Press
28. Yager R. R., *On the measure of fuzziness and negation. Part I: Membership in the unit interval*, Intern. J. of General Systems, Vol. 5, pp. 221–229, 1979
29. Yager R. R., *On the measure of fuzziness and negation. Part II Lattices*, Information and Control, Vol. 44(3), pp. 236–260, 1979
30. Zadeh L.A., *Fuzzy sets*, Information Control, Vol. 8, pp. 338–353, 1965

# Colour Image Comparison Using Vector Operators

**Dietrich Van der Weken, Valérie De Witte, Mike Nachtegael, Stefan Schulte and Etienne Kerre**

**Abstract** Objective quality measures or measures of comparison are of great importance in the field of image processing. These measures serve as a tool to evaluate and to compare different algorithms designed to solve particular problems, such as noise reduction, deblurring, compression, ... Consequently these measures serve as a basis on which one algorithm is preferred to another. In [15, 16] we constructed several new fuzzy similarity measures for grey-scale images that outperform the classical measures of comparison, like Root Mean Square Error or Peak Signal to Noise Ratio, in the sense of image quality evaluation. In this chapter we investigate the usefulness of these similarity measures for the comparison of colour images. First of all, we discuss the component-based approach in three different colour spaces, namely the RGB, HSV and Lab colour spaces. And secondly, we discuss a vector-based approach using vector morphological operators. Both approaches are compared by means of several experiments.

## 1 Basic Notions

### *1.1 Fuzzy Sets and Digital Images*

First of all it is important to observe that digital images and fuzzy sets are modelled in the same way. A fuzzy set $A$ in a universe $X$ is characterised by a $X - [0, 1]$ mapping $\chi_A$, which associates with every element $x$ in $X$ a degree of membership $\chi_A(x)$ of $x$ in the fuzzy set $A$. In the following, we will denote the degree of membership by $A(x)$. A digital image can be identified with a fuzzy set that takes its values on the grid points $(i, j)$, with $i, j \in \mathbb{N}, 0 \leq i \leq M$ and $0 \leq j \leq N$ $(M, N \in \mathbb{N})$. Given a digital image $A$, the quantity $A(i, j) \in [0, 1]$ (after rescaling) then represents the greyvalue of the image $A$ in the grid point $(i, j)$.

### *1.2 Similarity Measures*

In the literature a lot of measures are proposed to express the similarity or equality between fuzzy sets. There is no unique definition, but the most frequently used is the

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.* 

following. A similarity measure [21] is a fuzzy binary relation in $\mathcal{F}(X)$, i.e. a $\mathcal{F}(X) \times \mathcal{F}(X) \to [0, 1]$ mapping that is reflexive, symmetric and min-transitive. However, not every measure in the literature satisfies this definition. Therefore, a similarity measure will here be understood as a measure we can use to compare fuzzy sets, or objects which can be identified with fuzzy sets. We refer to the following section for some examples.

## *1.3 Colour Spaces*

A colour space is a way of representing colours and their relationship to each other. Different image processing systems use different colour models for different reasons. In this paper we will consider three commonly used colour spaces, namely the RGB colour space, the HSV colour space and the Lab colour space. For more details on these colour spaces we refer to [11, 12].

### RGB Colour Space

This technical colour model is very popular and is mostly used in digital image processing. The model is based on the trichromatic theory where colours are produced by a composition of the three basic colours: red, green and blue. These basic colours are called primaries. The RGB colour model is an additive colour display system, where the origin is black and the different colours are obtained by adding different amounts of the three basic colours. The grey-scales follow the line from black to white.

The RGB model simplifies the design of computer graphics systems but is not ideal for all applications. The red, green and blue colour components are highly correlated and therefore it is difficult to execute some image processing algorithms. Many processing techniques, such as histogram equalization, operate on the intensity component of an image only.

### HSV Colour Space

Sometimes the properties hue, saturation and value are used to describe colour. Therefore it seems logical that there is a corresponding colour model, namely the HSV colour space. In this case we do not see a certain colour as amounts of primary colours mixed in certain proportions. Hue is the colour as described by wavelength, for instance the distinction between red and yellow. Saturation is the amount of the colour that is present. The saturation describes how much a certain colour differs from white light, for instance the distinction between red (highly saturated) and pink (little saturated). The value is the amount of light, the distinction between a dark red and light red or between dark grey and light grey. So we don't need to know what percentage of blue or green is present in order to produce a colour. We can simply adjust the hue to get the colour we wish. To change the amount of white light, we adjust the saturation and to make the colour darker or lighter we alter the value.

The HSV colour space can be modelled with cylindrical coordinates. The hue is represented as the angle, varying from $0^\circ$ to $360^\circ$. Saturation corresponds to the radius, varying from 0 to 1. The value varies along the $z$ axis with 0 being black and 1 being white.

### Lab Colour Space

The RGB model has a technical origin and has a few drawbacks. Similar colours are 3-dimensional subspaces of the colour cube and are therefore difficult to represent. The borders of these subspaces are not straight lines, and consequently hard to define. Therefore the *Commission Internationale de l'Eclairage* (CIE) defined another colour space in 1931. The Lab colour model was developed to be completely independent of any device or other means of emission or reproduction and is based as closely as possible on how humans perceive colour. Equal Euclidean distances in these colour space correspond to roughly equal perceived colour differences. We refer to [12] for more detailed information about this colour space.

## 2 Overview of the Similarity Measures for Greyscale Images

### *2.1 Direct Application of Similarity Measures to Digital Images*

In the literature a lot of measures are proposed to express the similarity or equality between two fuzzy sets [2, 3, 4, 7, 9, 10, 21, 22, 23, 24]. From a total of more than 40 different similarity measures, 14 similarity measures satisfy a list of relevant properties for image processing [15]. Six of these measures were already discussed in [13], but meanwhile 8 other similarity measures turned out to be appropriate for the comparison of images. We will not discuss the whole list of similarity measures, but as example we will shortly repeat the expressions of three similarity measures.

The first similarity measure is based on the fuzzy Minkowski distance , and the observation that the smaller the distance between $A$ and $B$, the greater the similarity between $A$ and $B$. This observation leads to the following two similarity measures $M_1$ [3] and $M_2$ [10] (for $r \to \infty$):

$$M_1(A, B) = 1 - \left( \frac{1}{MN} \sum_{(x,y)\in X} |A(x, y) - B(x, y)|^r \right)^{\frac{1}{r}}, r \in \mathbb{N}\backslash\{0\} \quad (1)$$

$$M_2(A, B) = 1 - \max_{(x,y)\in X} |A(x, y) - B(x, y)|. \quad (2)$$

For the next similarity measure we also need the notion of cardinality of a fuzzy set. The cardinality of a finite crisp set is given by the number of elements in that set. This concept can be extended to fuzzy sets using the sigma count . The sigma count of a fuzzy set $A$ (with finite support) in a universe $X$ is defined as:

$$|A| = \sum_{x \in X} A(x). \tag{3}$$

The similarity measure $M_3$ [4] is based on this notion of cardinality:

$$\begin{aligned} M_3(A, B) &= \frac{|A \cap B|}{|A \cup B|} \\ &= \frac{\sum\limits_{(x,y) \in X} \min(A(x, y), B(x, y))}{\sum\limits_{(x,y) \in X} \max(A(x, y), B(x, y))}. \end{aligned} \tag{4}$$

Note that we use the minimum to model the intersection of two fuzzy sets and the maximum to model the union of two fuzzy sets.

## 2.2 Construction of Neighbourhood-based Similarity Measures

In order to improve the perceptual character of the similarity measures we have incorporated an important property of the human visual system, namely the contrast or homogeneity sensitivity [20]. Therefore, we review shortly how to construct these neighbourhood-based similarity measures that also incorporate homogeneity in the different neighbourhoods. We start with calculating the similarity between disjoint image parts of two images $A$ and $B$. Therefore we divide both images $A$ and $B$ in disjoint $8 \times 8$ image parts and we calculate the similarity between each of the $8 \times 8$ image parts (see Fig. 1). To calculate the similarity between two image parts we simply apply the above pixel-based similarity measures to both image parts. So, suppose the image is divided in $N$ image parts of size $8 \times 8$, and the similarity between the image part $A_i$ of image $A$ and the image part $B_i$ of image $B$ is denoted by $M(A_i, B_i)$, then the similarity between the two images $A$ and $B$ is given by the weighted average of the similarities in the corresponding disjoint image parts. So, we have that

$$M^h(A, B) = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot M(A_i, B_i), \tag{5}$$

where the similarity $M(A_i, B_i)$ is calculated using a pixel-based similarity measure, discussed in subsection 3.1, restricted to the image parts $A_i$ and $B_i$ and the weight $w_i$ is defined as the similarity between the homogeneity $h_{A_i}$ of image part $i$ in image $A$ and the homogeneity $h_{B_i}$ of image part $i$ in image $B$. The homogeneity $h_{A_i}$ of an image part $i$ in image $A$ is computed as the similarity between the grey value of

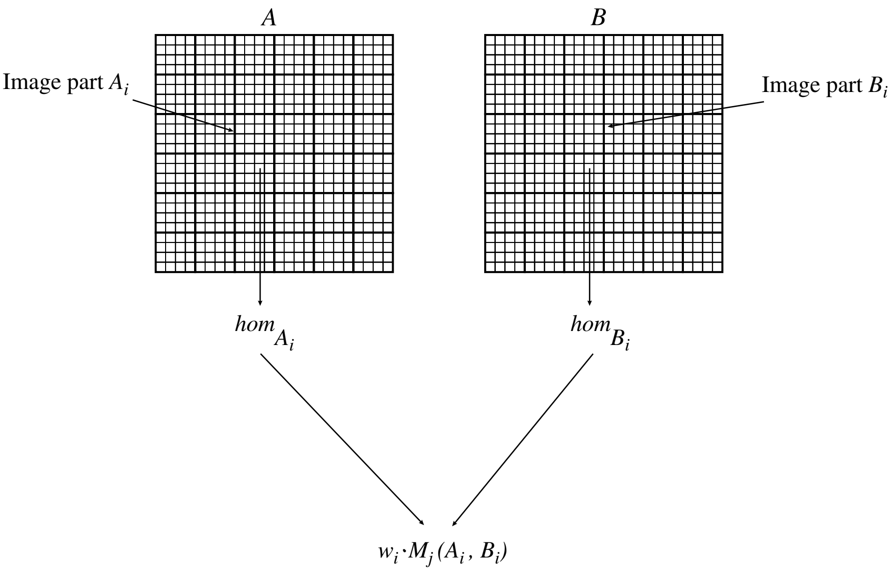



**Fig. 1** For the calculation of the global similarity value between two images $A$ and $B$ we consider a weighted average with weights $w_i$ which are defined as the similarity between the homogeneity $hom_{A_i}$ of the image part $A_i$ in image $A$ and the homogeneity $hom_{B_i}$ of the image part $B_i$ in image $B$

the pixel in the image part with maximum intensity and the grey value of the pixel in the image part with minimum intensity (see Fig. 2), using a resemblance relation [5, 15]. In order to decide which resemblance relation we will use, we make use of the following property [5]:

**Preposition** *For every pseudo-metric d on X, for a in* $[1, +\infty[$*, b in* $]0, +\infty[$ *we have that: E defined as*

$$E(x, y) = \min(1, \max(0, a - b \cdot d(x, y))) \tag{6}$$

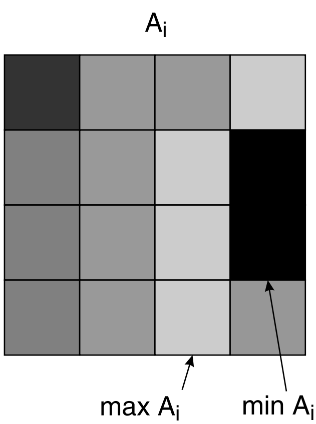



**Fig. 2** The homogeneity in an image part $A_i$ is defined as the similarity between the grey value of the pixel with minimum intensity (min $A_i$) in this image part and the grey value of the pixel with maximum intensity (max $A_i$) in this image part

*for every x and y in X, is a* [0, 1]*-valued resemblance relation on X w.r.t. d. Furthermore we have that*

$$\begin{aligned} E(x, y) = 1 \text{ if and only if } d(x, y) \leq \tfrac{a-1}{b} \\ E(x, y) = 0 \text{ if and only if } d(x, y) \geq \tfrac{a}{b}. \end{aligned} \tag{7}$$

Since every subset $G$ of $\mathbb{R}$ can be transformed to a metric space by using the absolute value metric as distance function and since normalized grey values and consequently the homogeneity in a certain image part belong to the unit interval, we make use of the absolute value metric to construct the resemblance relation $s$. Furthermore we prefer that

$$\begin{aligned} s(x, y) = 1 \text{ if and only if } d(x, y) \leq 0.1 \\ s(x, y) = 0 \text{ if and only if } d(x, y) \geq 0.6. \end{aligned} \tag{8}$$

Using the above property we obtain then the following values for $a$ and $b$:

$$a = \frac{6}{5} \text{ and } b = 2, \tag{9}$$

such that we obtain the following expression for the resemblance relation $s$ (see Fig. 3) which will be used for the calculation of the similarity between two values:

$$s(x, y) = \min(1, \max(0, \frac{6}{5} - 2|x - y|)). \tag{10}$$

So we have that

$$h_{A_i} = s(\max_{(i,j)\in A_i} A(i, j), \min_{(i,j)\in A_i} A(i, j)) \tag{11}$$

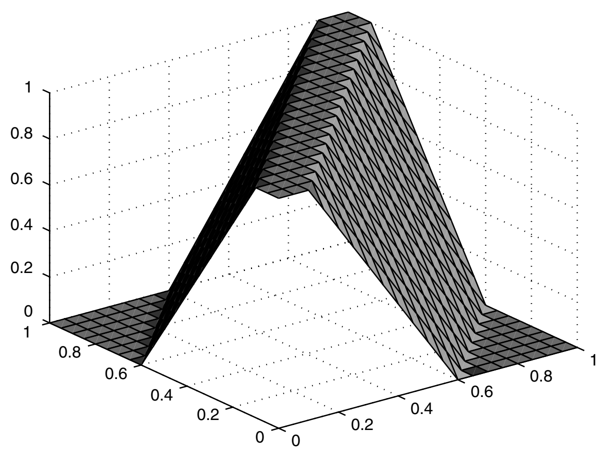



**Fig. 3** The resemblance relation $s$ with $a = \frac{6}{5}$ and $b = 2$

and the weight $w_i$ is then defined as follows:

$$w_i = s(h_{A_i}, h_{B_i}). \tag{12}$$

Using a similarity measure $M_i$ from subsection 3.1 to calculate the similarity between the image parts, we obtain a neighbourhood-based similarity measure. These neighbourhood-based similarity measures will be denoted as $M_i^h$.

## *2.3 Application of Similarity Measures to Histograms of Digital Images*

In this section we illustrate how fuzzy similarity measures can be useful for the comparison of histograms of digital images. We will show that similarity measures can be applied to two different types of histograms: normalized histograms and ordered normalized histograms.

### Direct Application to Histograms

First of all, it is meaningful to compare two histograms in the framework of fuzzy set theory, because the histogram of an image can be transformed into a fuzzy set in the universe of grey levels by dividing the values of the histogram in every grey level by the maximum amount of pixels with the same grey value. In this way the most typical grey value gets membership degree 1 in the fuzzy set associated with the histogram and every other less typical grey value gets a smaller membership degree. Consequently, a normalized histogram is in accordance with the intuitive idea behind a fuzzy set: the most typical element in the universe gets membership degree 1 and all other less typical elements belong to the fuzzy set to a less extent which can be expressed by membership degrees smaller than 1. In this way we obtain the following expression for the membership degree of the grey value $g$ in de fuzzy set $Fh_A$ associated with the histogram $h_A$ of the image $A$:

$$Fh_A(g) = \frac{h_A(g)}{h_A(g_M)} \tag{13}$$

with $G$ the universe of grey levels and $h_A(g_M) = \max_{g \in G} h_A(g)$. As histograms of digital images can be identified with fuzzy sets in the universe of grey values, it is interesting to investigate whether similarity measures, originally introduced to express the degree of comparison between two fuzzy sets, can be applied to normalized histograms in a meaningful way. In this way we compare two images on a histogram-level, and the frequencies of the different grey values are mutually compared.

### Application to Ordered Histograms

Similarity measures can be applied in a second way to associated histograms of digital images. The values of a histogram can be ordered in such a way the least occurring grey value is placed in the first position of the histogram and the remaining frequencies are ordered in increasing order. Again, the histogram is normalized analogously to the first case, and consequently the most typical grey value gets membership degree 1 in the fuzzy set associated with the histogram and all the other membership degrees are smaller than or equal to 1 and are ordered in increasing order. Again, we can apply the different similarity measures to these ordered and normalized histograms. In contrast to the first application of similarity measures to normalized histograms, where the frequencies of the different grey levels are mutually compared, in this case the frequency of the most occurring grey level in the image $A$ is compared with the frequency of the most occurring grey level in the image $B$, the frequency of the second most occurring grey level in the image $A$ is compared with the second most occurring grey level in the image $B$, ... So, the frequencies of the different grey values are compared frequency by frequency, with respect to an increasing order of the different frequencies. So, it can happen that two frequencies of two different grey values are compared to each other, depending on the place they take in the ordered histogram. If the ordered histogram of an image $A$ is denoted as $o_A$, we obtain the following expression for the fuzzy set associated with the ordered histogram of the image $A$. For $i = 1, ..., |G|$, with $G$ the universe of grey levels:

$$Oh_A(i) = \frac{o_A(i)}{o_A(|G|)}, \tag{14}$$

with $o_A(|G|) = \max_{g \in G} h_A(g)$.

### Appropriate Measures of Comparison for Histograms

A profound experimental study of the applicability of similarity measures to normalized histograms resulted in 15 similarity measures [14, 17] which are appropriate for histogram comparison, i.e. they satisfy the list of relevant properties we impose to a similarity measure in order to be applicable in image processing. As example, we repeat the expression of two appropriate similarity measures:

$$\begin{aligned} H_3(A, B) &= \frac{|Fh_A \cap Fh_B|}{|Fh_A \cup Fh_B|} \\ &= \frac{\sum_{g \in G} \min(Fh_A(g), Fh_B(g))}{\sum_{g \in G} \max(Fh_A(g), Fh_B(g))}, \end{aligned} \tag{15}$$

$$H_4(A,B) = \frac{|(Fh_A \triangle Fh_B)^c|}{\max(|(Fh_B \backslash Fh_A)^c|, |(Fh_A \backslash Fh_B)^c|)}$$
$$= \frac{\sum_{g\in G} 1 - \max\begin{pmatrix}\min(Fh_A(g), 1 - Fh_B(g)),\\ \min(Fh_B(g), 1 - Fh_A(g))\end{pmatrix}}{\max(\begin{pmatrix}\sum_{g\in G} 1 - \min(Fh_B(g), 1 - Fh_A(g)),\\ \sum_{g\in G} 1 - \min(Fh_A(g), 1 - Fh_B(g))\end{pmatrix}} \tag{16}$$

with $G$ the universe of grey levels.

If the similarity measures are applied to ordered histograms we obtained 22 similarity measures which satisfy the list of relevant properties [17]. Besides the 15 similarity measures which were appropriate for direct application, we found 7 extra similarity measures which are appropriate for application to ordered histograms. Also in this case we only recollect the expression of two appropriate similarity measures:

$$OH_3(A,B) = \frac{|Oh_A \cap Oh_B|}{|Oh_A \cup Oh_B|}$$
$$= \frac{\sum_{i=1}^{|G|} \min(Oh_A(i), Oh_B(i))}{\sum_{i=1}^{|G|} \max(Oh_A(i), Oh_B(i))} \tag{17}$$

$$OH_5(A,B) = \frac{\min(|Oh_A \backslash Oh_B|, |Oh_B \backslash Oh_A|)}{\max(|Oh_A \backslash Oh_B|, |Oh_B \backslash Oh_A|)}$$
$$= \frac{\min\begin{pmatrix}\sum_{i=1}^{|G|} \min(Oh_A(i), 1 - Oh_B(i)),\\ \sum_{i=1}^{|G|} \min(Oh_B(i), 1 - Oh_A(i))\end{pmatrix}}{\max\begin{pmatrix}\sum_{i=1}^{|G|} \min(Oh_A(i), 1 - Oh_B(i)),\\ \sum_{i=1}^{|G|} \min(Oh_B(i), 1 - Oh_A(i))\end{pmatrix}} \tag{18}$$

### Application to Local Histograms

Applying the similarity measures to partitioned images resulted in neighbourhood-based similarity measures with a higher perceptual performance. Therefore, it is also very interesting to investigate whether the performance of the histogram similarity measures can be improved by applying them to larger image parts. In this respect we can investigate whether the similarity measures can be applied to local histograms.

Therefore, we again divide both images $A$ and $B$ in disjoint $8 \times 8$ image parts and we determine the histogram for each of the image parts. The similarity between the images $A$ and $B$ is then calculated by applying a histogram similarity measure $H_j$ (or $OH_j$) to each of the pairs $(h_{A_i}, h_{B_i})$, with $h_{A_i}$ and $h_{B_i}$ the local normalized histograms of respectively image part $A_i$ in the image $A$ and image part $B_i$ in the image $B$ (or to each of the pairs $(o_{A_i}, o_{B_i})$ with $o_{A_i}$ and $o_{B_i}$ the local normalized ordered histograms of respectively image part $A_i$ in the image $A$ and image part $B_i$ in the image $B$).

## 2.4 Combined Similarity Measures

In order to optimally incorporate the image characteristics we proposed combined similarity measures which consist of a combination of neighbourhood-based similarity measures and similarity measures which are applied to ordered histograms (see Fig. 4). The combined image quality measures [16] are given by

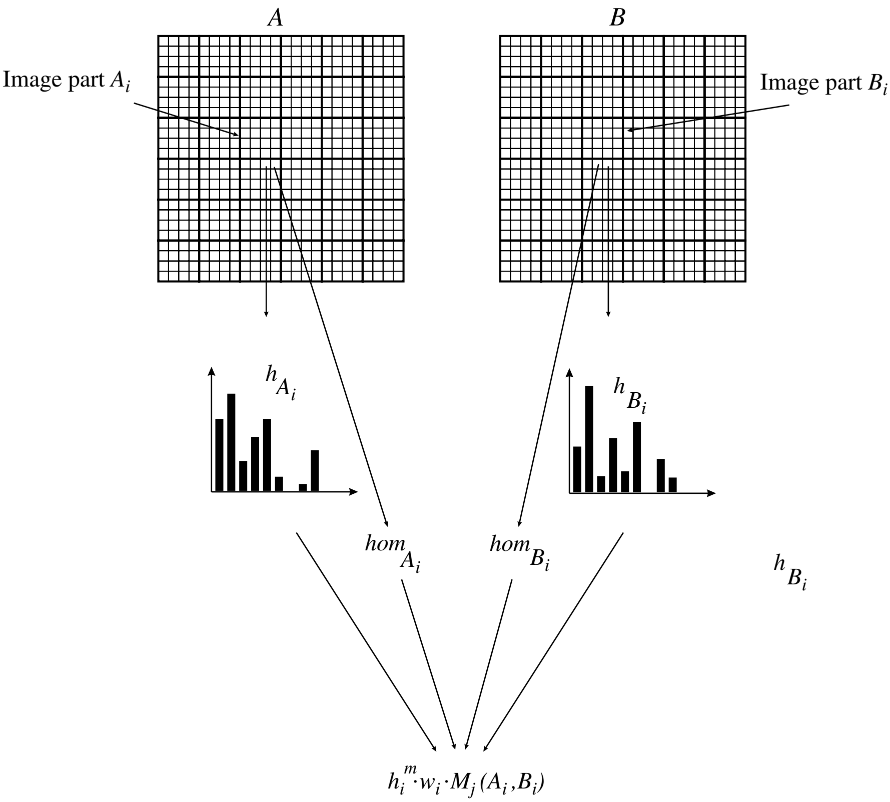



**Fig. 4** A diagram of the combined method which uses local histograms and neighbourhood-based similarity measures which also incorporate homogeneity for the calculation of the similarity value between two digital images $A$ and $B$

the following weighted average:

$$Q_{m,n}(A, B) = \frac{1}{N} \sum_{i=1}^{N} h_i^m \cdot w_i \cdot M_n(A_i, B_i), \tag{19}$$

where $h_i^m$ is calculated by applying a similarity measure $OH_m$ to the local ordered histograms of the image parts $A_i$ and $B_i$.

# 3 Applicability of the Similarity Measures to Colour Images

If we want to apply similarity measures to colour images, a problem arises because in case of colour images the value of a pixel is not a single value but a colour vector consisting of three values. For greyscale images we can calculate the minimum, maximum, sum or difference between two grey values of two pixels, but this is no longer straightforward in case of colour pixels. So we need a mechanism to extend the different similarity measures to colour images. A straightforward extension is based on the application of the different similarity measures to each component of the colour vectors. Besides this approach, we also discuss a vector-based approach [19] in this paper.

## *3.1 Component-based Approach*

A straightforward extension to colour images is applying the different similarity measures to each component of the colour vectors and then merge the result of the different components in a suitable way. We will consider the arithmetic mean for merging the results in the different components. Note that there exist other aggregation operators [1] that could be useful for this merging. So, for example in the RGB colour space, we have that

$$Q_{i,j}^{RGB}(A, B) = \frac{Q_{i,j}^{R}(A, B) + Q_{i,j}^{G}(A, B) + Q_{i,j}^{B}(A, B)}{3} \tag{20}$$

This approach can be applied in a similar way in the HSV colour space and the Lab colour space.

## *3.2 Vector-based Approach*

### Vector Operators for Colour Images

In order to define the minimum and maximum between two colours we need an ordering in the different colour spaces. Of course the ordering can be dependent on the choice of the colour space. We start with the RGB colour space. In this case we use the new ordering $\leq_{RGB}$ of vectors introduced in [6]. If we consider two colours $c(r_c, g_c, b_c)$ and $c'(r_{c'}, g_{c'}, b_{c'})$ in the RGB cube, we have that:

$$\begin{aligned}
& c <_{RGB} c' \\
& \Leftrightarrow d(c, Bl) < d(c', Bl) \\
& \quad \text{or} \begin{pmatrix} (d(c, Bl) = d(c', Bl)) \\ \text{and } (d(c, Wh) > d(c', Wh)) \end{pmatrix} \\
& c >_{RGB} c' \\
& \Leftrightarrow d(c, Wh) < d(c', Wh) \\
& \quad \text{or} \begin{pmatrix} (d(c, Wh) = d(c', Wh)) \\ \text{and } (d(c, Bl) > d(c', Bl)) \end{pmatrix} \\
& c =_{RGB} c' \\
& \Leftrightarrow d(c, Bl) = d(c', Bl) \\
& \quad \text{and } d(c, Wh) = d(c', Wh)
\end{aligned} \tag{21}$$

with $Bl(0, 0, 0)$ and $Wh(1, 1, 1)$ representing black and white, and $d$ the Euclidean distance.

In the HSV colour space we use the ordering $\leq_{HSV}$ introduced in [8]. For two colours $c(h_c, s_c, v_c)$ and $c'(h'_c, s'_c, v'_c)$ in the HSV colour space, this ordering is defined as follows:

$$\begin{aligned}
& c <_{HSV} c' \\
& \Leftrightarrow v_c < v'_c \\
& \quad \text{or } (v_c = v'_c) \text{ and } (s_c > s'_c) \\
& \quad \text{or } (v_c = v'_c) \text{ and } (s_c = s'_c) \text{ and } (h_c < h'_c) \\
& c =_{HSV} c' \\
& \Leftrightarrow (v_c = v'_c) \text{ and } (s_c = s'_c) \text{ and } (h_c = h'_c)
\end{aligned} \tag{22}$$

In the Lab colour space we can use a similar ordering as in the RGB colour space, which means that we use the distance to black and white to define the ordering.

Besides the minimum and maximum between two colours, we also need to extend the notion of cardinality. A colour image can be considered as a $L$-fuzzy set, with $L$ the RGB cube (resp. the HSV colour space, or Lab colour space) or, after normalisation, the unit cube $[0, 1]^3$. So, in this case we can't sum up membership degrees because the membership degree of a certain pixel in a colour image is a triplet $(c_x, c_y, c_z)$. Therefore we replace the colour values $c(c_x, c_y, c_z)$ by the norm (or distance to black) of the colour, and we use these values to define the sigma count. So for a colour image $C$ we have that

$$|C| = \sum_{c \in X} \sqrt{c_x^2 + c_y^2 + c_z^2}, \tag{23}$$

with $c_x$, $c_y$ and $c_z$ the colour values of the pixel $c$.

### Local Colour Histograms

We also have to pay attention to the histogram of a colour image. The histogram of a greyscale image is a chart that shows the frequency distribution of the different grey levels. The value of the histogram of a greyscale image $A$ in a grey level $g$ equals the total amount of pixels in the image $A$ with grey level $g$. If we consider the histogram of a colour image in the same way, we obtain a histogram, in case of 8-bit images, of dimension $2^{8^3}$. It is obvious that the dimension and variance of such histograms is far too large. If we however consider local ordered histograms, for example histograms of $8 \times 8$ image parts, the dimension of the histogram will decrease significantly. So, if we consider two colour images $A$ and $B$, we divide both images by $N$ $8 \times 8$ image parts $A_i$ and $B_i$. In this way, the local histograms of the image parts $A_i$ and $B_i$ will consist of at most 64 different colours. We now sort for each image part the frequencies of the different colours in descending order such that the most typical colour takes the first place in the colour histogram of the image part. The ordered histogram of an image part $A_i$ will be denoted as $o_{A_i}$ and can be transformed into a fuzzy set by dividing the values of the histogram by the maximum number of pixels with the same colour (the colour at position 1 in the ordered histogram). In this way the most typical colour has membership degree 1 in the fuzzy set associated with the ordered histogram and we obtain the following expression for the membership degree of the colour on the $i$-th position in the fuzzy set $Oh_{A_i}$ associated with the ordered histogram $o_{A_i}$ of the image part $A_i$:

$$Oh_{A_i}(c) = \frac{o_{A_i}(c)}{o_{A_i}(1)}, \tag{24}$$

where $o_{A_i}(1) = \max_c h_{A_i}(c)$.

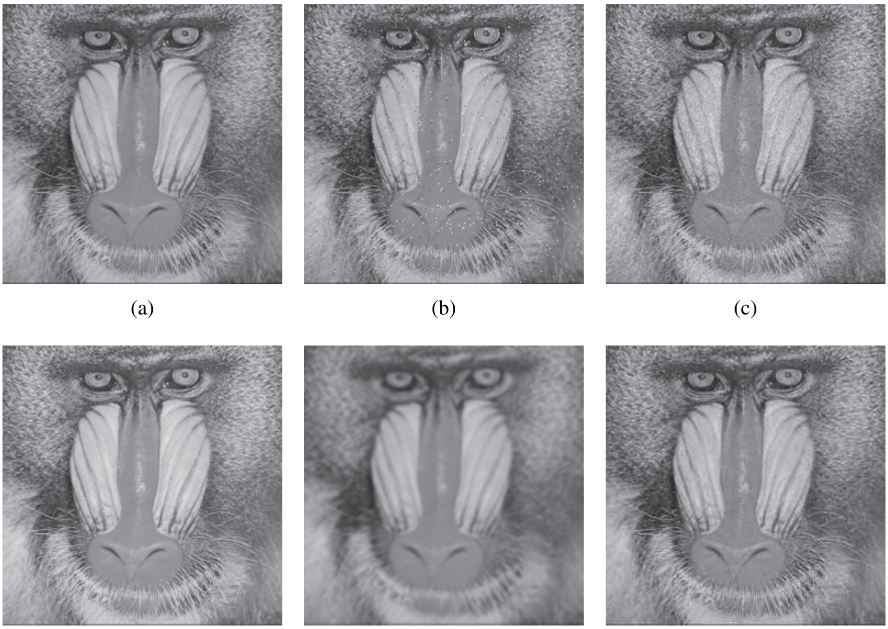
## 4 Experiments

In this Section we will illustrate the applicability of the similarity measures for the comparison of colour images. We compare the performance of the vector-based approach with the component-based approach from several experiments where we consider a certain image to which we add a variety of corruptions: impulsive salt & pepper noise, additive gaussian noise, enlightening, blur and JPEG compression. Furthermore, we tuned all the corruptions to yield more or less the same $RMSE$ relative to the original image. As example, we choose the following two similarity measures:

- $Q_{3,3}(A, B) = \frac{1}{N} \sum_{i=1}^{N} h_i^3 \cdot w_i \cdot M_3(A_i, B_i)$, where $h_i^3$ is calculated by applying the similarity measure $OH_3$ to the local ordered histograms of the image parts $A_i$ and $B_i$;

- $Q_{5,1}(A,B) = \frac{1}{N}\sum_{i=1}^{N} h_i^5 \cdot w_i \cdot M_1(A_i, B_i)$, where $h_i^5$ is calculated by applying the similarity measure $OH_5$ to the local ordered histograms of the image parts $A_i$ and $B_i$.

The different similarity measures are applied to the different versions of the "monkey" image (see Fig. 5). The numerical results of this experiment are displayed in Table 1 and Table 2. From these results we can observe that, especially in the RGB and Lab colour space, the values of the vector-based measures are significantly higher than the values of the component-based measures (except for the comparison of the original image with the blurred version of it). This is also in accordance with the visual quality of the different images. Especially when we compare these results with the experiments on greyscale images (see [15]). The results of the component-based measures applied to the colour versions of the "monkey" image are comparable with the results in case of greyscale images. This is of course not suprisingly, since in this case a colour image is considered as three separate greyscale images. However, the different distortions seem to have less influence on colour images, and this observation is confirmed by the numerical results. The results in the HSV colour space are less convincing. This can possibly be explained by the different ranges of the different components and the lack of uniformity in this colour space.



(a) (b) (c) (d) (e) (f)

**Fig. 5** (**a**) Original "monkey" image; (**b**) salt & pepper noise; (**c**) gaussian noise; (**d**) enlightening; (**e**) blur; (**f**) JPEG compression

**Table 1** Results of the similarity measures $Q^c_{3,3}$ (component-based approach) and $Q^v_{3,3}$ (vector-based approach) if they are applied to the different greyscale and colour versions of the "monkey" image

| "monkey" image | | | | | |
|---|---|---|---|---|---|
| greyscale version of the "monkey" image | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $RMSE$ | 15.44 | 15.20 | 15 | 15.43 | 15.24 |
| $Q_{3,3}$ | 0.636 | 0.47125 | 0.8903 | 0.30274 | 0.54401 |
| RGB model | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $RMSE$ | 27.40 | 27.75 | 27.58 | 28.22 | 28.71 |
| $Q^c_{3,3}$ | 0.68146 | 0.46831 | 0.84274 | 0.29957 | 0.56389 |
| $Q^v_{3,3}$ | 0.81464 | 0.62370 | 0.90783 | 0.28802 | 0.67571 |
| HSV model | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $Q^c_{3,3}$ | 0.62613 | 0.48000 | 0.88542 | 0.45296 | 0.54730 |
| $Q^v_{3,3}$ | 0.70394 | 0.45466 | 0.84532 | 0.31426 | 0.42458 |
| Lab model | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $Q^c_{3,3}$ | 0.57223 | 0.47646 | 0.81492 | 0.46401 | 0.47174 |
| $Q^v_{3,3}$ | 0.67102 | 0.54879 | 0.81996 | 0.48405 | 0.50101 |

**Table 2** Results of the similarity measures $Q^c_{5,1}$ (component-based approach) and $Q^v_{5,1}$ (vector-based approach) if they are applied to the different greyscale and colour versions of the "monkey" image

| "monkey" image | | | | | |
|---|---|---|---|---|---|
| greyscale version of the "monkey" image | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $RMSE$ | 15.44 | 15.20 | 15 | 15.43 | 15.24 |
| $Q_{5,1}$ | 0.62891 | 0.50782 | 0.94118 | 0.30333 | 0.57409 |
| RGB model | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $RMSE$ | 27.40 | 27.75 | 27.58 | 28.22 | 28.71 |
| $Q^c_{5,1}$ | 0.67363 | 0.51253 | 0.90234 | 0.30890 | 0.60256 |
| $Q^v_{5,1}$ | 0.74704 | 0.57669 | 0.95844 | 0.25829 | 0.61223 |
| HSV model | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $Q^c_{5,1}$ | 0.53301 | 0.32498 | 0.72638 | 0.37757 | 0.45536 |
| $Q^v_{5,1}$ | 0.70341 | 0.55238 | 0.82729 | 0.32624 | 0.43872 |
| Lab model | | | | | |
| | (a) vs. (b) | (a) vs. (c) | (a) vs. (d) | (a) vs. (e) | (a) vs. (f) |
| $Q^c_{5,1}$ | 0.57667 | 0.46822 | 0.83610 | 0.39587 | 0.40732 |
| $Q^v_{5,1}$ | 0.68788 | 0.52680 | 0.92844 | 0.42721 | 0.42819 |

**Table 3** Results of the ordering according to image quality for the measures $Q^c_{3,3}$ and $Q^v_{3,3}$ in the three colour models

| RGB | | HSV | | Lab | |
|---|---|---|---|---|---|
| $Q^c_{3,3}$ | $Q^v_{3,3}$ | $Q^c_{3,3}$ | $Q^v_{3,3}$ | $Q^c_{3,3}$ | $Q^v_{3,3}$ |
| d | d | d | d | d | d |
| b | b | b | b | b | b |
| f | f | f | c | c | c |
| c | c | c | f | f | f |
| e | e | e | e | e | e |

We can also order the different images according to image quality; the results in the three colour models are summarized in Table 3 and Table 4. From these results we can observe that for the measures $Q^c_{3,3}$ and $Q^c_{5,1}$ the results are dependent on the choice of the colour space. We obtain the same ordering for the RGB colour space and the HSV colour space, but a different ordering for the Lab colour space. For the measure $Q^v_{3,3}$ we obtain the same ordering for the HSV colour space and the Lab colour space, but a different ordering for the RGB colour space. Also for the measures $Q^c_{5,1}$ and $Q^v_{5,1}$ the orderings are dependent on the choice of the colour space. It is of course very hard and possibly subjective to decide which ordering leans closest against the visual quality of the different images.

Finally, we applied the similarity measures to similar distorted versions of the other images. From the numerical results of these experiments we can make similar conclusions as from the results of the "monkey" experiment.

**Table 4** Results of the ordering according to image quality for the measures $Q^c_{5,1}$ and $Q^v_{5,1}$ in the three colour models

| RGB | | HSV | | Lab | |
|---|---|---|---|---|---|
| $Q^c_{5,1}$ | $Q^v_{5,1}$ | $Q^c_{5,1}$ | $Q^v_{5,1}$ | $Q^c_{5,1}$ | $Q^v_{5,1}$ |
| d | d | d | d | d | d |
| b | b | b | b | b | b |
| f | f | f | c | c | c |
| c | c | e | f | f | f |
| e | c | c | e | e | e |

## 5 Conclusion

In this paper we have investigated whether similarity measures, which are designed for the comparison of greyscale images, can also be applied to colour images. Since the minimum and maximum are indispensable operators for several similarity measures, we needed a new definition for the minimum and maximum between two colours. We used a vector ordering in three different colours spaces (RGB, HSV and Lab) to define the minimum and maximum between two colours. In contrast to a component-based approach, a vector-based approach is more natural to deal with colour images. We illustrated both approaches with several examples. The results of these experiments showed that the vector-based similarity measures perform better than the component-based similarity measures.

## References


1. G. Beliakov, *Definition of general aggregation operators through similarity relations*. Fuzzy Sets and Systems, Vol. 114, pp. 437–453, 2000.
2. T. Chaira, A.K. Ray, *Fuzzy Measures for Color Image Retrieval*. Fuzzy Sets and Systems, Vol. 150 (3), pp. 545–560, 2005.
3. S.M. Chen, *Measures of similarity between vague sets*. Fuzzy Sets and Systems, Vol. 74, pp. 217–223, 1995.
4. S.M. Chen, M.S. Yeh, and P.Y. Hsiao, *A comparison of similarity measures of fuzzy values*. Fuzzy Sets and Systems Vol. 72, pp. 79–89, 1995.
5. M. De Cock and E.E. Kerre, *On (un)suitable fuzzy relations to model approximate equality*. Fuzzy Sets and Systems, Vol. 133 (2), pp. 137–153, 2003.
6. V. De Witte, S. Schulte, M. Nachtegael, D. Van der Weken, E. Kerre, *Vector Morphological Operators for Colour Images*. Lecture Notes in Computer Science, Vol. 3656, Image Analysis and Recognition: Second International Conference, ICIAR 2005, Toronto, Canada, September 2005, Proceedings, pp. 667–675.
7. J. Fan, W. Xie, *Some Notes on Similarity Measure and Proximity Measures*. Fuzzy Sets and Systems, Vol. 101, pp. 403–412, 1999.
8. G. Louverdis, M.I. Vardavoulia, I. Andreadis, Ph. Tsalides, *A New Approach to Morphological Color Image Processing*. Pattern Recognition, Vol. 35, pp. 1733–1741, 2002.
9. C. Mencar, G. Castellano, A.M. Fanelli, *Distinguishability Quantification of Fuzzy Sets*. Information Sciences, in press.
10. C.P. Pappis, and N.I. Karacipilidis, *A comparative assessment of measures of similarity of fuzzy values*. Fuzzy Sets and Systems, Vol. 56, pp. 171–174, 1993.
11. John C. Russ, *The Image Processing Handbook*, CRC Press, 1999.
12. G. Sharma, *Digital Color Imaging Handbook*, CRC Press, Boca Raton, 2003.
13. D. Van der Weken, M. Nachtegael, and E.E. Kerre, *An overview of similarity measures for images*. Proceedings of ICASSP'2002 (IEEE International Conference on Acoustics, Speech and Signal Processing), Orlando, United States, 2002, pp. 3317–3320.
14. D. Van der Weken, M. Nachtegael, and E.E. Kerre, *Using Similarity Measures for Histogram Comparison*. Lecture Notes in Artificial Intelligence, Vol. 2715, pp. 396–403, 2003.
15. D. Van der Weken, M. Nachtegael, and E.E. Kerre, *Using Similarity Measures and Homogeneity for the Comparison of Images*. Image and Vision Computing, Vol. 22 (9), pp. 695–702, 2004.
16. D. Van der Weken, M. Nachtegael, and E.E. Kerre, *Improved Image Quality Measures Using Ordered Histograms*. Proceedings of MMSP'2004 (IEEE Signal Processing Society International Workshop on Multimedia Signal Processing, Siena, Italy, September 2004), pp. 67–70.
17. D. Van der Weken, M. Nachtegael, E.E Kerre, *Some New Similarity Measures for Histograms*. Proceedings of ICVGIP'2004 (4th Indian Conference on Computer Vision, Graphics and Image Processing, Kolkata, India, December 2004), pp. 441–446.
18. D. Van der Weken, M. Nachtegael, V. De Witte, S. Schulte, E.E. Kerre, *Constructing Similarity Measures for Colour Images*. Proceedings of IFSA'2005 (International Fuzzy Systems Association World Congress on Fuzzy Logic, Soft Computing and Computational Intelligence Theories and Applications, Beijing, July 28–31), pp. 1020–1025.
19. D. Van der Weken, V. De Witte, M. Nachtegael, S. Schulte, E.E. Kerre, *Colour Image Comparison Using Vector Operators*. Proceedings of CIMCA'2005 (International Conference on Computational Intelligence for Modelling, Control and Automation, Vienna, Austria, November 2005), pp. 295–300.
20. B.A. Wandell, *Foundations of Vision*. Sinauer Associates, Inc., 1995.
21. L.A. Zadeh, *Similarity Relations and Fuzzy Orderings*. Information Sciences, Vol. 3, pp. 177–200, 1971.
22. W. Zeng, H. Li, *Relationship Between Similarity Measure and Entropy of Interval Valued Fuzzy Sets*. Fuzzy Sets and Systems, Vol. 157 (1), pp. 1477–1484, 2006.

23. H. Zhai, P. Chavel, Y. Wang, S. Zhang, Y. Liang, *Weighted Fuzzy Correlation for Similarity Measures of Color-histograms*. Optics Communications, Vol. 247 (1–3), pp. 49–55, 2005.
24. C. Zhang, H. Fu, *Similarity Measures on Three Kinds of Fuzzy Sets*. Pattern Recognition Letters, Vol. 27 (12), pp. 1307–1317, 2006.

# A Fuzzy-based Automated Cells Detection System for Color Pap Smear Tests -FACSDS-

**Pilar Sobrevilla, Eduard Montseny and Enrique Lerma**

**Abstract** There is a compelling need for automated cervical smear screening systems to improve the quality and cost/efficiency screening rate. Computer-assisted devices can reduce false negative Pap smear interpretations using computerized systems to assist the cytotechnologist in identifying Pap smear abnormalities and providing added value in their ability to consistently and objectively analyze all cells on slides without fatigue. However, automation of the process is a challenging problem due to the large variability in conventional Pap smears exhibiting no standard appearance and tremendous amount of data to be processed. Moreover, smear diagnostic may be obscured by benign conditions, overlapping cells, debris, inflammation, and no uniform staining.

Here we propose an efficient and fast Fuzzy-based Automated Cells Screening Detection System -**FACSDS**-, which can be useful for future Automatic Cells Screening System **ACSS**-. Because of detecting abnormal cells in a Pap smear can be refereed to as a "**rare**" event problem due to the normal cells and artifacts outnumber the intraepithelial lesions, the proposed algorithm has been divided into two steps. At first step the Areas of Interest **AOI**-, or best areas for screening in the smear, are detected and the degree to which these areas are interesting is given by means their evaluation goodness degree. At second step the **AOI**s are analyzed, taking into account their evaluation goodness degree, for detecting the cell nucleus. First step is carried out on monochrome images obtained using a 2.5X objective, and the results obtained have provided a high concordance degree with the cytotechnologist evaluation. The automatic system implemented at second step for detecting the nuclei is based on color information. We have considered color images because cells nuclei appear as dark regions within the images, hardly detected on monochrome images. Moreover, as color-order systems based on perceptual variables are somehow correlated with human beings color perception, and their coordinates are highly independent, what makes possible to treat achromatic and chromatic information separately, the proposed algorithm first convert **RGB** into **Hue**, **Saturation** and **Intensity** (**HSI**) color components. In addition, we make use of fuzzy techniques to face up the problems due to low saturation and illumination.

H. Bustince et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.*  

# 1 Introduction

Cervical cancer is an important cause of death in women worldwide. It originates on the surface of the uterine cervix or in the endocervical canal. Precursor lesions are epithelial dysplasia and/or Carcinoma In Situ (CIS). Some of these early lesions develop into invasive cervical cancer. Generally, the progression to invasive cancer is a slow, predictable pattern. Longitudinal studies indicate that 30 to 70 percent of untreated patients with Cervical Intraepithelial Neoplasia (CIN) will develop invasive carcinoma in 10–12 years. In about 10 percent of patients with CIN, lesions progress to invasive carcinoma in less than one year. Prognosis for the disease is influenced by the stage, size and grade of tumor at the time of detection, the tumors histologic type, and whether or not it has spread via the blood or lymphatic systems [41, 50].

Pap smear is the primary screening method for the detection of cervical cancer and its precursor [14, 33, 34, 36]. The Pap test is a subjective method that has remained substantially unchanged for more than 60 years, and there are several concerns regarding its performance [23]. When properly obtained and prepared, Pap smears alert practitioners to possible abnormalities of the cervix. Manual microscopic screening of Pap smears by cytotechnologist continues being the most widely used and accepted method of identifying irregular cells on cellular abnormalities of the cervix. During manual screening, at a first step, cytotechnologists search for what may be a few pathologic cells inspecting many fields on a slide under the microscope. They microscopically examine the morphologic features of the cells, relate these findings to the patient's clinical history, and render a cytological impression. In a second step cytopathologists diagnose disease by analyzing cells previously selected by the cytothechnologist manipulating the microscope stage through visual inspection of approximately 50.000 to 300.000 epithelial and inflammatory cells per smear. Even smears with serious abnormalities may contain only a few dozen malignant or pre-malignant cells among hundreds of thousands of normal cells [19, 31].

Over the last 60 years several modifications of the terminology for cytological diagnosis has been introduced. Today the most accepted one is The Bethesda System [15], proposed initially in 1989, and actualized in 2001 [40]. In this system, cytologies are divided into two classes: negatives and pathological. In the latest group there are, according to the severity of the lesion, four diagnostic groups: **atypia**, **low grade lesion**, **high grade lesion** and **carcinoma**. The main objective of screening is in identifying these pathological groups, from **atypia** until **carcinoma**. Most **atypias** can be resolved spontaneously, but up to 30% can progress to a **carcinoma** if they are not adequately controlled [16]. In addition, most **lesions** would progress to a **carcinoma** [12, 13, 15].

Medical community and consumers are concerned about the number of "**false negative**" results (i.e. a normal cytologic report for a woman with existing dysplasia, pre-malignant or malignant lesions of the cervix [24]) generated by the manual technique [32]. It is estimated that manual Pap smear test suffers from a high number of "**false negative**" which rate varies from 5% to over 55%, although most are in the range of 20–30%. When these cells are missed, resulting delays in diagnosis can lead to progression of the disease and the need for more aggressive treatment. "**False negative**" can occur under following conditions:

- **Cell sampling and preparation**: At the time of collection, samples of cells collected from the surface of the cervix must contain both ectocervical and endocervical cells. If abnormal cells are not included in the sample, abnormalities cannot be detected. "**False negatives**" can also occur if blood, vaginal secretions, or inflammatory cells in the sample conceal the abnormal cells making them undetectable by the human eye.
- **Screening and interpretation**: "**False negative**" results due to screening are caused by failure to identify a relatively small number of cells in the screening process. A trained cytotechnologist typically searches hundreds of cells per smear on up to 100 slides per day [20].

About two-thirds of "**False negative**" are a result of sampling error and the remaining one-third a result of detection error [45]. Techniques have been introduced to improve screening and interpretation accuracy through improvements in sampling and slide preparation methods. Each of the new technologies is directed at one of these components of "**False negative**". So, while Thin-Layer cytology aims primarily to fix sampling error [17, 18], computerized re-screening PAPNET and AutoPap systems target detection error [1, 46]. This implies that neither technology will be able to reduce "**False negative**" beyond a certain threshold. Moreover, while the added value of these technologies in improving the net health outcome of women and preventing cervical cancer has not been determined, questions have been raised concerning their cost-effectiveness [10, 22].

Factors that can lead or contribute to a misdiagnosis include habituation, cytotechnologist fatigue, incomplete screening of a specimen, and cytotechnologist shortage. All of these factors contribute to human error. It is because of the likelihood of human error that there is a compelling need for low-cost automated systems to improve the quality and efficiency of screening and diagnosis. Computer-assisted devices can reduce false negative Pap smear interpretations. The use of computerized systems could both, assist the cytotechnologist in identifying Pap smear abnormalities and provide added value in their ability to consistently and objectively analyze all cells on slides without fatigue.

However, automation of the process is challenging due to the tremendous amount of data to be processed. Detecting abnormal cells in a Pap smear can be refereed to as a "**rare**" event problem or a needle-in-a-haystack problem. Typically 9600 images (measuring 400 microns by 400 microns) have to be processed for each conventional Pap smear slide. 95 percent of all slides that are processed will be normal and have no abnormal cells. Of the 5 percent of slides that have abnormal cells, perhaps only 100 of the 300,000 cells will show signs of abnormality.

To carry out the Pap Smear Screening process technicians must first discard the areas wherein cells can't be analyzed because are superimposed, there is blood or other artifacts, there are no enough cells in good conditions to be identified, or other problems that make impossible the correct cells visualization and the posterior cells identification. This discarding process is performed using a 2.5X objective.

In a second stage the technician proceeds to a more specific analysis on the images obtained using a 10X objective. This stage has two essential and final aims: locate the cells feasible of being abnormal and to account up the identified normal cells. The potentially abnormal cells have to be located to be analyzed and identified in a following stage. Cells normality or abnormality is determined based on geometric features of their nuclei. As normal cells are counted but won't be undergone to further analysis, in a first step the features of the nuclei matching with normal cells are specified. The cells whose nuclei don't satisfy those features will be labelled as Abnormal/Doubtful, and the shape, size and characteristics of their cytoplasm are analyzed.

Some times the cytoplasm's analysis allows to the cytopatologist give a diagnostic of the cells Normality/Abnormality. If it is not possible, in order to observe the cells' features more precisely and be able of identifying the Doubtful cells, the technician uses a 40X objective.

Since many of the structural and molecular changes occur within the cell's nucleus, the ability to segment the individual nuclei is an important and basic technical capability. Segmentation performance can be greatly improved by incorporating priory knowledge about the specific type of images being segmented.

Considering the problems involved within manual Pap smear tests we propose an efficient and fast Fuzzy-based Automated Cell Screening Detection System -**FACSDS**- suitable of interacting with the human technologist, which can be useful for future Automatic Cells Screening System -**ACSS**-. Taken into consideration the process followed by the technicians the process is carried out in two differentiated steps:

1. **Detection and evaluation of the areas wherein cells can be identified**. This step will be carried out on monochrome images obtained using a 2.5X objective. To intelligently eliminating normal images (the "**hay**") a degree of interest will be also associated to each region. So, normal images can be eliminated and suspicious images saved and evaluated for further processing.
2. **Nuclei detection**. To do it the regions to which a high interesting degree has been associated at previous step will be analyzed on color images obtained using a 10X objective. We will consider a color representation that prevents non-homogeneity problems due to illumination and shadows, allowing color recognition process to be independent of illumination [29]. Moreover, the process is based on fuzzy techniques, so that the natural variability of color data is well accommodated.

The paper is organized as follows: After laying, in Sect. 2, the foundations of the considered color perceptual, at Sect. 3, after introducing the method considered for representing the vagueness of both, color pixels and pixels' illumination intensity, we obtain the membership degree of each pixel to the Regions of Interest. Then, the values obtained at Sect. 3 are considered for detecting and locating the cells nuclei. Finally, Sect. 3 presents some results obtained by the proposed approach and the conclusions.

## 2 Color Perceptual System

A color space is a way of representing colors and their relationship to each other. Different image processing systems use different color models for different reasons. However, as color is not an absolute characteristic of an object, but a human perception, color-order systems based on perceptual variables are more convenient for computer vision applications, as they are somehow correlated with human being's color perception [28]. It can be said that the perceptual variables **hue**, **saturation** and **lightness** are related to the psychophysical variables **dominant wavelength**, **purity** and **luminance**, respectively.

The majority of color segmentation algorithms work with the **RGB** values of pixels, what may be misleading, especially when using distances within the **RGB** space to compare color differences, as explained by Chien and Cheng in [8]. Although a lot of color segmentation algorithms work on the **RGB** color space [6], the number of researchers that make use of the **Hue-Saturation-Intensity** (**HSI**) components attending to their perceptual meaning is slowly increasing [4, 5, 7, 26, 35, 43, 44]. Moreover, two principal factors make the **HSI** model ideal for developing machine vision applications: decoupling the intensity component from the color information, and the close relationship between chromaticity and how humans perceive color.

When working with **HSI** representation of color it must be taken into account the varying noise-sensitivity within the color space. Some authors have studied and modelled the propagation of constant noise through several **RGB** to **HSI** transformations, and Gevers [9] proved that the intrinsic error of each **RGB** channel can be modelled as a normal zero-mean distribution with nearly constant deviation for the whole space. Moreover, **HSI** components variability has to be considered taking into account following statements [3, 11, 44]:

– **Low Saturation -S-** increases the **Hue -H-** standard deviation: if $S = 0$ then the value of **H** gets undefined.
– **Low Intensity -I-** increases the **Hue** and **Saturation** standard deviations: if $I = 0$ then both **H** and **S** values get undefined.

With regard the **H** and **S** invariance in front of intensity variations of illumination, typically produced by shadows, shading and highlights, some **HSI** formulations, like Smith's and Tenenbaum's ones [37, 42], provide stable Hue and Saturation values from a colored surface illuminated with different intensities. In this sense Perez and Koch [26], based on algebraic manipulations of its formulation, proved that **Hue** is invariant to uniform shifting and scaling in **RGB** space. Moreover, in the case of the Smith's color space, we have proved [29] that **Saturation** is invariant to uniform scaling in **RGB** space.

From these properties, and taking into account that shadowing and shading produce uniform $RGB$ scaling while highlights produce uniform **RGB** shifting [26], we can conclude that **Hue** keeps invariant when both artifacts appear, but **Saturation** only keeps invariant for shadowing and shading. Other authors [2, 25] experimented on the **Hue** invariance of color samples under controlled illumination level variations, and tested the **Saturation** invariance for several color models, founding that **Saturation** is less invariant than **Hue**.

A great many color transformation formulae have been proposed in order to convert the basic $RGB$ coordinates into several **HSI** spaces [27, 37, 42, 49], that can appear with different notations (**HSV**, **HSB**, **IHS**, **HIS**, etc.). Among the available **HSI** models [47] we have chosen Smith's one [37] because: assures a high independence degree among the three components, is easy to compute, shows the maximum component reliability under illumination variations, and allows to avoid the variability due to the **RGB** representation of color characteristics.

In present work, we will consider the chromatic components (**Hue** and **Saturation**) for representing the color feature of the scene, while the achromatic component (**Intensity**) will be interpreted as a color-independent feature measuring the intensity of light reflected by those objects. Moreover, we really believe that **Intensity** has to be handled separately from the color characteristic, because it does not intrinsically belong to the chromaticity of a color and we are very confident of the fact that chromatic features are very stable on illumination-varying environments. A positive aspect of this color representation is that our algorithm will be aware from the non-homogeneity due to illumination and shadows problems, because these artifacts modify only the Smith's **I** component, and should let **H** and **S** components invariant.

As we pretend to develop a system robust with regard to variable illuminant level conditions, but taking into account color components' variability degrees, we will make use of the **Reliability Functions** introduced in [30] for predicting the expected reliability of the Smith's **Hue** and **Saturation** data for any color pixel. So, if $H_{ij}$, $S_{ij}$, and $I_{ij}$ represent the **Hue**, **Saturation**, and **Intensity** values of a pixel color $p_{ij}$, its **Hue** and **Saturation Reliability values** are given by expressions (1) and (2), wherein the scalar factors $\phi_H$ and $\phi_S$ are chosen to fit conveniently the behavior of real data.

$$R_H(p_{ij}) = Min\left(1, \frac{(\phi_H \cdot S_{ij} \cdot I_{ij})}{20000}\right) \tag{1}$$

$$R_S(p_{ij}) = Min\left(1, \frac{(\phi_S \cdot I_{ij})}{500}\right) \tag{2}$$

Previous **Reliability Functions** are thought for taking into account the variability of color components, and are defined so that the higher the deviation from the real component value, the lower the stability, and vice-versa. Moreover, these functions have been limited to vary within the unit interval.

Working on real color samples captured within a varying illumination level environment, some reliability behaviors have been observed. On the one hand, it has been noticed that **Hue** and **Saturation Reliability Functions** depend on the intensity value in such a way that Intensity values lower than 120 imply low **Reliability Functions** values. On the other hand, even for high Intensity values, the **Reliability Functions** values are dependant on the **Saturation** values because low **Saturation** values correspond to a "light" dye feature.

# 3 Areas of Interest -AOI- Detection and Evaluation

Bearing in mind the problems involved within manual Pap smear tests, and taken into consideration the process followed by the technicians, the first step for developing an automatic and cost-effective cells detection system will be to carry out an **Areas of Interest -AOI- detection and evaluation process**. On the one hand, this step will allow marking the areas that must not be later analyzed by the technician and, on the other hand, the process will provide with the importance degree with which the rest of regions will have to be analyzed, so giving a priority degree very useful for later system steps.

To carry out this step at a reduced cost/efficiency screening rate, but considering cells images properties, it has been applied to grey-scale images because of they are less sensitive to variations of lighting conditions and staining quality than those of color images. So, due to original images are color ones, to avoid a further monochrome additional capture, and to increment the overall cost, the monochrome information of each pixel $p_{ij}$ will obtained converting the **RGB** values ($R_{ij}$, $G_{ij}$, $B_{ij}$), obtained using a 2.5X objective, to grey-level ones ($Gl(p_{ij})$) applying the expression:

$$Gl(p_{ij}) = 0.6 \cdot R_{ij} + 0.3 \cdot G_{ij} + 0.1 \cdot B_{ij}$$

## *3.1 Region Segmentation by Pixel Classification*

To incorporate as much as possible cyto-technologists' knowledge and experience avoiding their subjectivity in such a way that the results provided by the **AOI** detection be independent of human bias and error, our system makes use of segmentation by pixel classification, wherein local features are analyzed using fuzzy techniques [21, 38].

Following the usual structure of pixel classification systems, the detection and location of the areas wherein cells can be identified (**AOI**), will be based on region segmentation [21]. This is why, after segmenting the image into distinguishably different regions, in a second step the results obtained will be analyzed for recognizing the **AOI**s.

### 3.1.1 Regions Description and Evaluation

To implement these steps in a suitable way, vagueness factors introduced by vaginal secretions, inflammatory cells, cell-clumbing, blood staining, air-drying and other cell sampling and preparation artifacts have to be taken into consideration.

So, following the technician way of work, after an analysis of the images focused on grey level and texture characteristics, the regions of the cell images that will be considered by our pixel classification system are:

- **Cells**,
- **Background**,
- **Region 1**: that includes inflammatory cells, cell-clumbing and bloodstaining,
- **Region 2**: that contains vaginal secretions, air-drying and other cell sampling and preparation artifacts, and
- **Contours**: which are transition areas with special characteristics.

For detecting these regions within the cell images, we carry out an analysis based on their local grey level lightness and homogeneity degree that will provide the degree to which the pixels belong to the regions. To do it, first the regions' description is as given in Table 1.

Then, following the framework of [38], and considering variability and vagueness within the images, as well as the linguistic descriptions of the elements, we represent the fuzzy knowledge by the fuzzy rule base given by:

$$\mathrm{R}_{\mathrm{Cell}} : if\ \mathrm{Gl}(p_{ij})\ is\ \mathrm{LM}\ and\ \mathrm{Hom}(p_{ij})\ is\ \mathrm{M},\ then\ p_{ij}\ is\ Cell.$$

$$\mathrm{R}_{\mathrm{B}} : if\ \mathrm{Gl}(p_{ij})\ is\ \mathrm{L}\ and\ \mathrm{Hom}(p_{ij})\ is\ \mathrm{H},\ then\ p_{ij}\ is\ Background.$$

$$\mathrm{R}_{\mathrm{BCont}} : if\ \mathrm{Gl}(p_{ij})\ is\ \mathrm{L}\ and\ \mathrm{Hom}(p_{ij})\ is\ \mathrm{VS},\\ then\ p_{ij}\ is\ Contour\ or\ Background.$$

$$\mathrm{R}_{\mathrm{R1}} : if\ \mathrm{Gl}(p_{ij})\ is\ \mathrm{DM}\ and\ \mathrm{Hom}(p_{ij})\ is\ \mathrm{VH},\\ then\ p_{ij}\ is\ Region1.$$

$$\mathrm{R}_{\mathrm{R1Cont}} : if\ \mathrm{Gl}(p_{ij})\ is\ \mathrm{DM}\ and\ \mathrm{Hom}(p_{ij})\ is\ \mathrm{VS},\\ then\ p_{ij}\ is\ Contour\ of\ Region1.$$

$$\mathrm{R}_{\mathrm{R2}} : if\ \mathrm{Gl}(p_{ij})\ is\ \mathrm{LM}\ and\ \mathrm{Hom}(p_{ij})\ is\ \mathrm{VH},\\ then\ p_{ij}\ is\ Region2.$$

To get the degree to which each pixel belongs to the regions, and so the membership functions of the consequent fuzzy sets, the grey level lightness and homogeneity degrees are evaluated by means of the average and standard deviation of the grey values measured over a 3x3 window, respectively. Then, the membership functions $\mu_{GlL}$, $\mu_{GlL-M}$, $\mu_{GlD-M}$, $\mu_{HomVS}$, $\mu_{HomM}$, $\mu_{HomH}$, and $\mu_{HomVH}$ are

**Table 1** Regions' description

| Regions | Grey level | Homogeneity |
|---|---|---|
| Cell (**Cell**) | Light-Medium (**LM**) | Medium (**M**) |
| Background (**B**) | Light (**L**) | High (**H**) |
| Back-Contour (**B-Cont**) | Light (**L**) | Very Small (**VS**) |
| Region 1 (**R1**) | Dark-Medium (**DM**) | Very high (**VH**) |
| Region 1-Contour (**R1-Cont**) | Dark-Medium (**DM**) | Very Small (**VS**) |
| Region 2 (**R2**) | Light-Medium (**LM**) | Very high (**VH**) |

obtained starting from the probability density and distribution functions of these values, evaluated over the training images following the process explained at [21]. The fuzzy sets so obtained are aggregated by the minimum, according to previous rules to get the regions fuzzy sets determined by their membership functions:

$$\mu_{Cell}(p_{ij}),\ \mu_B(p_{ij}),\ \mu_{BCont}(p_{ij}),\ \mu_{R1}(p_{ij}),\ \mu_{R1Cont}(p_{ij}),\ \mu_{R2}(p_{ij})$$

### 3.2 AOI Detection

To provide the cytotechnologist with an accurate **Areas of Interest** (**AOI**) detection and location, the output of the system must give information about all the best areas for screening. The identification of the Areas of Interest can be performed locating them directly, or finding and eliminating the Areas Of No Interest -**AONI**-. Considering the high conceptual level of cell images, and to be sure that all best areas for screening are obtained, the output of the proposed **FACSDS** will be the optimum obtained comparing the outputs provided by both methods.

Therefore, next step will consist in detecting the Areas of no Interest. This process will be carried out applying next rule:

$$\text{R}_{\text{AONI}} : if\ (p_{ij})\ is\ Bacground,\ or\ Region1,\ or\ Region2\ then\ p_{ij}\ is\ AONI.$$

And the **AONI** membership function will be obtained implementing this rule as follows:

$$\mu_{AONI}(p_{ij}) = Max\{\mu_B(p_{ij}), \mu_{R1}(p_{ij}), \mu_{R2}(p_{ij})\}$$

Once the **AONI** has been obtained, we go on to obtain the **AOI**, which is accomplished applying the classic complement to $\mu_{AONI}$.

### 3.3 Areas of Interest Evaluation

For deciding the importance degree of each area for its evaluation, the proposed approach checks that the **AOI**s obtained by the direct -**Cells**- and indirect -**AOI**- methods are concordant. To do it, given an image **I**, after obtaining the **Cells** and **AOI** membership degrees of each pixel $p_{ij}$ ( $\mu_{Cell}(p_{ij})$, and $\mu_{AOI}(p_{ij})$ ), to decide if this image has to be analyzed, as well as the membership value with which this analysis has to be performed, next rule will be applied:

$$\textbf{RCells}: if\ \ \forall p_{ij} \in I\ \ \mu_{AOI}(p_{ij}) \overset{>}{\approx} \mu_{Cell}(p_{ij})\ \ then\ I\ is\ analyzed,$$
$$else\ \ I\ is\ rejected.$$

If I is considered to be analyzed, $\mu_{AOI}$ will be the **Area of Interest** fuzzy set to be considered.

Then, to get the final Areas that will be given to the cytotechnologist, and their importance degree, a regional analysis will be performed. In this analysis a **Fuzzy Morphological Structural Element** -**FMSE**- [39] is used to detect the areas of the image in which there exist interesting cells in a high enough number that will ponder next conditions:

1. The region containing cells must be compact.
2. The presence of cells is important in the inner as well as in the periphery of the region.

In accordance with these conditions, and considering that an element without edges matches better the regions to be segmented, and a small element reduces the computational cost allowing detecting small regions containing cells, a circle **FMSE** having a 2.9 pixels radius has been chosen. Moreover, based on the small number of pixels within the considered element, it has been divided into one inner region $R_1$, having 1.6 pixels' radius, and four outer ($R_2$ to $R_5$) sub-regions. This division has been performed with the aim of getting similar number of pixels in all regions, allowing analyzing the homogeneity of the presence of cells.

Bearing in mind the second aforementioned condition, and that we are looking for regions in the image with a minimum of cells, the same level of demand is required to all sub-regions.

As the **FMSE** moves over the image, and all the pixels covered by it have a membership degree to the **Area of Interest** fuzzy set, $\mu_{AOI}(p_{ij})$, these membership degrees must be aggregated bearing in mind the second aforementioned condition, and that we are looking for regions in the image with a minimum of cells. This is why we have considered OWA operators to aggregate the membership degrees of the pixels inside each sub-region.

So, in the case of Region $R_1$ we have considered the OWA operator [48] of dimension 9, whose associated vector is given by (3), and for Regions $R_2$ to $R_5$ we have used an operator of dimension 4 with associated vector (4). This way, a membership degree $\mu_{R_r}$ will be assigned to all the pixels covered by sub-region $R_r$, $1 \leq r \leq 5$.

$$\{W_1 = \{0, 0.4, 0.25, 0.25, 0.1, 0, 0, 0, 0\}\} \tag{3}$$

$$\{W_2 = \{0.25, 0.25, 0.25, 0.25\}\} \tag{4}$$

Moreover, as the same level of demand is required to all sub-regions, the five membership degrees are associated by the minimum and the value obtained is assigned to all the pixels covered by the **FMSE**, i.e. saying $\forall p_{ij} \in I_{FMSE}$ the set of pixels covered by the **FMSE**, then:

$$\eta_{I_{FMSE}}(p_{ij}) = \min_{1 \leq r \leq 5}\{\mu_{R_r}\} \tag{5}$$

In order to homogenize the **Area of Interest** membership degrees, and use the **FMSE** as a Morphological Closing, previous membership values have been considered to modify the original **AOI** membership degrees such that $\forall p_{ij} \in I_{FMSE}$:

$$if\ \mu_{AOI}(p_{ij}) < \eta_{I_{FMSE}}(p_{ij})\ then\ \mu_{AOI}(p_{ij}) < \eta_{I_{FMSE}}(p_{ij}) \tag{6}$$

## 4 Fuzzy Nuclei Detection

Nuclei detection and location has been carried out on color images due to these images hold more information, what make easier both, nuclei detection and the analysis that will be performed in subsequent stages for identifying the kind of cells appearing within the image. Moreover, it has to be taken into consideration that within these images there is blood, which appears as red stains that in a monochrome image is converted into a dark grey-level very similar to the one of the cells nuclei, so obstructing their location. Moreover, as color-order systems based on perceptual variables are somehow correlated with human being's color perception, and their coordinates are highly independent, what makes possible to treat achromatic and chromatic information separately, the Smith's chromatic components.

The process followed for achieving the nuclei detection works as follows: after introducing the method considered for representing the vagueness of each color pixel (subsection 4.1), we represent the vagueness of the pixels' illumination intensity (subsection 4.2). Finally (Sect. 4.3), making use of the information obtained at previous steps, we obtain the membership degree of each pixel to the Cells Regions.

### *4.1 Color Pixels' Vagueness Representation*

As we consider the **Hue** and **Saturation** for representing the color feature of the scene, the system is based on a palette of 92 colors, $C^k (1 \leq k \leq 92)$, defined by their **H** and **S** values $(H_i, S_i)$ and distributed over the **H-S** map. The selection of these color patterns has been carried out subject to next commitment: **"The number of color patterns must be big enough for assuring that two pixels are classified to different pattern colors if the pixels' components differ, but no so big as to classify to different pattern colors pixels belonging to the same color or make more difficult the algorithm"**. For a suitable distribution map, bearing in mind the particular behavior of color components, the Saturation interval has been divided into five subintervals labelled: **Very Low**, **Low**, **Medium**, **High**, and **Very High**. Then, considering the Saturation/Hue color behavior, and their relation, we have chosen one color with very low saturation, 5 colors with low saturation, 9 with medium, 13 with high, and 64 equidistant distributed colors with very high saturation.

For representing color vagueness, each pattern color $C^k$ has associated a fuzzy set $\mu_{C^k}$, given by:

$$\mu_{C^k}(p_{ij}) = \sqrt{\mu^H_{C^k}(p_{ij}) \cdot \mu^S_{C^k}(p_{ij})} \tag{7}$$

such that, if $(H_{ij}, S_{ij}, I_{ij})$ are the $Hue$, $Saturation$, and $Intensity$ values of pixel $p_{ij}$, and $(H^k, S^k)$ are the $Hue$ and $Saturation$ of color $C^k$, then $\mu^H_{C^k}(p_{ij})$ and $\mu^S_{C^k}(p_{ij})$ represent the degrees to which the $Hue$ and $Saturation$ of the pixel are similar to the color ones. These membership degrees are obtained taking into account the $Hue$ and $SaturationReliabilityvalues$, ($R_H$ and $R_S$) according to next expressions:

$$mu^{H}_{C^k}(p_{ij}) = \mu^{H}_{C^k}(H^k - (H^k - H_{ij}) \cdot R_H(p_{ij})) \tag{8}$$

$$\mu^{S}_{C^k}(p_{ij}) = \mu^{S}_{C^k}(S^k - (S^k - S_{ij}) \cdot R_S(p_{ij})) \tag{9}$$

### 4.2 Pixels' Illumination Intensity Vagueness Representation

As a previous step for getting the Nuclei Regions we must take into account that these regions appear within the images as poorly illuminated areas. Therefore, for each pattern color $C^k$, we have defined the fuzzy set $C^k - shadow$, whose membership function, $\mu^{Shadow}_{C^k}$, is obtained according to next process:

1. $\forall p_{ij}$ such that $\mu_{C^k}(p_{ij}) > 0.5$ we obtain the histogram of intensity frequencies.
2. Within the histograms we get the minimum and maximum values: $I^k_{min}$, $I^k_{max}$.
3. Then, the fuzzy set $C^k - shadow$ is defined by the membership function:

$$\mu^{Shadow}_{C^k}(p_{ij}) = \begin{cases} 1 - \dfrac{I_{ij} - I^k_{min}}{(I^k_{max} - I^k_{min}) \cdot 0.4} & if\ \mu_{C^k}(p_{ij}) > 0.5\ and\ C^k \in CN \\ 0 & else \end{cases}$$

where $CN$ is the set of colors associated with the nucleus that has been experimentally determined and is constituted by the colors that can present the nuclei when react with the preservative liquids. Moreover, previous membership value is restricted to belong to the unit interval.

### 4.3 Obtaining the Nuclei Fuzzy Sets

Finally, the membership degree to which each pixel belongs to the Nuclei fuzzy set is obtained by:

$$\mu_{Nuclei}(p_{ij}) = \max_{1 \le k \le 92} \{mu_{C^k}(p_{ij}) \cdot mu^{Shadow}_{C^k}(p_{ij})\};\ \forall p_{ij}$$

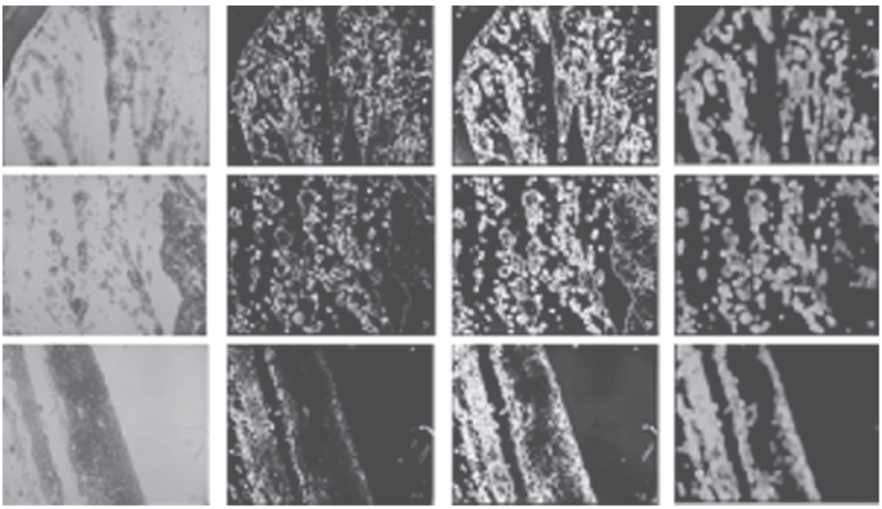
## 5 Results and Conclusions

As presented the system has been developed within two differentiated parts, one devoted to the Areas of Interest detection and the other addressed to the cells nuclei detection. These two parts have been verified in an independent way in order to guarantee their good performance. The results of the tests carried out have been very satisfactory, providing very accurate results in the case of $AOI$ detection, and promising results in the case of nuclei detection. However, it must be pointed out that, as nuclei detection tests have been carried out considering not only the $AOI$ but the whole sample, the majority of false detections are located on the areas don't marked as $AOI$s at previous step.

Next, some results obtained at each part of the proposed system are going to be presented.

## 5.1 Areas of Interest Detection

This part has been tested on more than 100 images obtained by a system formed by a JVC TK-C1381 video color camera attached to a BH2 Olympus microscope and connected to a PC. In each slide, the screening field detection was tested with the 2.5X objective. Afterwards, the system has been tried out by expert cytopatologists for a period of three months, during which they verified its good operation, proving that the system versus expert concordance degree has been between 80 and 95 percent considering both $AOI$ detected and goodness degree, but greater than 95% if only the areas that must be analyzed are considered. Moreover, isolated cells and bidimensional groups were clearly differentiated from a-cellular areas or those with marked cell crowding that usually makes screening impossible. Dense mucus and contaminants, such as dust or air bubbles, were correctly discarded. Some results are presented at Fig. 1.



a) b) c) d)

**Fig. 1** **a**) Original image; **b**) results obtained for the Cells Fuzzy set -Direct ROI obtaining-; **c**) results obtained for the final ROI after comparison; and **d**) ROI after applying the Fuzzy Morphological Structural Element

Pictures a) of each row correspond to the original image over which the algorithm has been applied. The pictures labelled b) show the $Cells$ fuzzy set ($AOI$ using the direct method) obtained as explained at Sect. 3.2 by the direct method. In these pictures, higher grey level corresponds to pixels belonging with higher degree to the **Cells region**. Pictures labelled c) depict the $AOI$ fuzzy set obtained after applying the comparison between the **Cells region** and the $AOI$ obtained by the indirect method. As in b), higher grey level corresponds to pixels belonging with higher degree to $AOI$.

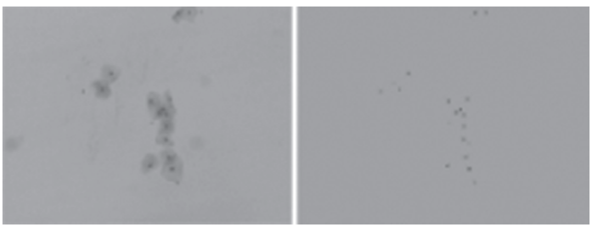
Finally, pictures d) correspond to the **Cells region** obtained applying the $FMSE$ to the images shown by pictures c) detected from original image, these images will be used by the cytotechnologist to perform the final analysis.

## *5.2 Nuclei Detection*

This step has been tested on 20 cervical smear color images from a database collected at the Hospital de Sant Pau de Barcelona, Spain. These images were obtained with the system previously described, but using a 10X objective. As previously said, the results are very promising considering that the majority of false detections are located on areas of the images that should be classified at previous step as no interesting regions. Here we present the results obtained after applying the proposed system to three real data images. It mist be taken into consideration that the images sides are no analyzed.

Having a look at the image of Fig. 2-a) it can be observed that cells appear in good conditions to be analyzed, and the results provided by the system (Fig. 2-b) show that almost all the nucleus have been detected and doesn't appear any false detection.

Within the original images of Fig. 3 the marked regions correspond to areas with cells that cannot be analyzed. Having a look at the results obtained (second column of Fig. 3) by the Nuclei detection algorithm, the system detects a large number of nuclei that in many cases correspond to false detections. However, it must be noticed that these "no good results" lack importance because, if the $AOI$ step was first applied, these regions never would be considered.



a) b)

**Fig. 2** **a**) Original image; **b**) Nuclei detected applying the process explained at Sect. 4

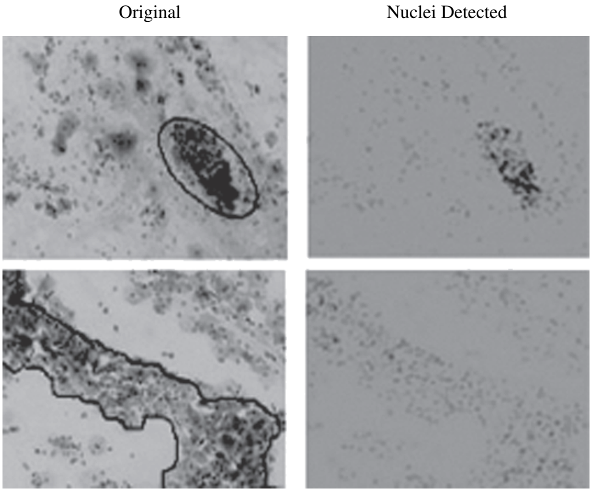



**Fig. 3** First column: Regions with cells that cannot be analyzed are surrounded. Second column: Nuclei detected applying the process explained at section 4

If we have a look at the two amplified areas of interest of image Fig. 4, depicted at first column of Fig. 5, and the results obtained for these two regions (second column of Fig. 5), it can be observed that the nucleus have been correctly detected.

Table 2 gives the details of the nuclei detection in images 3-a), 5-b), and 5-d). As can be seen most of the nuclei were detected, and no false alarm appear.

**Table 2** Results of the nulei detection step for the images of Figs. 3-a), 5-b), and 5-d)

| Image | Nuclei Number $NN$ | Nuclei Detected $ND$ | False Alarms $FA$ | Perform. $\rho$ | Efficiency $\xi$ |
|---|---|---|---|---|---|
| Fig.3-a | 20 | 19 | 0 | 0.95 | 0.95 |
| Fig.6-a | 49 | 45 | 0 | 0.918 | 0.918 |
| Fig.6-c | 53 | 49 | 0 | 0.924 | 0.924 |

The performance ($\rho$) and efficiency ($\xi$) measures, defined as $\rho = ND/NN$ and $\xi = ND/(NN + FA)$, are given at fifth and sixth columns of this table. As can be observed these values are very high and coincide due to no false has been detected.

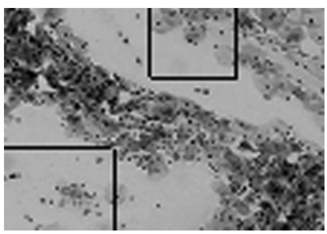


**Fig. 4** Original Image. Surrounded areas correspond to Regions of interest

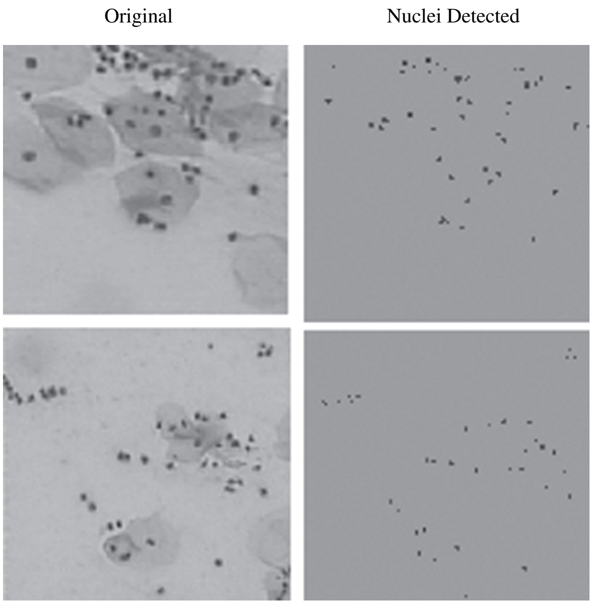



**Fig. 5** First column: Original image. Second column: Nuclei detected applying the process explained at Sect. 4

## *5.3 Conclusions*

We have proposed an efficient and very fast Fuzzy-based Automated Cells Screening Detection System algorithm, useful for future Automatic Cells Screening System, which simulates technologists' process, simplifies their work, and provides a very high Areas of Interest detection concordance percentage with the cytotechnologist.

The results obtained by the first step of the proposed approach for detecting $AOI$ within the smears show the effectiveness of fuzzy techniques in vagueness treatment. Furthermore, the Structural Element has seen to be very effective for segmentation based on pixel classification, when uniformity and gray level must be analyzed locally. On the other hand, the way this element has been applied allows us to obtain convex and compact regions.

Although it has not been the case in our tests, an open problem is to decide what to do if there are not concordances between the $AOIs$ obtained by the direct -**Cells Regions**- and indirect -**Regions Of Interest**- methods.

The proposed algorithm for Nuclei Detection has proved a good performance and efficiency, specially considering that, in general, the number of false detections has been lower than the 0.5% for all the images.

As a future work, the Nuclei detection step will be implemented starting from the information provided by the $AOI$ detection step.

**Acknowledgment** This work has been partially supported by the Spanish CICYT Project TIC2002-02791.

## References


1. *Assessment of automated primary screening on PAPNET of cervical smears in the PRISMATIC trial. PRISMATIC Project Management Team*, Lancet, 353, pp. 1381–1385, 1999
2. Berry, D. T., *Color recognition using spectral signatures*, Patten Recognition Letters, Vol.6, pp. 69–75, 1987
3. Carron, T. and Lambert, P., *Color edge detector using jointly Hue, Saturation and Intensity*, International Conference on Image Processing, pp. 977–981, 1994
4. Carron, T. and Lambert, P., *Symbolic fusion of Hue-Chroma-Intensity features for region segmentation*, International Conference on Image Processing, Vol.1, pp. 971–974, 1996
5. Celenk, M., *A color clustering technique for image segmentation*, Graphical Models and Image Processing, 52 Vol.3, pp. 145–170, 1990
6. Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J.L., *Color image segmentation: advances and prospects*, Pattern Recognition, 34 Vol.12, pp. 2259–2281, 2001
7. Cheng, H.D. and Sun, Y., *A hierarchical approach to color image segmentation using homogeneity*, IEEE Transactions on Image Processing, 9 Vol.12, pp. 2071–2082 2000
8. Chien, B.C. and Cheng, M.C., *A color image segmentation approach based on fuzzy similarity measure*, IEEE Int. Conf. on Fuzzy Systems, pp. 449–454, 2002
9. Gevers, T., *Robust histogram construction from color invariants*, IEEE Int. Conf. on Computer Vision, Vol.1, pp. 615-620, 2001
10. Grohs, D.H., *Impact of automated technology on the cervical cytologyc smear. A comparison of cost*, Acta Cytology, Vol.42, pp. 165–170, 1998
11. Kender, J., *Saturation, hue and normalized color: calculation, digitization effects and use*, Technical Rapport. Carneguie Mellon University, 1976
12. Koss, L.G., *The Papanicolaou test for cervical cancer detection: a triumph and a tragedy*, JAMA, Vol.261, pp. 737–743, 1989
13. Koss, L.G., Schreiber, K., Elberg, P. and Mango, L., *Evaluation of the PAPNET Cytologic system for quality control of cervical smears*, Am. J. Clin. Pathol, Vol.101, pp. 220–229, 1994
14. Kurman, R.J., Henson, D.E., Herbst, A.L., Noller, K.L. and Schiffman, M.H., *Interin guidelines of management of abnormal cervical cytology*, JAMA, Vol.271, pp. 1866–1869, 1994
15. Kurman, R.J. and Solomon, D., *The Bethesda System for reporting cervical/vaginal cytologyc diagnosis*, JAMA, Springer-Verlag, N. York, 1994
16. Lerma, E., Colomo, Ll., Carreras, A., Esteva, E., Quilez, M. and Prat, J., *Re-screening of atypical vaginal smears with PAPNET*, Cancer Cytopathology, Vol.84, pp. 361–365, 1998
17. Lin, W.M., et al., *Molecular Papanicolau tests in the twenty-first century: molecular analyses with fluid-based Papanicolau technology*, Am. J. Obstet. Gynecol., Vol.183, pp. 39–45, 2000
18. Malle, D., Pateinakis, P., Chakka, E. and Destonni, C., *Experience with a thin-layer, liquid-based cervical cytologic screening method*, Acta Cytology, Vol.47, pp. 129–134, 2003
19. Mango, L.J., *Reducing false negatives in clinical practice: the role of neural network technology*, Am J Obstet Gynecol, Vol.75, pp. 1114–1119, 1996
20. Maremont, M., *Cytyc test for cervical cancer offers minimal advancement study says*, Wall Street Journal, 1998
21. Montseny, E. and Sobrevilla, P., *Application of fuzzy techniques to the design of algorithms in computer vision*, Mathware and Soft Computing, Vol.2-3, pp. 223–230, 1998
22. Moseley, R.P. and Paget, S., *Liquid-based cytology: is this the way forward for cervical screening?*, Cytopathology, Vol.13, pp. 71-82, 2002
23. Nanda, K., et al., *Accuracy of the Papanicolau test in screening for and follow up of cervical cytologic abnormalities: a systematic review*, Ann. Intern. Med., Vol.132, pp. 810–819, 2000
24. *National Institutes of Health Consensus Development Panel: National Institutes of Health consensus conference statement: cervical cancer.*, J Natl Cancer Inst Monographs, pp. vii-xix, 1996
25. Ohta, Y., Kanade, T. and Sakai, T., *Color information for region segmentation*, Computer Graphics and Image Processing, Vol.13, pp. 222–241, 1981

26. Perez, F. and Koch, C., *Toward color image segmentation in analog VLSI: algorithm and hardware*, International Journal of Computer Vision, Vol.12(1), pp. 17–42, 1994
27. Pratt, W., *Digital image processing*, John Wiley and Sons, 1978
28. Robertson, A.R., *Color Perception*, Physics Today, pp. 24–29, 1992
29. Romani, S., Montseny, E. and Sobrevilla, P., *An Approach to a Fuzzy Color Detection Method, Robust with Regard to Variable Illuminant Level Conditions*, Proceedings NAFIPS 2002 Conference, pp. 69-74, 2002
30. Romani, S., Montseny, E. and Sobrevilla, P., *Obtaining the Relevant Colors of an image through Stability-based Fuzzy Color Histograms*, Proc. of the 12th IEEE Int. Conf. on Fuzzy Systems, pp. 914–919, 2003
31. Rosenthal, D.L., Acosta, D. and Peters R.K., *Computer-assisted re-screening of clinically important false negative cervical smears using PAPNET testing system*, Acta Cytology, Vol.40, pp. 120–126, 1996
32. Rosenthal, D.L., *Computerized scanning devices for Pap smear screening: current status and critical review*, Clinics in Laboratory Medicine, Vol.17, pp. 263–284, 1997
33. Sawaya, G.F., Brown, A.D., Washington, A.E. and Garber, A.M. *Clinical practice. Current approaches to cervical-cancer screening*, N. Eng. J. Med., Vol.344, pp. 1603-1607, 2001
34. Selvaggi, S.M. *Implications of low diagnostic reproductibility of cervical and histologic diagnosis*, JAMA, Vol.285, pp. 1506–1507, 2001
35. Shafarenko, L., Petrou, M. and Kittler, J. *Histogram-based segmentation in a perceptually uniform color space*, IEEE Trans. on Image Processing, Vol.7(9), pp. 1354–1358, 1998
36. Sherlaw-Johnson, C., Gallivan, S. and Jenkis, D. *Low risk women from cervical screening programmes: mathematical modeling study*, B. M. J., Vol.318, pp. 356–361, 1999
37. Smith, A.R. *Color gamut transform pairs*, IEEE Transactions on Computer Graphics, Vol.12, pp. 12–19, 1978
38. Sobrevilla, P., Montseny, E. and Keller, J. *White Blood Cell Detection in Bone Marrow Images*, Proc. of the 18th NAFIPS International Conference, pp. 403-407, 1999
39. Sobrevilla, P., Keller, J. and Montseny, E. *Using a Fuzzy Morphological Structural Element for Image Segmentation*, Proc. of 19th NAFIPS International Conference, pp. 95–99, 2000
40. Solomo, D., Davey, D., Kurman, R., Moriarty, A., O'Connor, D., Drey, M., et al. *The 2001 Bethesda System. Terminology for reporting results of cervical cytology*, JAMA, Vol.287, pp. 2114-2119, 2001
41. Stehman, F.B., Bundy, B.N. and DiSaia, P.J. *Carcinoma of the cervix treated with radiation therapy I: a multi-variate analysis of prognostic variables in the Gynecologic Oncology Group*, Cancer, Vol.67, pp. 2276–2785, 1991
42. Tenenbaum, J.N. *An interactive facility for scene analysis research*, Technical Note 87. Artificial Intelligence Center. Stanford Research Institute, 1974
43. Tominaga, S. *Color image segmentation using three perceptual attributes*, Proc. CVPR'86, pp. 628–630, 1986
44. Tseng, D.C. and Chang, C.H. *Color segmentation using perceptual attributes*, Proc. 11th International Conference on Pattern Recognition, Vol.3(C), pp. 228–231, 1992
45. Vassilicos, P. *Management of suboptimal cytologic smears*, Acta Cytol, Vol.42, pp. 1481, 1998
46. Wilbur, D.C., Prey, M.U., Miller, W.M., Pawlick, G.F. and Colgan, T.J. *The AutoPap System for primary screening in cervical cytology: comparing the results of a prospective, intended-use study with routine manual practice*, Acta Cytol, Vol.42, pp. 214–220, 1998
47. Wyszecki, G. and Stiles, W.S. *Color science, concepts and methods, quantitative data and formulas*, John Wiley and Sons, 1967
48. Yager, R.R. *Families of OWA operators*, Fuzzy Sets and Systems, Vol.59, pp. 125–148, 1993
49. Yagi, D., Abe, K. and Nakatani, H. *Segmentation of color aerial photographs using HSV color models*, IAPR Workshop on Machine Vision Applications, 1992
50. Zaino, R.J., Ward, S., Delgado, G., et al. *Histological predictors of the behaviour of surgically treated stage 1B squamous cell carcinoma of the cervix: A Gynecologic Oncology Group study*, Cancer, Vol.69, pp. 1750–1758, 1992